STUDIA PAEDAGOGICA

VOLUME 26 / NUMBER 4 / YEAR 2021

2021

MASARYK
UNIVERSITY

Studia paedagogica cooperates with the Emerging Researchers' Group
of the European Educational Research Association (EERA).

Citation: *Studia paedagogica* 26, 4, 2021

ISSN 1803-7437 (print)
ISSN 2336-4521 (online)

CONTENTS

Editorial

Studies

EDITORIAL

The quality of educational decision making greatly influences the extent to which schools succeed in developing the talents of all students, in being agile and responsive to change, and in providing a supportive climate for students and teachers. Teachers' decisions profoundly affect students' lives, as they weigh important matters such as retention, promotion, grouping, and tracking. Inequalities in education are predominantly related to decision bias, such as stereotyping or self-fulfilling prophecies. The decisions of school leaders have a tremendous impact on learning, development, and well-being in schools. Although decision making is at the heart of issues of school effectiveness, improvement, and equity, our insight into how educators make decisions in practice is still limited (Earl & Katz, 2006; Harteis et al., 2008). The key aim of this special issue is to broaden our understanding of decision making in education by investigating and discussing different perspectives.

For a long while in education, researchers and practitioners had great trust in teachers' intuitive judgment derived from experience within the teaching profession (Elbaz, 1993; Verloop et al., 2001). During the past decade, the trustworthiness of teachers' intuitive judgment has been questioned. Studies have showed a lack of validity and reliability when the accuracy of teacher judgment was compared with objective measures such as standardized tests (Brookhart, 2001, 2011). Mostly, these studies showed that intuitive teacher judgment disadvantaged low achievers, students with special educational needs, and those from lower social classes (e.g., Brookhart, 2011). This has led to a counter movement with the expectation that decisions would become more standardized and data driven (Mandinach et al., 2008; Schildkamp & Lai, 2013). The initial body of data use research mainly conceptualized data as quantitative indicators of students' cognitive output (Hubbard et al., 2014). More recently, scholars have critiqued this narrow view because it inhibits a full understanding of student competences and has led to undesirable practices (Brown, 2017; Ehren & Swanborn, 2012).

Even more recently, researchers have broadened their view on data and data use. Schildkamp (2019) discussed both formal data (collected deliberately and systematically) and informal data (collected on the fly). Data-based decision making has evolved to data-informed decision making–decisions do not have to be based on data; they should be informed by data. Or, as Earl

(2012) put it: data do not provide answers, they provide tools for thinking. Models of research-informed practices have described how research can be used to improve teaching practices and student outcomes, ultimately leading to improvement at the system level (Brown, 2017). In this special issue, the article by Groß Ophoff and Egger reflects on Educational Research Literacy (ERL) as the ability to access, comprehend, and reflect scientific information as well as to apply the resulting conclusions to problems with respect to educational decisions. The article discusses how crucial the engagement with research is for the process of data-based decision making. This coincides with the idea that both data and research are important for evidence-informed school improvement (Brown et al., 2017).

The rise of data, big data, and data use has also raised new questions related to data ethics. Responsible data use has emerged in education as an important concept. In their article within this journal, Mandinach and Jimerson couple data literacy with an ethical approach to using data—to be an ethical data user means using the right data in the right ways for the right purposes.

In their study, Gutwirth, Goffin, and Vanhoof investigate how Flemish middle school mathematics teachers make sense of school performance feedback data from external standardized tests. They show that the availability of school performance feedback data does not spontaneously spark sensemaking, nor does it necessarily lead to improvements in instructional practice. It appears that teachers' sensemaking of school performance feedback data is a largely intuitive process, grounded in external attributions and often lacking triangulation.

In education, judgment is mostly studied either from a data use or a teacher (tacit) knowledge perspective. However, in the broader field of decision making, recent theories on dual-process approaches indicate that both data-driven and intuitive processes are important for human judgment, and that both have merits and pitfalls (Hogarth, 2014; Klein, 2008). Professional decision making implies a combination of evidence (data and research) and intuitive expertise (Vanlommel, 2018, 2021; Vanlommel et al., 2017). In this special issue, the article by Vanlommel and Pepermans reports on the validation of a Teacher Decision-Making Inventory that combines both data-driven and intuitive dimensions in the different steps of the decision process.

An interesting message is conveyed in the article by Van Gasse and Mol, who explore how teachers use data for student guidance decisions at team meetings. Their qualitative analysis shows that data was only used sporadically, often not in a systematic way, and the depth of inquiry in formulating diagnoses on poor student functioning was low. This clearly implies the need to raise awareness and perhaps to provide adequate training to teachers involved.

Given our view that professional decision making requires a combination of data, research, and intuitive expertise, we also need to broaden the concept of data literacy. Judgment literacy would be more appropriate, describing the competences (knowledge, skills, and attitudes) to collect, combine, and weigh data, research, and intuition to reach informed decisions. In this special issue, Fjørtoft and Morud discuss a specific competence: the ability to make sound judgments about student learning processes, performances, and practical skills. They study assessment decision making in teaching as being highly complex, as teachers are faced with dilemmas such as tensions between different sets of goals (i.e., curriculum, business standards, and student goals) or between tacit and explicit dimensions of learning.

In her article on data-informed decision-making approaches to inform school improvement processes, Fernandes makes the effort to understand the "how" and the "why" of data-informed decision-making systems and their use in practice in the independent sector of Australian schooling. Fernandes concludes with recommendations for improved system capabilities and shows the important role school leaders play in the development of data-informed collaborative school cultures.

Overall, this special issue offers insights on broader competences needed for professional decision making and discusses findings with a dual-process starting point integrating data and intuition. In this special issue, you can find research that starts from student data and articles with a focus on professional capital related to decision making. We believe this broad view on decision making in education offers interesting and inspiring reading for a broad professional community.

Kristin Vanlommel and Milan Pol, Editors

References

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice, 8*(2), 153–169. https://doi.org/10.1080/09695940123775

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brown, C. (2017). Further exploring the rationality of evidence informed practice: A semiotic analysis of the perspectives of a school federation. *International Journal of Educational Research, 82*(1), 28–39. https://doi.org/10.1016/j.ijer.2017.01.001

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning.* Corwin Press.

Earl, L. M., & Katz, S. (Eds.). (2006). *Leading schools in a data-rich world: Harnessing data for school improvement.* Corwin Press.

Ehren, M. C., & Swanborn, M. S. (2012). Strategic data use of schools in accountability systems. *School effectiveness and school improvement, 23*(2), 257–280. https://doi.org/10.1080/09243453.2011.652127

Elbaz, F. (1993). Responsive teaching: a response from a teacher's perspective. *Journal of Curriculum Studies, 25*(2), 189–199. https://doi.org/10.1080/0022027930250208

Fullan, M. (2011). *The six secrets of change: What the best leaders do to help their organizations survive and thrive.* John Wiley & Sons.

Harteis, C., Koch, T., & Morgenthaler, B. (2008). How intuition contributes to high performance: *An educational perspective. US-China Education Review, 5*(1), 68–80.

Hogarth, R. M. (2014). *Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought.* Psychology Press.

Hubbard, L., Datnow, A., & Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation, 42*, 54–62. https://doi.org/10.1016/j.stueduc.2013.10.003

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456–460. https://doi.org/10.1518%2F001872008X288385

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. Data-driven school improvement: Linking data and learning. In E. B. Mandinach & M. Honey (Eds.), *Data-driven school improvement. Linking data and learning* (pp. 13–31). Teachers College Press.

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research, 61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai, & L. Erl. *Data-based decision making in education. Challenges and opportunities* (pp. 177–191). Springer.

Vanlommel, K. (2018). *Opening the black box of teacher judgement: The interplay of rational and intuitive processes.* University of Antwerp, Faculty of Social Sciences, Department of Training and Education Sciences.

Vanlommel, K. (2021). Drivers and obstacles for evidence-informed practice in an autonomous and decentralized educational system: Belgium. In *The handbook of evidence-informed practice in education: Learning from international contexts.* Emerald Publishing Limited.

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research, 83*, 75–83. https://doi.org/10.1016/j.ijer.2017.02.013

Verloop, N., Van Driel, J., & Meijer, P. (2001). Teacher knowledge and the knowledge base of teaching. *International journal of educational research, 35*(5), 441–461. https://doi.org/10.1016/S0883-0355(02)00003-4

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-1

DATA ETHICS IN EDUCATION: A THEORETICAL, PRACTICAL, AND POLICY ISSUE

ELLEN B. MANDINACH,
JO BETH JIMERSON

Abstract

Responsible data use has emerged as an important concept in education, especially in the wake of the COVID-19 pandemic, which continues to highlight inequities. The knowledge and skills to use data effectively and appropriately are at the heart of data ethics. Educators must tightly couple data literacy with an ethical approach to using data—that is, they must be thoughtful about what they choose to do with data, how they go about their work, and how they center their work to benefit, rather than to harm, those engaged in the work of schooling, including students, teachers, families, and other educators. In this article, intended to provoke thought around data ethics among educators, researchers, and policymakers, we take a broad view of what data are and assert that data ethics go far beyond protecting the privacy and confidentiality of data. To be an ethical data user means using the right data in the right ways for the right purposes. The article lays out a context for data ethics, demonstrates how ethics are coupled with data literacy, provides examples of data ethics in practice, and recommends steps for strengthening ethical data use in practice.

Keywords

data ethics, data literacy, data-driven decision making

Introduction

Ethics in education, leadership, and decision making have been a core concern for quite some time (Starratt, 2004, 2012). Data-driven decision making (DDDM) is similarly not new, having most pointedly accelerated with the birth of modern accountability movements that foregrounded broad-scale standardized testing regimes seeking to use data in comparative (and sometimes punitive) ways (Beadie, 2004; Darling-Hammond, 2007; Dorn & Ydesen, 2014; Kuhn, 2013; Militello et al., 2013). Despite ethics in data use being at least an implicit thread running through empirical research over the past decades (e.g., Booher-Jennings, 2005; Daly, 2009; DeMatthews & Serafini, 2019; Vasquez Heilig & Darling-Hammond, 2008), the role of ethics in data use has become a more explicit focus as of late. Perhaps because of what new technologies allow educators to do (apart from questions of what they should do), and in part due to crises like the 2020-21 pandemic, which pressed educators into broader and newer forms of technology use (and, for many, introduced questions about privacy, data security, and the degree to which the worlds of school, workplace, and home do or should overlap), ethical data use has emerged as a critical topic.

Long before the pandemic, data use was inextricably bound up with accountability systems, with accompanying pressures to "game the system" and use data inappropriately to performative ends (Aronson et al., 2016; Booher-Jennings, 2005; Nichols, 2021; Nichols & Berliner, 2007). "Gaming the system" in the educational context refers to the manipulation of student performance indices to maximize accountability outcomes, such as focusing only on students who are near passing, to the exclusion of all other students. The result of this myopia is the further marginalization of already challenged students, because accountability measures often fail to accurately reflect the performance of some student groups (Datnow, 2017; Datnow & Park, 2018; Garner et al., 2017). Mandinach and Schildkamp (2021) explored how the persistent linking of data use with accountability systems created misconceptions about data use and a need to shift the dialogue towards data for continuous improvement. Mandinach et al. (2019, 2021) have further asserted that educators need to use diverse data sources to address the whole child, and the fullness of data literacy requires educators to assume an asset and equity mindset.

Starratt (2004) noted that one conundrum facing education is that "… the issues that school leaders face tend to be presented and interpreted primarily as technical, rationalizable problems resolvable by technical, rational solutions" (p. 4). Focusing on the use of data without an ethical lens can also reify thinking that DDDM processes are "successful" when they effect desired changes in high-profile metrics. However, we have seen instance after instance of improvement in (short-term) metrics that derived from unethical acts (e.g., Daly, 2009; DeMatthews & Serafini, 2019; Vasquez Heilig & Darling-Hammond, 2008). In a later work, Starratt (2012, pp. 8-9) clarified:

> …being ethical refers to behaving in ways consistent with internal, self-appropriate principles that one can articulate and that, at least sometimes, lead persons to go beyond self-interest. […] Ethical persons behave in certain ways because they are convinced that it is the right thing to do, because doing the right thing is tied up with their identity.

Data ethics, in an educational context, mean that leaders not only use data to inform decision making in ways that affect improved academic, social, and behavioral outcomes, but that they also use data in ways that increase awareness of and acknowledge underlying issues such as inequities, political pressures, and barriers to change. Moreover, school leaders work to articulate their professional values, beliefs, and principles, and subsequently use data in ways that align with those values. This also means that ethical data use is not aimed at achieving "quick wins" in terms of improvements in reportable data if these come at the expense of the long-term good of students and communities; data use is purposed toward unearthing and addressing root causes of student underperformance and to centering the long-term good of the students and communities in data-driven dialogues and processes.

Data users too often have fallen into the trap of using narrowly defined data to inform the work of schooling mainly in regard to accelerated gains in metrics related to academic progress—and even then in areas privileged by test-driven accountability policies (Brighouse et al., 2018). Ethical data use can be a part—*should* be a part—of identifying and addressing the full needs of children beyond accountability metrics. To engage in the "profoundly moral work" of education requires an ethics of data use such that data are used to gain a comprehensive understanding of each student based not just on academic performance but also on the student's personal story, contextual factors, interests, and strengths (Mandinach et al., 2019, 2021).

In the United States, regulations such as the Family Educational Rights and Privacy Act (FERPA) protect educational student data in terms of privacy and confidentiality. But data ethics are broader than compliance with legal regulations. Data ethics are about using the right data for the decision at hand, properly analyzing the data and drawing accurate interpretations, outlining potential paths forward for action, articulating and weighing the ethical issues involved with potential actions based on data, and selecting actions that are ethically defensible (Mandinach & Gummer, 2021a). It is important to note, however, that interpretations are not straightforward and there may be no single "right" interpretation of some data. Wrestling with what is ethically defensible requires school leaders to collaborate with others (so they are not captive to their individual personal or professional desires or fears) in order to see issues and actions from different perspectives (Mandinach & Jimerson, 2021a). This enables leaders to more accurately weigh how both the short- and long-term benefits and burdens of their decisions fall on students, teachers, and others in the school and community. Acting on data without internal and external checks on ethics can lead leaders to engage in cognitive fallacies. Such cognitive fallacies include cherry picking which data to use or reject, using incomplete data, and privileging certain metrics and excluding others to provide a narrowed view of the phenomenon being examined (Geckoboard, n.d.). Such fallacies can lead to inaccurate interpretations and poor decisions; unethical data use centers the good of leader(s) at the expense of students, school culture, and learning—that is, leaders' careers can profit from improvements in highly publicized high-stakes metrics, even if the methods used to attain those improvements include practices that fail to enrich learning and long-term opportunities for the students they ostensibly serve (see Daly, 2009; Vasquez Heilig & Darling-Hammond, 2008).

In cases in which school leaders profit (in terms of prestige and/or promotion) by leveraging high test scores through "drill-and-kill" strategies as opposed to more robust instruction and engagement with rich curricular materials, students are, in effect, positioned as mere means to an end for leaders—as pawns in a larger game. When leaders practice ethical data use, the support and development of students is the goal in and of itself, and benefit to leaders is a mere byproduct of this means-ends positioning. Fundamentally, data ethics are about being data literate and using those skills, knowledge, and dispositions to use data effectively and responsibly first and foremost for the support and development of students.

Theoretical Background

Definitions

Throughout this paper, the terms *data use* and *data-driven decision making* refer to the systematic collection, analysis, and application of data to inform educational decision making. *Data literacy* is a composite of skills, knowledge, and dispositions that educators need to use data effectively and responsibly. It is the ability to use a skill set to make all sorts of educational decisions based on diverse data and actionable information to inform practice (Mandinach & Gummer, 2016).

By *data ethics*, we mean the ability not only to use appropriate data for appropriate purposes, but to apply reasons that prioritize the long-term benefit of students. This aligns with a definition included in the *Data Ethics Framework* provided by the US General Services Administration (2020; p. 9): "the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society, and maximizing the public good."

We acknowledge that our approach to data ethics is reflective of Kantian ethics (Fieser, 2003) in that we hold that ethical data use requires that leaders refrain from decisions that treat others (e.g., students, teachers) as merely a means to an end; instead, ethical data use should center efforts that reflect a valuing of students as worthy ends in and of themselves. This is in contrast with a utilitarian approach, which would judge the ethics of a decision or act solely upon the outcome of the act (Fieser, 2003). The judgment of the ethical nature of a data-driven decision, then, inheres in the decision itself, rather than in the outcome. Tenets of data ethics therefore include acting with integrity, being accountable, being transparent, and protecting privacy and confidentiality. Finally, we assert that data ethics are inextricably coupled with data literacy and in fact, in the ideal, data ethics are a requisite component of data literacy. But we question whether an educator can make an ethical, data-based decision without sufficient data literacy, and whether an educator can use data literacy skills to make a data-based decision that results in a good outcome but that is unethical. The interplay of skills is complex and requires exploration.

Ties to Data Literacy and Cognition

Mandinach and Gummer (2016) have conducted theoretical studies for over a decade to provide a definition of what it means for educators to be data literate. They identified 53 skills and types of knowledge that teachers need to use data effectively. In addition, they identified several habits of mind or dispositions that are generic to teaching but are essential to data literacy, such as the belief that all students can learn, the importance of collaboration, and communication. We position data ethics as a foundational component of data literacy that works in concert with other key educational dispositions and cognitive skills. According to the *data literacy for teachers* construct (Mandinach & Gummer, 2016), data ethics are seen as both a skill and a disposition that educators need in order to use data responsibly. For our purposes here, we generalize the construct from teachers to leaders, noting that the fourth component, transforming information into an instructional decision, would involve many kinds of administrative decisions for leaders. Data literacy includes the ethical use of data as well as honoring those to whom data belong through protection of data privacy. Mandinach and Nunnaley (2021) argued that, given a continuum of expertise, individuals who exhibit advanced or high-capacity data literacy for teaching (DLFT) skills, are by definition, using data effectively and responsibly. More novice users may be less likely to know how to invoke appropriate DLFT skills. DLFT skills are not used in isolation but most often are used as a composite (Beck & Nunnaley, 2021), with different subsets of skills being used at various times. This means that of the 53 DLFT skills, educators rarely engage one skill at a time but rather a composite of skills and knowledge.

When we look across the DLFT skills, a major emphasis would be on selecting the right data (more than just student performance indices) to address specific questions, drawing appropriate interpretations, and taking actionable steps. A key source of knowledge is understanding the principles of consequential validity drawn from the interpretation (Cronbach, 1988; Messick, 1989), and avoiding the many cognitive fallacies, such as detecting patterns where none exist, that Kahneman and Tversky (1973, 1984; Tversky & Kahneman, 1971, 1974) outlined in their work, as well as confirmation bias (Nickerson, 1998).

Further, DLFT includes the use of data with integrity. Data literate educators should understand the concept of data quality, meaning that the data they use have relevance, completeness, timeliness, the right granularity, and accuracy (Mandinach & Gummer, 2016). Data literate educators should know how to examine and analyze data appropriately and use data displays to represent their results without distorting or misrepresenting the findings. They should know which assessments to use for what purposes (e.g., Coburn & Talbert, 2006; Militello et al., 2013). They should know how to communicate accurately with data. They should understand that data use is an iterative rather than a finite process, one in which they examine their own findings and question them for accuracy and appropriateness. Educators should use data to interrogate their own implicit biases rather than confirm them, using an asset-based mindset that avoids deficit thinking (Bertrand & Marsh, 2021). Without the guardrails of data ethics, educators can misuse data and end up responding to accountability pressures in dysfunctional ways (Nichols, 2021), gaming the system (Booher-Jennings, 2005), and marginalizing challenged students (Datnow, 2017; Datnow & Park, 2018). Responsible data use is about an equity model, using data responsibly to address the diverse needs of all learners.

Data Ethics in Practice: Focal Scenarios

In this section, we offer six brief scenarios. These are illustrative in nature, composites drawn from both authors' personal and professional experiences working with education and educators, and are constructed to allow readers to consider broadly how data ethics impact educational practice. To help readers consider multiple levels of data ethics application, we provide scenarios situated at the district, school, and classroom levels that highlight data ethics issues pre- and intra-COVID-19. We purposefully do not provide interpretations of the scenarios, as our goal is to present scenarios that spark reflection and dialogue among readers. We do, however, note core ethical issues at the heart of each scenario. After presenting all six scenarios, we connect the concepts illustrated in each to the broader research literature in the "General Commentary" that concludes this section.

Classroom Decision Making

Third-grade teacher Sarah Shuster collects data on students' reading levels and skills through 1:1 assessments and read-alouds, whole-class assessments, talking with students about their assigned and self-selected reading choices, and via the school's new learning management system (LMS). The LMS has an assessment mode that engages students for 20 to 30 minutes per week; it uses these data to help the system "learn" and to link "recommended learning activities" to students' demonstrated knowledge and skills. Students access activities from school or home, and activities contribute to the overall reading grade.

One of her students, Lola, consistently selects books at the 5th grade and higher reading levels, and Ms. Shuster and Lola have engaging conversations about the material. Lola does well in class-wide assessments, but the LMS consistently reports Lola's performance at the 1.5 to 2.0 grade level range and assigns her tasks that Ms. Shuster thinks are below Lola's ability. The school expects students to use the LMS; teachers are expected to send reports to parents. Lola's reports from the LMS and from Ms. Shuster are consistently in conflict. Lola's parents want to meet with Ms. Shuster and the principal, as they now fear Lola is falling behind in reading and wonder if Ms. Shuster is able to address her needs. Ms. Shuster wants to tell Lola's parents they have nothing to worry about, but she also wonders if that is true. Maybe she is missing something, or worse—maybe she is not looking for evidence of Lola's gaps. After all, what would that indicate about students she had taught prior to this year—students she thought were flourishing?

This scenario highlights the inevitable coupling of data literacy and data ethics. In terms of data literacy, Ms. Shuster has to determine whether she should privilege her own data and observations or that of the LMS. Regarding data ethics, she must interrogate her motives behind which she chooses to privilege. Simply moving forward and assuming her interpretations are accurate centers her own comfort and well-being (e.g., she can avoid conflict, protect her reputation with the principal, and continue with her practice unchanged). Centering the long-term well-being of Lola may require that she seek out more data to determine whether her instruction is part of the problem, and this could require reflection, dedication to professional learning and growth, and substantial changes in practice.

Campus Decision Making

Turing Middle School Principal Eric West wants to find a new way to recognize students for their character, hard work in the classroom, and positive impact on the campus. The school has grade-based honor societies and student government (which often turns on a popular vote). He creates a program called the "Turing Ten"; ten students from each grade who are to be featured on a prominent bulletin board in the school's front hallway. Student pictures and profiles are featured, and students in each month's "Turing Ten" earn coupons to use in the school's cafeteria and athletic event concession stands. Parents receive a bumper sticker and cookies delivered to their home or workplace in recognition of their children.

At first, Mr. West operated the Turing Ten purely by teacher nomination, but he noticed the same students who were typically recognized and were well-liked quickly appeared among the Turing Ten. He knows that many teachers on campus already use an app that communicates merits and demerits, sometimes in real time. If all teachers used the app, he could simply collect those data and run a report to inform the monthly list. In fact, students could already see their own point totals—if they knew what the average point level of a Turing Ten student was and what the school average was, it could help motivate them to improve aspects of what the school identified as good citizenship, such as good attendance, turning work in on time, following the dress code, and moderating behavior.

Here, data ethics would require that Mr. West at least consider potential harmful or unintended consequences of using data to foster a competitive environment in the school—a contest where for some to win, others must inevitably lose. Data ethics would also require consideration of the potential effects of making students' data available to peers; even if masked, the premise that a students could improve their own performance, yet still encounter data outputs that continually show them falling below the school average or even near the bottom of the list could have detrimental effects. Ethics require Mr. West to at least look beyond the anticipated positive outcomes that will be afforded to some students to determine whether potential negative consequences for others outweigh the value of his plan.

District Decision Making

Superintendent Marcia Bales is leading the district through rezoning, as two new elementary schools and a new middle school will open in the next few school years. The school district was recently assessed a "B" overall in the state's school accountability system. Two elementary schools were rated "C" and one was rated “F”; all other schools were in the “B” or even “A” range. Dr. Bales has been at the forefront of updating curriculum and teacher training. She knows that some of her “wins” in the district have come because the communities that the schools serve have largely seen improvements. Only five years ago, the district was rated a “C” district, with five “C” campuses and two “F” campuses. She has received recognition from local realty companies and the local Chamber of Commerce, because as perceptions of school quality rise, so do property values across much of the district.

Dr. Bales wants to keep pressing for improvement at all schools, but also begins to wonder if the rezoning is an opportunity to combine some programs and schools to maximize performance and cluster lower-performing areas of the district at two main elementary campuses. That would ensure higher ratings for the district (possibly even enable the district to receive an “A” rating) and in the process allow her to target more resources (programs, personnel, facilities) to the two campuses. Maybe they would not remain low performing for long, she thinks, given the additional support.

The data ethics issue here is whether Dr. Bales is indeed aiming to make the targeting of resources more efficient (centering the learning needs of students) or whether she is manipulating the rating system to advantage the district (and possibly to bolster her reputation for effecting improved outcomes). If the latter, she is in effect using students as a means to an end that benefits her by using gaming strategies to obscure struggling schools and students; these strategies could actually put students’ long-term achievement in jeopardy.

Classroom Decision Making (Intra-pandemic)[1]

Just prior to the pandemic and the pivot to virtual learning, teacher Mr. Torres had assessed his 1st grade students in-person and via AIMSweb. Brooke came to him not knowing her alphabet; in half a year she had progressed to a mid-kindergarten reading level. The week before the shutdown, Mr. Torres had assessed Brooke at a Developmental Reading Assessment (DRA) level of 4,

1 To be sure, other forms of crisis can push schools to new applications of technology and data use. Natural disasters have forced schools to close, realign, or find new ways of operation for periods from a few weeks to months or even years. However, we use the COVID-19 crisis as it is unique in pushing so many schools to new ways of operating—and operating at distance—for so long and at the same time. As a pandemic in which so many pushed for normalcy in the midst of a wildly abnormal context, it allows us to uncover and examine some examples pertinent to data ethics.

and she had begun working through the *Biscuit* book series with assistance. Two months into virtual learning, Mr. Torres had students complete an online reading assessment, and Brooke's score came out at the 8th grade level. Brooke was making progress, but not *this* much. Mr. Torres called Brooke's mother, who insisted, "Brooke has been flourishing in virtual learning with us right here to help!"

Trying to give the parent an out for possibly helping Brooke a bit too much, Mr. Torres suggested, "perhaps one of one of her older siblings tried to help her a bit… it's really important to have an accurate assessment so I know how to help Brooke." But Brooke's mother was steadfast. She had watched Brooke take the test—it was just evidence that virtual learning was working well for her! During class, when Brooke read to Mr. Torres, he was pleased with her reading, but also recognized that she was coming closer to being "on level" for a 1st grade student; she was in no way reading on the level suggested by the online assessment. Mr. Torres paused, trying to determine how he would assign a progress grade for Brooke in reading and how he could get data to guide the next steps with Brooke without alienating her family.

Mr. Torres is demonstrating elements of data literacy; he is using multiple measures, collaborating with family, and working to move from interpretation to action. In terms of data ethics, he has a choice to make: he can press the issue of inaccurate data with the family or simply ignore the issue and move forward. Confronting the issue could alienate Brooke's family and result in hassles in terms of meetings with school administrators, but also focuses all collaborators on the importance of capturing accurate data to inform instruction and on the importance of everyone who has a stake in Brooke's progress accepting her current performance so that realistic plans can be made to support her in moving forward. Ignoring the issue is likely to result in less conflict for Mr. Torres, though he may have to develop workarounds for teaching and assessing Brooke, which could involve deceiving her family by allowing them to believe that Brooke is more advanced than she actually is as a reader or that their narrative has been accepted.

Campus Decision Making (Intra-pandemic)

Principal Nat Lawrence is frustrated. The year has been marked by virtual learning, then hybrid learning, then "in-person" learning with long absences as students and teachers move in and out of quarantine. Then the state determined all students still had to take the state-mandated accountability exam, and Mr. Lawrence spent a week trying to figure out how to get students rotated into the building safely to facilitate the test. Then, under pressure, the state decided that students participating in virtual learning would not have to take the test—only those attending in-person would be tested. Mr. Lawrence thinks of four families in particular—families whose children have already struggled with anxiety due to multiple quarantines and, in each family, more than one death of a loved one. "The last thing those children need to be doing is taking a standardized exam," he thinks. Mr. Lawrence believes that at least three of the five students across those families would do well anyway, but he looks at the phone and considers calling them to tell them that if they opt for virtual learning for a few weeks, they can effectively bypass this year's test. After all, what more could the test reveal that they do not already know about the children, given their internal assessment systems?

Principal Janelle Rogers is frustrated, too. She has seen the same issues Mr. Lawrence has. She was new to her school the year the pandemic hit and had only begun leading much-needed changes and improvement at the campus. She has lost 10 teachers to retirements and resignations this year. She fought for safety protocols to have students in school if they so chose, and she had Wi-Fi routers and laptops delivered to families of students (and to teachers) who needed resources to work from home. "After all we've been through, the state is going to judge us on a standardized test?" she thinks. "Compared to who, exactly, if nobody learning from home takes the exam? Like we aren't assessing with benchmarks and the learning platform every other week?" Ms. Rogers looks at in-person and virtual rosters and compares them to previous years' data. Many of the at-home learners posted high scores in previous years, and now they are out of the testing pool entirely. She wonders how her own performance will be judged this year, with so many students and teachers in and out of attendance. She looks at the phone and wonders if she can talk some of the virtual learners into coming in to take the exam to shore up campus scores and buy her another year to keep moving the campus forward.

Here we see competing scenarios with two principals both considering talking to families to sway them to have their children be tested or to opt out of testing. Apart from the ethical issue of potentially coercing families to make a choice they would otherwise not make, particularly in a situation where the health and safety of children and families may be compromised, these scenarios provide contrasting approaches. Which leader (if either) is centering the good of students? Which is using students as a means to an end? For what reasons would dissuading a family from having their child be tested, or persuading a family to test their child, be ethical? Under what circumstances are such actions unethical? Issues here include gaming accountability systems to ensure that high performing students get tested while others are excluded, with the assumption that a particular sample of students will make campus data look better.

District Decision-Making (Intra-pandemic)

As part of a new emphasis on mental health and well-being in the district, Assistant Superintendent Jac Elliot is considering the introduction of a well-being indicator to district-owned laptops (provided to all secondary students and employees); students and employees could also download an app and log in to the system from personal devices. The system would prompt users periodically to report in with a general indicator of mood via emojis and would also push well-being-oriented recommendations to users. A button in the app would allow users to indicate whether they were feeling particularly stressed or down and would like to be connected to a counselor via chat or phone. Users could also schedule counseling visits (school counselors for students, HR-related counselors for employees) through the app. Though the prompts would be pushed to all district-owned devices and app users daily, the choice of when and how to respond would be wholly up to students and employees.

Jac particularly likes that the district can run risk reports, so if a pattern emerges that is cause for concern, a counselor could, after approval from a risk assessment team, contact the student or employee to offer help. At the same time, Jac is uncertain of how the app stores data, and recognizes that to get participation and honest responses, people need to trust in the security of the system; having the ability to run the backdoor risk reports inherently compromises user trust. Jac wonders how to proceed, given dual concerns for security and privacy on one hand, and on wellness support and crisis prevention on the other. They wonder how they will present the pros and cons of the app and integration into district systems at the next board of trustees meeting.

This scenario suggests that Jac is working to apply data ethics by trying to balance student and employee well-being (and helpful intentions) with personal and data privacy. Another issue highlighted in this scenario is the responsibility of district leaders to ensure they understand how data are collected, stored, and used; if a vendor can access and use (e.g., sell) data, then users of the app should be informed as to when and under what conditions their data may be so used; transparency is key. Compromising confidentiality or brokering privacy—particularly without being transparent to app users so they can make informed decisions about entering their personal data—would be unethical, regardless of whether the app provided benefit to some students and employees.

Commentary on Examples

Though different threads run through the examples, they provide grist for discussing a range of issues pertinent to data ethics. One issue is that in each decision, the educator has to determine whether the data being used are those most appropriate to the articulated goals and whether accessing and using the data in the ways intended abides by privacy regulations. Even if those hurdles are cleared, educators must query whether robust and diverse data sources are being used to answer the questions asked, from “can Lola really read well or is she struggling?” to “does a merit/demerit app really give unbiased data on ‘good citizenship’?” to “under what circumstances should students be taking state assessments in a crisis, and how can those data be used?” to “are we helping people by capturing data to provide mental health support?”

The scenario with Ms. Shuster raises questions of how to fit educator judgment and diverse data sources alongside data generated via learning analytics (e.g. Lupton & Williamson, 2017). The scenario at Turing Middle invites questions of the appropriateness of data surveillance in fostering a competitive school culture or forcing routine reporting on what may be minor issues to parents throughout the workday (e.g. Lupton & Williamson, 2017; Manolev et al., 2019). The scenario involving potential rezoning raises questions related to gaming accountability systems in the pursuit of prestige or even in the pursuit of reform (e.g. Aronson et al., 2016). The scenarios dealing with mandated testing during the pandemic raise questions about the influence afforded a single (though state-sanctioned) data point (e.g. Roegman et al., 2021) and the appropriate uses of data in a context of complex and competing pressures on leaders (e.g. DeMatthews & Serafini, 2019).

All scenarios require educators to engage in thoughtful data use—from matching data to driving questions, through collecting and interpreting data, to identifying potential decisions and to determining actions—and to doing so hand-in-hand with the question: Is this process/action/decision ethical?[2]

---

2 Though we are tempted to ask if the process/action/decision is “ethical and equitable” to underscore the importance of centering equity within data ethics, our assertion is that if a “data-driven decision” does not move systems and practices towards equity, then it is inherently unethical, as it (in effect) further privileges some students at the expense of others.

Recommended Steps

We have provided a landscape view of why data ethics are important in educational practice, having given background, theoretical grounding, and examples. We conclude with some general, recommended steps that can be taken to bring awareness to and action in practice about the implementation of data ethics. Some of these recommendations are drawn from our prior work (Mandinach & Gummer, 2021b; Mandinach & Jimerson, 2021a). The intent of these recommendations is not only to bring awareness to the importance of data ethics but also to provide some concrete suggestions for actionable steps that can be taken to help educators and educational agencies use data more responsibly.

A first recommendation is to bring *awareness* to the importance of data ethics. Educators work in contexts where the "techno-friendly obsession within education encourages the prolific spread" of tools and strategies educators can use to identify, monitor, surveil, assess, and respond to perceived student needs (Manolev et al., 2019). However, being rooted in data ethics helps educators recognize when they ought to refrain from doing something that is technically do-able and technologically easy, or when they need to push further than is logistically simple to get the data needed to inform problems adequately. Building awareness of data ethics includes changing the messaging, particularly in expanding the notion of what data ethics are; namely, that data ethics are more than just the protection of data privacy and confidentiality; data ethics require appropriate and effective data use. Thus, the messaging includes moving the conversation to responsible data use, with how the data are being used and the validity of interpretation and action being central.

Following from the messaging is the need to build educator capacity to use data responsibility. Capacity building must begin in pre-service and be sustained throughout educators' careers, through professional development and technical assistance. To do this, educator preparation and educational agencies must recognize the importance, take action by requiring educators and candidates to be literate about data ethics, and provide opportunities for knowledge acquisition. Relatedly, this requires support from professional organizations and state education agencies to include data ethics in state and professional standards. Currently, there are limited resources to help educators acquire the needed skill set, so there must be an effort to develop relevant materials beyond those that exist for data privacy (Mandinach et al., 2021; Mandinach & Jimerson, 2021b), broadening the resources to data ethics and responsible data use.

Related to messaging and data literacy, there is a need for educators to confront issues around confirmation bias (Mandinach & Gummer, 2021b), the impact of accountability system pressures on appropriate data use (Nichols, 2021), the need to assume a whole child perspective and an equity mindset (Datnow, 2017; Datnow et al., 2021; Mandinach & Mundry, 2021), and the detrimental effects of deficit thinking/framing (Bertrand & Marsh, 2021).

There is much work to be done around data ethics in terms of both research and implementation in practice. This article provides both grounding in the issues and a springboard to further dialogue and progress in the field. For a more thorough examination of the issues at play related to data ethics, we recommend the book by Mandinach and Gummer (2021a) that examines theories that pertain to data ethics, the landscape of regulations, how accountability impacts data ethics, and use cases of how data ethics are being implemented in educational settings.

References

Aronson, B., Murphy, K. M., & Saultz, A. (2016). Under pressure in Atlanta: School accountability and special education practices during the cheating scandal. *Teachers College Record*, *118*(14), 1–26.

Beadie, N. (2004). Moral errors and strategic mistakes: Lessons from the history of student accountability. In K. A. Sirotnik (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 35–50). Teachers College Press.

Beck, J. S., & Nunnaley, D. (2021). A continuum of data literacy for teaching. *Studies in Educational Evaluation*, *69*(2), 100871. https://doi.org/10.1016/j.stueduc.2020.100871

Bertrand, M., & Marsh, J. (2021). How data-driven reform can drive deficit thinking. *Phi Delta Kappan*, *102*(8), 35–39. https://doi.org/10.1177/00317217211013936

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231–268. https://doi.org/10.3102/00028312042002231

Brighouse, H., Ladd, H. F., Loeb, S., & Swift, A. (2018). Good education policy making: Data-informed but values-driven. *Phi Delta Kappan*, *100*(4), 36–39. https://doi.org/10.1177/0031721718815671

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, *112*(4), 469–495. https://doi.org/10.1086/505056

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3–17). Lawrence Erlbaum.

Daly, A. J. (2009). Rigid response in an age of accountability: The potential of leadership and trust. *Educational Administration Quarterly, 45*(2), 168–216. https://doi.org/10.1177%2F0013161X08330499

Darling-Hammond, L. (2007). Standards and accountability movement needs to push, not punish. *Journal of Staff Development*, *28*(4), 47–50.

Datnow, A. (2017). *Opening or closing doors for students? Equity and data-driven decision-making.* https://research.acer.edu.au/cgi/viewcontent.cgi?article=1317&context=research_conference

Datnow, A., Lockton, M., & Weddle, H. (2021). When data use raises equity and ethical dilemmas in schools. In E. B. Mandinach & E. S. Gummer (Eds.), *The ethical use of data in education*: *Promoting responsible policies and practices* (pp. 216–231). Teachers College Press.

Datnow, A., & Park, V. (2018). Opening or closing doors for students? Equity and data use in schools. *Journal of Educational Change*, *19*(5), 131–152. https://doi.org/10.1007/s10833-018-9323-6

DeMatthews, D. E., & Serafini, A. (2019). Do good principals do bad things? Examining the bounds of ethical behavior in the context of high-stakes accountability. *Leadership and Policy in Schools, 20*(3), 1–20. https://doi.org/10.1080/15700763.2019.1668023

Dorn, S., & Ydesen, C. (2014). Towards a comparative and international history of school testing and accountability. *Educational Policy Analysis Archives*, *22*(115), 1–7. http://doi.org/10.14507/epaa.v22.1913

Fieser, J. (2003). *Ethics*. The internet encyclopedia of philosophy. https://iep.utm.edu/ethics/#SH2a 5/27/2021

Garner, B., Thome, J. K., & Horn, I. S. (2017). Teachers interpreting data for instructional decisions: Where does equity come in? *Journal of Educational Administration*, *55*(4), 407–426. https://doi.org/10.1108/JEA-09-2016-0106

Geckoboard. (n.d.). *Data fallacies*. https://www.geckoboard.com/best-practice/statistical-fallacies/

General Services Administration. (2020, September). *Data ethics framework*. Federal Data Strategy. https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. https://doi.org/10.1037/h0034747

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kuhn, K. (2013). *Test and punish: How the Texas education model gave America accountability without equity*. Park Place Publications.

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society, 19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Mandinach, E. B., Cotto, J., Rastrick, E., Siegl, J., Vance, A., & Wayman, J. C. (2021, October). *Student data privacy and data ethics scenarios*. Washington, DC and San Francisco, CA: Future for Privacy Forum and WestEd. https://studentprivacycompass.org/resource/scenarios-user-guide/

Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators: Making it count in teacher preparation and practice*. Teachers College Press.

Mandinach, E. B., & Gummer, E. S. (Eds.). (2021a). *The ethical use of data in education: Promoting responsible policies and practices*. Teachers College Press.

Mandinach, E. B., & Gummer, E. S. (2021b). The landscape of data ethics in education: What counts as responsible data use. In E. B. Mandinach & E. S. Gummer (Eds.), *The ethical use of data in education: Promoting responsible policies and practices* (pp. 35–55). Teachers College Press.

Mandinach, E. B., & Jimerson, J. B. (2021a). The role of the district, school, and classroom to ensure ethical use of data: It's more than just FERPA. In E. B. Mandinach & E. S. Gummer (Eds.), *The ethical use of data in education: Promoting responsible policies and practices* (pp. 125–143). Teachers College Press.

Mandinach, E. B., & Jimerson, J. B. (2021b). *User's guide for data privacy and data ethics scenarios for leaders*. WestEd.

Mandinach, E. B., & Mundry, S. E. (2021). Data-driven decision making and its alignment with educational psychology: Why data are more than student performance results. In

S. L. Nichols & D. Varier (Eds.), *Teaching on assessment* (pp. 269–291). Information Age Publishing.

Mandinach, E. B., & Nunnaley, D. (2021). The role of professional development providers in training data ethics. In E. B. Mandinach & E. S. Gummer (Eds.), *The ethical use of data in education: Promoting responsible policies and practices* (pp. 144–172). Teachers College Press.

Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, *69*, 100842. https://doi.org/10.1016/j.stueduc.2020.100842

Mandinach, E. B., Warner, S., & Mundry, S. E. (2019, November 6). *Using data to promote culturally responsive teaching: Session 1* [webinar]. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands, Washington, DC, USA. https://www.youtube.com/watch?v=ltEf8M4RUoQ

Mandinach, E. B., Warner, S., & Mundry, S. E. (2020, June 4). *Using data to promote culturally responsive teaching: Session 2* [webinar]. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands, Washington, DC, USA. https://www.youtube.com/watch?v=2ZZuyFXpqZM

Manolev, J., Sullivan, A., & Slee, R. (2019). The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media and Technology*, *44*(1), 36–51. https://doi.org/10.1080/17439884.2018.1558237

Messick, S. J. (1989). Validity. In. R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103.). Macmillan.

Militello, M., Bass, L., Jackson, K. T., & Wang, Y. (2013). How data are used and misused in schools: Perceptions from teachers and principals. *Education Sciences*, *3*(2), 98–120. https://doi.org/10.3390/educsci3020098

Nichols, S. L. (2021). Educational policy contexts and the (un)ethical use of data. In E. B. Mandinach & E. S. Gummer (Eds.), *The ethical use of data in education: Promoting responsible policies and practices* (pp. 81–97). Teachers College Press.

Nichols, S. L, & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools.* Harvard Education Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Roegman, R., Kenney, R., Maeda, Y., & Johns, G. (2021). When data-driven decision making becomes data-driven test taking: A case study of a midwestern high school. *Educational Policy, 35*(4), 535–565. https://doi.org/10.1177%2F0895904818823744

Starratt, R. J. (2004). *Ethical leadership* (1st ed.). Jossey-Bass.

Starratt, R. J. (2012). *Cultivating an ethical school* (1st ed.). Routledge.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105–100. https://doi.org/10.1037/h0031322

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vasquez Heilig, J., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, *30*(2), 75–110. https://doi.org/10.3102%2F0162373708317689

Corresponding authors
Ellen B. Mandinach
WestEd, Arizona, USA
E-mail: emandin@wested.org

Jo Beth Jimerson
Educational Leadership, College of Education, Texas Christian University, Texas, USA
Email: j.jimerson@tcu.edu

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-2

ASSESSMENT OF GERMAN AND AUSTRIAN STUDENTS' EDUCATIONAL RESEARCH LITERACY: VALIDATION OF A COMPETENCY TEST BASED ON CROSS-NATIONAL COMPARISONS

JANA GROß OPHOFF,
CHRISTINA EGGER

Abstract

Educational Research Literacy (ERL) is the ability to access, comprehend, and consider scientific information and to apply the resulting conclusions to problems connected with educational decisions. It is crucial for the process of data-based decision making and–corresponding to the consecutive phases–defined as the conglomeration of different facets of competence, including information literacy, statistical literacy, and evidence-based reasoning. However, the engagement with research in educational contexts appears to have some difficulties. This is even more remarkable as the state of knowledge about actual teacher competency levels remains unsatisfactory, even though test instruments for assessing research literacy have been developed in recent years. This paper addresses the question of whether such a test developed in the specific context of German study programs in (teacher) education can be applied to other national contexts, in this case to Austrian teacher education. An investigation of the construct validity under consideration of the psychometric structure and group differences on item level is necessary for ensuring the fairness of cross-national comparisons. Based on multidimensional item response theory models, samples from Germany (n = 1360 students, 6 universities) and Austria (n = 295 students, 2 universities) are investigated in terms of measurement invariance between the two countries. A comparable psychometric structure and at least partial measurement invariance with no particular advantage for either sample could be demonstrated. This is an indication that the presented test instrument can be validly applied to assess the research literacy of teacher training students in both countries.

Keywords

Educational Research Literacy, teacher education, construct validity, measurement equivalence, differential item functioning

Introduction

As early as 1999, Davies stated that educational professionals at all levels should be able to (a) to pose answerable questions; (b) search for relevant information; (c) read and critically appraise evidence; (d) evaluate; and (e) use the resulting conclusions for educational decision making. These requirements correspond to the stages of research engagement in the sense of a complex, cognitive, knowledge-based problem-solving cycle. Thus, it is not surprising that corresponding process descriptions can be found in conceptual frameworks of data-based decision making (e.g., Groß Ophoff & Cramer, in press; Mandinach et al., 2008; Marsh, 2012; Schildkamp & Kuiper, 2010; Schildkamp & Poortman, 2015; Schratz et al., 2018). These models consider teachers' competent engagement with and the use of research in the various forms of data and evidence available to teachers (cf., Wiesner & Schreiner, 2019) as crucial for quality improvement and professionalization in educational practice. Accordingly, there is some evidence that if educators engage with evidence to make or change decisions, embark on new courses of action, or develop new practices, this can have a positive impact on both teaching and learning (Bach et al., 2014; Cain, 2015; Richter et al., 2014; van Geel et al., 2016). However, there is evidence that teachers still struggle to transform data from performance tests, and also from classroom records, classroom assessments, program descriptions, and school statistics, etc. into useful knowledge (Groß Ophoff & Cramer, in press; Hamilton & Reeves, 2021; Schildkamp & Lai, 2013). Instead, teachers appear to rely on intuition, which is prone to bias and mistakes (Dunn et al., 2019; Fullan, 2005). Even attempts to develop the capacity of school leaders and practitioners to engage in reflective problem solving, such as Research Learning Networks or Data Teams (Brown et al., 2017; Mintrop & Zumpe, 2019), seem to fail in their attempts to facilitate deep research engagement. Against this backdrop, this paper addresses the issue of the assessment of the necessary and crucial competencies that should enable teachers and educational practitioners in general to engage and use research deliberately and (more) systematically.

Theoretical Background

In the field of educational assessment, the widely called-for research-related competencies include Educational Research Literacy (ERL, Groß Ophoff, Schladitz, et al., 2017; cf., Shank & Brown, 2007). This is conceptually related to assessment literacy (referring to the selection and use of student assessments, cf., DeLuca et al., 2016), data literacy (referring to drawing instructional conclusions from statistical information, cf., Mandinach, 2012; van Geel et al., 2017a), and statistical literacy (SL, referring to organizing/working with different data representations and understanding statistical concepts, cf., Ben-Zvi & Garfield, 2004; Watson & Callingham, 2003). In order to capture the research cycle as a whole, the concept of ERL also incorporates concepts from adjacent research fields, including information literacy (IL, referring to formulating research questions and information searches, e.g., Blixrud, 2003) and evidence-based reasoning (ER, referring to interpreting and critically evaluating evidence, e.g., Kuhn et al., 2008; Halonen, 2008).

However, despite the global movement toward accountability, evaluation, and assessment in education (DeLuca & Johnson, 2017) and the (theoretically assumed) importance of educational practitioners' proficiency in the engagement with research (see above), the state of knowledge about actual competency levels is unsatisfactory, and not only in Germany and Austria. However, there is some evidence that German in-service teachers are less proficient in ERL than pre-service teachers, even though the required abilities can be imparted or fostered during initial training or through professional development (e.g., Kittel et al., 2017). Accordingly, university (teacher) education is viewed as central because it allows connecting research and teaching (Healey, 2005). Because of the reorganization and change processes associated with the Bologna Reform (e.g., German Federal Ministry of Education and Research, 2015), a theoretical and empirical foundation for developing and implementing sustainable and psychometrically sound measures for quality assurance and development is regarded as crucial (Blömeke & Zlatkin-Troitschanskaja, 2013). In particular, psychometrically sound test instruments are expected to support the criterion-referenced interpretation with regard to the aspired competencies–which in turn can stimulate curriculum development and facilitate feedback about learning goals and gains (Wilson & Scalise, 2006) in initial education and as part of continuing teacher education.

Research literacy is usually assessed based on self-reports (e.g., Braun et al., 2008; Ntuli & Kyei-Blankson, 2016), but in general, correlations between subjective and objective competency measures are rather low (Lowman & Williams, 1987). Empirical approaches via assigned test instruments can be found, but have been scarce and psychometrically weak (e.g., Reeves & Honig, 2015). Recently developed test instruments focus either on particular steps of the research cycle (e.g., ER: Münchow et al., 2019; SL: Zeuch et al., 2017), or were developed in the context of specific interventions (e.g., Ebbeler et al., 2017; van Geel et al., 2016). Regarding the investigation of the psychometric structure, more often than not, one-dimensional models are applied without further comparison to other theoretically plausible multidimensional models (e.g., Watson & Callingham, 2003; van Geel et al., 2016). As one approach to investigating construct validity (Cronbach & Meehl, 1955), Groß Ophoff, Schladitz, et al. (2017) compared theoretically plausible one- and multi-dimensional models of a test instrument for the (more comprehensive) assessment of ERL based on a sample of 1360 students at six German universities. Even though it could be demonstrated that ERL consists of one generic factor of ERL and three secondary factors representing specific aspects in relation to the requirements of the research cycle (IL, SL, ER), the authors recommended applying a one-dimensional model because–due to the dominance of the general factor–essential unidimensionality can be assumed (Stout, 1987). Additionally, there is some evidence that even though social sciences share a certain methodological repertoire (Dietrich et al., 2015), the different research traditions represented in study programs involved in teacher education (e.g., sociology, educational science, psychology) appear to have a differential impact on performance in comprehensive assessments of research competency (Gess et al., 2017). In line with conceptual frameworks of data-based decision making (see above), the acquisition of competencies is shaped by the research-related opportunities-to-learn during initial and continuing teacher education (based on the institution- and discipline-specific curriculum) and also by its national and cultural contexts (Larcher & Oelkers, 2004). This perspective has been adopted in the current contribution.

For example, the educational systems in Germany and Austria (and Switzerland, for that matter) share cultural and linguistic commonalities (Gonon, 2011). In recent history, both countries were faced with an empirical shift in their education systems after disappointing results in international large-scale assessments became public in the early 2000s (Altrichter et al., 2005; Bos et al., 2010). In the aftermath, as early as 2004, research literacy was explicitly identified as a requirement in the so-called Standards for Teacher Education by the German Standing Conference of the Ministers of Education and Cultural Affairs. Accordingly, teachers in training should be able to consider and evaluate evidence from educational research and practicing teachers should be able to use evidence-based insights for instructional and school development. In Austria, reforms came to fruition later, especially as stakeholders in education policy did not present a united front (Olano, 2010). As late as 2010, the Austrian Federal Ministry of Science and Research and the Federal Ministry for Education, Arts and Culture (2010) published an expert view on the future of pedagogical professions in which the recommendation was expressed that science and research need to be established as constitutive elements of teacher education. Further reforms in teacher education followed in later years, like the legal adoption of a reform in 2013 (Hofmann et al., 2020) asserting that all teachers in training must obtain an academic degree (bachelor's or master's degree), and renouncing the previously parallel organization of teacher education in universities of education (UE, German: Pädagogische Hochschulen; with a focus on primary and lower secondary education) and universities (with a focus on higher secondary education) in favor of founding development networks or clusters in which universities and UE collaborate.

Research Questions

Against this backdrop, this paper addresses the question of whether a test instrument developed in the specific context of German study programs in (teacher) education can be applied to other national contexts, in this case to Austrian teacher education. This approach to investigating construct validity under consideration of the psychometric structure and group differences on item level (Cronbach & Meehl, 1955) is a necessary step in ensuring the fairness of cross-national comparisons (Davidov et al., 2014; Förster et al., 2015).

For this purpose, results about the dimensionality of ERL in the large-scale German study (e.g., Groß Ophoff, Wolf, et al., 2017) are compared to a study at two Austrian UE (Haberfellner, 2016). In both studies, the same ERL test instrument was used. According to Prenzel et al. (2007), probabilistic test theory, which is the basis for the reported analyses in this paper, makes it possible to validate theoretically plausible assumptions about the dimensional structure of a construct (e.g., by comparing competing models). Accordingly, for the data from the German sample (Study 1), a bifactor model (Model 3, see 4.3) with one dominant general factor and the three secondary factors (IL, SL, ER) turned out to be the best fit (Groß Ophoff, Wolf, et al., 2017). This model served as acceptable compromise between the one-dimensional model (Model 1) and the three-dimensional model (Model 2 with the subdimensions IL, SL, ER) that were applied in preliminary analyses (e.g., Haberfellner, 2016; Schladitz et al., 2015). These findings serve as a reference for the analysis of the Austrian sample in this paper. As the invariance of the measurement instrument is crucial for the valid comparison of samples from different countries (e.g., Davidov et al., 2014), the following question will be pursued:

1. Can the psychometric structure of the ERL test instrument for the sample of German students in Study 1 also be applied to the sample of Austrian students in Study 2 (Model 3), and can configural invariance therefore be assumed? If not, which of the two other theoretically plausible models (Model 1, Model 2) fits better?
2. Are different probabilities for a correct response in single items identifiable (so-called differential item functioning; DIF)? If so, is one of the samples consistently disadvantaged (uniform DIF) or does this vary across the item sample (non-uniform DIF)?

Methods

Analyses were conducted utilizing data sets from two studies: the first from the large-scale main study in Germany (Study 1: winter semester 2012/2013 and summer semester 2013), and the second from a study at two Austrian UE (German: Pädagogische Hochschulen = UE) in the summer semester of 2015 (Study 2). In both studies, participants were recruited upon request in lectures (convenience samples). Participation was voluntary and anonymous.

Data collection and samples

In Study 1 (see Table 1), 1360 students in the field of educational science at six German higher education institutions from five federal states were investigated between 2012 and 2013. Because of the German federal constitution, the federal states are predominantly responsible for education, science, and culture, but cross-nationally coordinate and collaborate in education and training (to some extent) through the Standing Conference established in 1948 (Standing Conference, 2019, 2020). The sample includes one UE in Baden-Württemberg, and one university (reformed former UE) in Rhineland-Palatinate that both are rather small universities with a strong focus on educational science and related disciplines as well as on subject-related didactics. In other federal states, teacher education institutions were integrated into the educational science departments of state universities by the 1970s (Meissner et al., 2012). This is the case for the other four large universities in this sample that offer a wide range of study programs and are characterized by a strong research orientation. In teacher training, these comprehensive universities typically tend to focus on subject-related studies. In this study, teacher training students (for all school forms) represented the largest group, followed by educational studies students (23%), and other study programs (e.g., early education, health education, educational psychology).

Table 1
Descriptive statistics of the samples from Study 1 (five German states, winter semester 2012/2013 and summer semester 2013) and Study 2 (two development networks: summer semester 2015)

	Germany (*Study 1*): winter semester 2012/2013 and summer semester 2013	Austria (*Study 2*): summer semester 2015
N	1360 students	295
Age, *M* (*SD*)	22.9 years (3.95)	22.9 years (4.32)
Gender (% female)	75.9%	77.6%
Teacher Training students	62%	100%

Note. Abbreviations: *N* = number of study participants.

Study 2 investigated 295 teacher training students (primary and lower secondary education) from two Austrian UE from the cluster "West" (Tirol, Vorarlberg) and "Mitte" (Salzburg, Upper Austria). At the time of the study, teacher education for primary and lower secondary schools was located at UE that were established as late as 2007 from post-secondary schools (German: Pädagogische Akademien). To this day, Austrian UE are not authorized to award doctoral and postdoctoral degrees, and the link between research and teaching is by no means a given for teaching staff (Haberfellner, 2016; Hofmann et al., 2020).

Test instrument and booklet design

The main focus of the research program in Study 1 was on the development of a test instrument for the assessment of ERL (see Table 2) covering the steps of the research process, such as search strategies for problem-specific research information, the comprehension of different types of academic documents, the formulation of adequate research questions (IL), the analysis and interpretation of descriptive statistics (SL), and the critical evaluation of research-based assumptions (ER; cf., Groß Ophoff, Wolf, et al., 2017). The resulting item pool was reviewed by content experts, pre-tested comprehensively, and subsequently deployed with the goal of test standardization between 2012 and 2013 (see Table 1). During implementation, 40 minutes were allotted by the test administrators to complete the ERL test. In the remaining 20 minutes, participants were asked to provide personal and professional background information, and further characteristics were surveyed. During data analysis in Study 1, poor fitting items ($0.80 \geq$ Infit/Outfit ≥ 1.20, cf., Adams & Wu, 2002) and items with low discrimination ($r < 0.20$) were excluded. The foundation of the results reported here is the reduced item pool of 193 items (119 stems, see Table 2).

Table 2

Distribution of test items to the competence facets information literacy, statistical literacy, and evidence-based reasoning in the standardization study (Germany) and the study in summer semester 2015 (Austria)

	Germany (DE: Study 1): winter semester 2012/2013 and summer semester 2013	Austria (AT: Study 2): summer semester 2015
Competence facets		
IL	30 (15.5%)	8 (20.0%)
SL	71 (36.8%)	14 (35.0%)
ER	92 (47.4%)	18 (45.0%)

Note. Abbreviations: IL = information literacy; SL = statistical literacy; ER = evidence-based reasoning; n_i = number of test items.

In contrast to Study 1, the focus of Study 2 was on investigating the effect of the subjective value of research on pre-service teachers' research-oriented stance and on their level of ERL (Haberfellner, 2016). The 40 test items (referring to 17 stems, see Table 2) were selected from the item pool from Study 1 and then arranged in a single test booklet set up for a processing time of 40 minutes. Again, personal and professional background information were collected, and research-related attitudes were assessed. The test booklet for Study 2 had to be compiled before the data analysis in Study 1 was concluded. Therefore, no standardized parameter estimates were available for six of the selected items because they were excluded from analysis in Study 1 (see above). In Study 2, all items showed good item fit and were retained in the separate investigation of the dimensional structure of ERL reported here. These items could not be used to investigate differential item functioning (DIF, see 4.3), testing for (partial) measurement equivalence. The same applies to eight other items that were slightly modified for Study 2. Therefore, the in-depth analysis of the item-by-country interaction was based on 26 items.

Statistical analysis

Psychometric models popular in the field of competency assessment are based on item response theory (IRT), which rests upon stringent statistical assumptions (i.e., monotonicity, local independence, and unidimensionality). Multidimensional IRT models (Hartig & Höhler, 2009) assume that several latent dimensions are represented by item clusters. But it has been questioned whether the assumption of strict unidimensionality is applicable to, for example, educational and psychological assessment where, in addition to one dominant latent trait, other minor latent factors likely influence participant responses (e.g., Gustafson, 2001). Bifactor models are a solution to this, as they allow each item response to be explained by both a dominant factor in the sense of a common latent trait (e.g., ERL), and additional, orthogonal (therefore uncorrelated) factors caused by "parcels" of items drawing from similar aspects of the underlying traits (Reise et al., 2010).

As mentioned above, valid comparisons between groups–like the samples from Study 1 and Study 2–require cross-national invariance of the measurement instrument (Tay et al., 2015). The identification of a comparable dimensional structure for the ERL instrument ("configural invariance") was the first step in warranting comparability. Therefore, in each of the two samples (Study 1, Study 2), three competing models (see 3) were compared with the R package Test Analysis Modules (TAM, Kiefer et al., 2016). The best fitting model was identified separately for each sample based on the lowest values in the information criteria AIC, BIC, and CAIC (Schermelleh-Engel et al., 2003). The precision of person estimates was reported by the EAP/PV (expected a posteriori/ plausible value) reliability coefficient, which represents the explained variance in the estimated model divided by total person variance (Bond & Fox, 2007). This coefficient is comparable with Cronbach's α, for which values of at least 0.55 are deemed satisfactory for group comparisons (Rost, 2013). For multidimensional constructs like the bifactor model, Green and Yang (2009) recommended reporting Omega (ω) as a model-based reliability estimate that combines higher-order and lower-order factors, and Omega-hierarchical (ωh) as model-based reliability estimate of one target construct with others removed.

To gain further insights into measurement equivalence (or the lack thereof) of single items, DIF was investigated. To this end, group specific item parameters were compared based on the deviation of the group mean from the overall mean in relation to the standard error (Critical Ratio, cf., Holland & Wainer, 1993). Accordingly, values for a certain item below $z = -1.96$ or above $z = 1.96$ indicate meaningful DIF (Wu et al., 2007). In this case, respondents with the same proficiency level, but from different countries, showed different probabilities for a correct response in an item (Wirtz & Böcker, 2017). However, emerging DIF should be interpreted with caution here because smaller samples lead to higher standard errors, thus more frequently to significant results. This is particularly the case in Study 1, where single items were usually assigned to approximately 200 students due to the applied incomplete block design (Groß Ophoff, Wolf, et al., 2017).

Results

At first glance, the samples from Study 1 and Study 2 appear to demonstrate a different dimensionality of ERL (see Table 3). In Study 1, the bifactor model solution in Model 3 shows better fit than the one- or the three-dimensional models because the corresponding values of AIC, BIC, and CAIC were lowest. The information criteria values of the one-dimensional and the bifactor model were closer to each other than to the three-dimensional model. This is the same in Study 2, even though only the AIC indicates the four-dimensional model as better-fitting, whereas the BIC- and CAIC-values favored the more parsimonious one-dimensional model of ERL. Overall, the model results from both samples indicate that the three secondary factors of IL, SL, and ER can be distinguished from a general factor of ERL. Although more pronounced in the Study 1 sample, the general factor in Model 3 was dominant in both samples (DE: $\omega h = 0.85$; AT: $\omega h = 0.65$). Accordingly, it is reasonable to apply a one-dimensional model without further differentiation of the three competence facets (Groß Ophoff, Wolf, et al., 2017b). For the one-dimensional model, the reliability of the test instrument was found to be satisfactory for both the German (EAP-reliability = 0.61, cf., Böttcher-Oschmann et al., 2019) and the Austrian sample (EAP-reliability = 0.59).

Table 3
Goodness-of-fit statistics for competing models in Study 1 and Study 2

Sample	n_i	Model	Factors	Final Deviance	n_p	AIC	BIC	CAIC
Study 1: DE	193	1	1 (G)	43,049.0	194	43,437	44,449	44,643
		2	3 (IL, SL, ER)	43,052.4	199	43,450	44,488	44,687
		3	4 (G, IL, SL, ER)	43,020.1	197	**43,414**	**44,442**	**44,639**
Study 2: AT	40	1	1 (G)	10,760.7	41	10,843	**10,994**	**11,035**
		2	3 (IL, SL, ER)	10,742.7	46	10,835	11,004	11,050
		3	4 (G, IL, SL, ER)	10,744.5	44	**10,832**	10,995	11,039

Note. Study 1 (Germany): winter semester 2012/2013 & summer semester 2013. *Study 2* (Austria): summer semester 2015. Sample size: *N (Study 1)* = 1360; *N (Study 2)* = 295. abbreviations: n_i = number of test items included; n_p = *number of estimated parameters; G = general factor Educational Research Literacy; IL* = information literacy; SL = statistical literacy; ER = evidence-based reasoning; The parameters of the respective best fitting solution are indicated in bold.

On closer inspection of the 26 test items included in the DIF analysis (see 4.3), 12 items showed no meaningful DIF between the two samples. In Table 4, the results for the remaining 14 items are reported. Critical Ratio values (last column) below z = −1.96 indicate that the teacher training students in Study 1 showed a higher probability for a correct response than those in Study 2 (upper half of Table 4), which is the case for six items; conversely, values above z = 1.96 indicate an advantage for participants in Study 2 (eight items, see lower half of Table 4). In the third and fourth column from left, task content and the required competencies are briefly stated. It should be stressed that for item 6.1, the slight advantage for the Austrian sample might be explained by the compilation of the test booklet. In Study 2, this item was preceded by item 5, referring to the same graph; in Study 1, these two items were located in different test booklets. The obvious assumption is that the close reading required for the solution of item 5 lead to a slight advantage in this item. But overall, a mixed picture emerges.

Table 4

Overview of items with Differential Item Functioning in Study 1 and Study 2

In favor of …	Item position (Study 2)	Task content	Required competencies	M (Study 1)	SE (Study 1)	Critical Ratio
Study 1	10	Study abstract	Identification of adequate follow-up research question	−0.643	0.082	−7.84
	11	Literature search	Identification of suitable search terms	−0.246	0.076	−3.24
	16.4	Comparison of two study abstracts	Evaluation of study designs	−0.312	0.079	−3.95
	16.5		Identification of study with control group	−0.331	0.079	−4.19
	21.3	Bar chart (degree aspiration of male vs. female students)	Recognition of inadmissible conclusion	−0.513	0.097	−5.29
	21.2		Calculation of percentages for appraising a statement	−0.273	0.081	−3.37
Study 2	4.1	Description of different research procedures	Assessment of suitability for research objective	0.592	0.081	7.31
	4.2			0.408	0.084	4.86
	4.3			0.201	0.084	2.39
	6.1	Integrated bar chart	Graph interpretation	0.163	0.075	2.17
	8	Venn diagram	Interpretation of intersections	0.352	0.099	3.56
	16.2	see 16.4 (above)	Appraisal of conclusions	0.191	0.078	2.45
	16.6			0.228	0.082	2.78
	19	bibliographical reference	Identification of source	0.423	0.078	5.42

Note. Study 1 (Germany): winter semester 2012/2013 & summer semester 2013. *Study 2* (Austria): summer semester 2015. Abbreviations: IL = information literacy; SL = statistical literacy; ER = evidence-based reasoning. All items reported show significant DIF (Critical Ratio below z = −1.96 or above z = 1.96).

Conclusions

Even though these reported results appear somewhat inconclusive with a view to the dimensionality of ERL, they perpetuate the previously described structural ambiguity of the test instrument (Groß Ophoff, Wolf, et al., 2017). Because of the dominance of the general factor in Model 3 in both samples (DE: $\omega_h = 0.85$; AT: $\omega_h = 0.65$), the recommendation to use a one-dimensional model of ERL (Model 1) for the assessment and feedback of proficiency on the individual level (research question 1, see 3) could be substantiated.

Overall, the results indicate that the presented ERL test can be validly applied to assess the research literacy of teacher training students in both countries, even though it is worthwhile to take Differential Item Functioning into account. The DIF analysis of the two samples further revealed that at least partial equivalence can be assumed (research question 2), even though the issue of whether the identified violations are problematic for meaningful comparisons is still controversial (Davidov et al., 2014). Nevertheless, the identification of non-uniform DIF indicates that neither of the two samples was consistently disadvantaged. Given that both studies are based on convenience samples (which is a common challenge for research in higher education, cf., Zlatkin-Troitschanskaia et al., 2016), items with DIF might–interpreted with due caution–hint at some advantage in research-methodological issues for the sample in Study 1, and in appraising research-based conclusions for Study 2. But it should be remembered that even though more items showed an unexpected higher probability for a correct response for the students in Study 2, they showed an overall lower proficiency in ERL ($b_{Study2} = -.22$; 95%-CI: −.30, −.14) than teacher training students in Study 1 (NLA = 841). However, particularly in Study 1, the ratio of persons per item was comparatively small due to the incomplete block design. To gain a better understanding of the reasons for the identified DIF (benign vs. adverse DIF, cf., Gierl, 2005), larger, specifically selected item samples need to be investigated based on larger samples, and curricular content experts should be involved.

It should be noted, too, that it was not necessary to translate the test items in the two studies here. The transfer to other cultural contexts and the necessary translation to ensure linguistic and cultural equivalence will probably present greater challenges (e.g., Grisay et al., 2007). For example, there are currently translations for a selection of ERL test items either available (English, Arabic) or in the making (Italian, Spanish). But due to the use of the translated ERL tests for course or curricular evaluations in specific higher education institutions and the resulting small samples, the translated versions have not yet been analyzed with regard to measurement equivalence.

In-depth analysis on a meso-level revealed differences between the institutions included in the two studies here. For example, the more proficient teacher training students in Study 1 were located at large German universities with a traditionally strong research orientation, whereas students with the lowest proficiency came from a university that did not explicitly identify ERL as a study objective in the curriculum at that time (Groß Ophoff, Schladitz, et al., 2017). In Study 2, both institutions offered only introductory research-based courses (scientific working methods, applied research, and evaluation), which is probably why no significant differences in ERL between the two emerged. Presumably, these differences are related to the embedding and amount of research in teacher education study programs. While knowledge about the extent of research-orientation in Austrian education is still rather limited (Jesacher-Roessler & Kemethofer, in press), more is known about the current research-related practices in German teacher education (Groß Ophoff & Cramer, in press): Evidently mainly research-led (focused on engagement with research data) or—to a lesser extent—research-oriented courses (focused on imparting research methods, e.g., Rueß et al., 2016; Stelter & Miethe, 2019) are available. Inquiry-based courses in which students are scaffolded to absolve certain phases or full research projects have been established in recent years, particularly as part of long-term school internships (e.g., Ulrich & Gröschner, 2020). But findings about the intended (research-related) effects have thus far been rather sobering (e.g., van Ophuysen et al., 2017). Based on the reconstructive analysis of inquiry-based course concepts in teacher education, Katenbrink and Goldmann (2020) pointed out that rather superficial "one fits all"-concepts appear to dominate in German initial teacher education, in which practical procedures are trained and inquiry is loosely imparted as the evaluation of educational practices.

Further limitations of the study presented here are that the ERL test strongly (but not exclusively) focuses on quantitative-methodological topics. Furthermore, the presented ERL test operationalizes only a subsidiary, that is cognitive, aspect of research competence. In recent years, there has been an increased awareness that affective-motivational factors also play an important role for the depth of engagement with research information (Wessels et al., 2018), which highlights the importance of further research on meso- (courses in teacher training) and micro-level (competency development of pre- and in-service teachers). This might shed light on the much-needed advanced understanding of how to support or facilitate competent research engagement in teacher education (cf., Brown et al., 2021). According to Katenbrink and Goldmann (2020), inquiry-based learning in particular, with a focus on the assumption of the fundamental and unresolvable difference between theory and practice (concepts of difference), has the potential to empower teachers to reflect on their own educational practice with professional distance (Cramer et al., 2019; Helsper, 2016) and to use research information as an opportunity for deep learning or even conceptual change (Gregoire, 2003), and also to convince them about the usefulness of research for quality development in education (Prenger & Schildkamp, 2018).

References

Adams, R. J., & Wu, M. (2002). *PISA 2000 Technical Report.* Organisation for Economic Cooperation and Development. https://www.oecd.org/pisa/data/33688233.pdf

Altrichter, H., Brüsemeister, T., & Heinrich, M. (2005). Merkmale und Fragen einer Governance-Reform am Beispiel des österreichischen Schulwesens. *Österreichische Zeitschrift für Soziologie*, *30*(4), 6–28. https://doi.org/10.1007/s11614-006-0063-0

Bach, A., Wurster, S., Thillmann, K., Pant, H. A., & Thiel, F. (2014). Vergleichsarbeiten und schulische Personalentwicklung—Ausmaß und Voraussetzungen der Datennutzung. *Zeitschrift für Erziehungswissenschaft*, *17*(1), 61–84. https://doi.org/10.1007/s11618-014-0486-5

Ben-Zvi, D., & Garfield, J. (2004). The challenge of developing statistical literacy, reasoning and thinking (1st ed.). Springer. https://doi.org/10.1007/1-4020-2278-6

Blixrud, J. C. (2003). Project SAILS: Standardized assessment of information literacy skills. In G. J. Barret (Ed.), *ARL: A bimonthly report on research library issues and actions from ARL, CNI, and SPARC. Special double issue on new measures. Number 230-231, October-December 2003* (pp. 18–19). Association of Research Libraries. https://files.eric.ed.gov/fulltext/ED498293.pdf

Blömeke, S., & Zlatkin-Troitschanskaia, O. (2013). Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs (KoKoHs Working Papers No. 1). Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor. https://www.kompetenzen-im-hochschulsektor.de/files/2018/05/KoKoHs_WP1_Bloemeke_Zlatkin-Troitschanskaia_2013_.pdf

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.

Bos, W., Postlethwaite, T. N., & Gebauer, M. M. (2010). Potenziale, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (3rd ed., pp. 275–295). Springer.

Böttcher-Oschmann, F., Groß Ophoff, J., & Thiel, F. (2019). Validierung eines Fragenbogens zur Erfassung studentischer Forschungskompetenzen über Selbsteinschätzungen – Ein Instrument zur Evaluation forschungsorientierter Lehr-Lernarrangements. *Unterrichtswissenschaft*, *47*(4), 495–521. https://doi.org/10.1007/s42010-019-00053-8

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, *54*(1), 30–42. https://doi.org/10.1026/0012-1924.54.1.30

Brown, C., Poortman, C., Gray, H., Groß Ophoff, J., & Wharf, M. (2021). Facilitating collaborative reflective inquiry amongst teachers: What do we currently know? *International Journal of Educational Research*, *105*, 101695. https://doi.org/10.1016/j.ijer.2020.101695

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational Research*, *59*(2), 154–172. https://doi.org/10.1080/00131881.2017.1304327

Bundesministerium für Unterricht, Kunst und Kultur, & Bundesministerium für Wissenschaft und Forschung [Federal Ministry for Education, the Arts and Culture & Federal Ministry of Science and Research]. (2010). *LehrerInnenbildung NEU – Die Zukunft der pädagogischen Berufe: Die Empfehlungen der ExpertInnengruppe* [Teacher Education NEW – The future of pedagogical professions]. https://www.qsr.or.at/dokumente/1870-20140529-092820-Empfehlungen_der_ExpertInnengruppe_Endbericht_092010_2_Auflage.pdf

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

Cramer, C., Harant, M., Merk, S., Drahmann, M., & Emmerich, M. (2019). Meta-Reflexivität und Professionalität im Lehrerinnen- und Lehrerberuf. *Zeitschrift für Pädagogik*, *65*(3), 401–423.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40*, 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 121–126. https://doi.org/10.1080/0969594X.2017.1297010

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

de Olano, D. (2010). Gewinner, Verlieren und Exoten - PISA in sieben weiteren Staaten. In P. Knodel, K. Martens, D. de Olano, & M. Popp (Eds.), *Das PISA-Echo. Internationale Reaktionen auf die Bildungsstudie* (pp. 251–300). Campus.

Dietrich, H., Zhang, Y., Klopp, E., Brünken, R., Krause, U.-M., Spinath, F. M., Stark, R., & Spinath, B. (2015). Scientific competencies in the social sciences. *Psychology Learning & Teaching*, *14*(2), 115–130. https://doi.org/10.1177/1475725715592287

Dunn, K. E., Skutnik, A., Patti, C., & Sohn, B. (2019). Disdain to acceptance: Future teachers' conceptual change related to data-driven decision making. *Action in Teacher Education*, *41*(3), 193–211. https://doi.org/10.1080/01626620.2019.1582116

Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2017). The effects of a data use intervention on educators' satisfaction and data literacy. *Educational Assessment, Evaluation and Accountability*, *29*(1), 83–105. https://doi.org/10.1007/s11092-016-9251-z

Federal Ministry of Education and Research Germany. (2015). *Bericht der Bundesregierung über die Umsetzung des Bologna-Prozesses 2012–2015 in Deutschland*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_02_12-NationalerBericht_Umsetzung_BolognaProzess.pdf

Förster, M., Zlatkin-Troitschanskaia, O., Brückner, S., Happ, R., Hambleton, R. K., Walstad, W. B., Asano, T., & Yamaoka, M. (2015). Validating test score interpretations by cross-national comparison: Comparing the results of students from Japan and Germany on an American test of economic knowledge in higher education. *Zeitschrift Für Psychologie*, *223*(1), 14–23. https://doi.org/10.1027/2151-2604/a000195

Fullan, M. (2005). The meaning of educational change: A quarter of a century of learning. In A. Lieberman (Ed.), *The roots of educational change: International handbook of educational change* (1st ed., pp. 202–216). Springer. https://doi.org/10.1007/1-4020-4451-8_12

Gess, C., Wessels, I., & Blömeke, S. (2017). Domain-specificity of research competencies in the social sciences: Evidence from differential item functioning. *Journal for Educational Research Online*, *9*(2), 11–36. https://doi.org/10.25656/01:14895

Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice*, *24*(1), 3–14. https://doi.org/10.1111/j.1745-3992.2005.00002.x

Gonon, P. (2011). Die Bedeutung des internationalen Arguments in der Lehrerbildung. *Beiträge zur Lehrerbildung, 29*(1), 20–26. https://doi.org/10.25656/01:13763

Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. https://doi.org/10.1007/s11336-008-9099-3

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*(2), 147–179. https://doi.org/10.1023/A:1023477131081

Grisay, A., de Jong, J. H. A. L., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, *8*(3), 249–266.

Groß Ophoff, J., & Cramer, C. (in press). The engagement of teachers and school leaders with data, evidence and research in Germany. In C. Brown & J. R. Malin (Eds.), *The Emerald international handbook of evidence-informed practice in education: Learning from international contexts*. Emerald.

Groß Ophoff, J., Schladitz, S., & Wirtz, M. (2017). Differences in research literacy in educational science depending on study program and university. In J. Domenech i de Soria, M. C. Vincent-Vela, E. de la Poza, & D. Blazquez (Eds.), *Proceedings of the HEAd'17. 3rd international conference on higher education advances* (pp. 1193–1202). Congress UPV. http://dx.doi.org/10.4995/HEAD17.2017.6713

Groß Ophoff, J., Wolf, R., Schladitz, S., & Wirtz, M. (2017). Assessment of educational research literacy in higher education: Construct validation of the factorial structure of an assessment instrument comparing different treatments of omitted responses. *Journal for Educational Research Online*, *9*(2), 37–68. https://doi.org/10.25656/01:14896

Gustafson, J.-E. (2001). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (1st ed., pp. 87–111). Routledge.

Haberfellner, C. (2016). Der Nutzen von Forschungskompetenz im Lehramt: Eine Einschätzung aus der Sicht von Studierenden der Pädagogischen Hochschulen in Österreich. Klinkhardt.

Halonen, J. S. (2008). Measure for measure: The challenge of assessing critical thinking. In D. S. Dunn, J. S. Halonen, & R. A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (1st ed., pp. 61–76). Wiley-Blackwell.

Hamilton, V. M., & Reeves, T. D. (2021). Relationships between course taking and teacher self-efficacy and anxiety for data-driven decision making. *The Teacher Educator*, 1–19. https://doi.org/10.1080/08878730.2021.1965682

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, *35*(2-3), 57–63. https://doi.org/10.1016/j.stueduc.2009.10.002

Healey, M. (2005). Linking research and teaching: Exploring disciplinary spaces and the role of inquiry-based learning. In R. Barnett (Ed.), *Reshaping the University: New Relationships between Research, Scholarship and Teaching* (1st ed., pp. 67–78). Open University Press.

Helsper, W. (2016). Lehrerprofessionalität – der strukturtheoretische Ansatz. In M. Rothland (Ed.), *Beruf Lehrer/Lehrerin: Ein Studienbuch* (pp. 103–125). Waxmann.

Hofmann, F., Hagenauer, G., & Martinek, D. (2020). Entwicklung und Struktur der Lehrerinnen- und Lehrerbildung in Österreich. In C. Cramer, J. König, M. Rothland, & S. Blömeke (Eds.), *Handbuch Lehrerinnen- und Lehrerbildung* (pp. 227–236). Klinkhardt.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning* (1st ed.). Erlbaum.

Jesacher-Roessler, L., & Kemethofer, D. (in press). Evidence-informed practice in Austrian education. In C. Brown & J. R. Malin (Eds.), *The Emerald international handbook of evidence-informed practice in education*. Emerald.

Katenbrink, N., & Goldmann, D. (2020). Varianten Forschenden Lernens—Ein konzeptbasierter Typisierungsvorschlag. In M. Basten, C. Mertens, A. Schöning, & E. Wolf (Eds.), *Forschendes Lernen in der Lehrer/innenbildung: Implikationen für Wissenschaft und Praxis* (pp. 195–202). Waxmann.

Kiefer, T., Robitzsch, A., & Wu, M. (2013). *TAM (Test analysis modules)*. http://www.edmeasurementsurveys.com/TAM/Tutorials/

Kittel, D., Rollett, W., & Groß Ophoff, J. (2017). Profitieren berufstätige Lehrkräfte durch ein berufsbegleitendes weiterbildendes Studium in ihren Forschungskompetenzen? *Bildung & Erziehung*, *70*(4), 437–452.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, *23*(4), 435–451. https://doi.org/10.1016/j.cogdev.2008.09.006

Larcher, S., & Oelkers, J. (2004). Deutsche Lehrerbildung im internationalen Vergleich. In S. Blömke, P. Reinhold, G. Tulodziecki, & J. Wildt (Eds.), *Handbuch Lehrerbildung* (pp. 128–150). Klinkhardt.

Lowman, R. L., & Williams, R. E. (1987). Validity of self-ratings of abilities and competencies. *Journal of Vocational Behavior, 31*(1), 1–13. https://doi.org/10.1016/0001-8791(87)90030-3

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In M. Honey (Ed.), *Data-Driven School Improvement: Linking Data and Learning*, (pp. 13–31). Teachers College Press.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48.

Meissner, D., Vogel, E., & Horn, K.-P. (2012). Lehrerausbildung in Baden-Württemberg seit 1945. Angleichungs-und Abgrenzungsprozesse. In C. Cramer (Ed.), *Lehrerausbildung in Baden-Württemberg. Historische Entwicklungslinien und aktuelle Herausforderungen,* (pp. 33–62). IKS Garamond.

Mintrop, R., & Zumpe, E. (2019). Solving real-life problems of practice and education leaders' school improvement mind-set. *American Journal of Education*, *125*(3), 295–344. https://doi.org/10.1086/702733

Münchow, H., Richter, T., von der Mühlen, S., & Schmid, S. (2019). The ability to evaluate arguments in scientific texts: Measurement, cognitive processes, nomological network, and relevance for academic success at the university. *The British Journal of Educational Psychology*, *89*(3), 501–523. https://doi.org/10.1111/bjep.12298

Ntuli, E., & Kyei-Blankson, L. (2016). Improving K-12 online learning: Information literacy skills for teacher candidates. *International Journal of Information and Communication Technology Education*, *12*(3), 38–50.

Prenger, R., & Schildkamp, K. (2018). Data-based decision making for teacher and student learning: A psychological perspective on the role of the teacher. *Educational Psychology*, *38*(6), 734–752. https://doi.org/10.1080/01443410.2018.1426834

Prenzel, M., Walter, O., & Frey, A. (2007). PISA misst Kompetenzen. *Psychologische Rundschau*, *58*(2), 128–136. https://doi.org/10.1026/0033-3042.58.2.128

Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, *50*, 90–101. https://doi.org/10.1016/j.tate.2015.05.007

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Richter, D., Böhme, K., Becker, M., Pant, H. A., & Stanat, P. (2014). Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. *Zeitschrift für Pädagogik*, *60*(2), 225–244. https://doi.org/10.25656/01:12846

Rost, D. H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien* (3. completely revised ed.) [Interpretation and evaluation of educational-psychological studies]. Klinkhardt.

Rueß, J., Gess, C., & Deicke, W. (2016). Forschendes Lernen und forschungsbezogene Lehre – empirisch gestützte Systematisierung des Forschungsbezugs hochschulischer Lehre. *Zeitschrift für Hochschulentwicklung, 11*(2), 23–44.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online*, *8*(2), 23–74.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482–496. https://doi.org/10.1016/j.tate.2009.06.007

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. Kuin & L. Earl (Eds.), *Data-based decision making in education* (pp. 177–191). Springer.

Schildkamp, K., & Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers College Record*, *117*(4), 1–42.

Schladitz, S., Groß Ophoff, J., & Wirtz, M. (2015). Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz. *Zeitschrift für Pädagogik*, *61*, 167–184. https://doi.org/10.25656/01:15509

Schratz, M., Wiesner, C., Rößler, L., Schildkamp, K., George, A. C., Hofbauer, C., & Pant, H. A. (2018). Möglichkeiten und Grenzen evidenzorientierter Schulentwicklung. *Nationaler Bildungsbericht Österreich*, *2*, 403–454. http://doi.org/10.17888/nbb2018-1.4

Shank, G., & Brown, L. (2007). *Exploring educational research literacy*. Routledge.

Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. (2019). The education system in the Federal Republic of Germany 2016/2017. A description of the responsibilities, structures and developments in education policy for the exchange of information in Europe. https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-engl-pdfs/dossier_en_ebook.pdf.

Standing Conference. (2020). Ländervereinbarung über die gemeinsame Grundstruktur des Schulwesens und die gesamtstaatliche Verantwortung der Länder in zentralen bildungspolitischen Fragen (Beschluss der Kultusministerkonferenz vom 15.10.2020). https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_10_15-Laendervereinbarung.pdf.

Standing Conference. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/ 2004_12_16-Standards-Lehrerbildung.pdf

Stelter, A., & Miethe, I. (2019). Forschungsmethoden im Lehramtsstudium – aktueller Stand und Konsequenzen. *Erziehungswissenschaft*, *30*(58), 25–33. https://doi.org/10.3224/ezw.v30i1.03

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589–617. https://doi.org/10.1007/BF02294821

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, *18*(1), 3–46. https://doi.org/10.1177/1094428114553062

Ulrich, I., & Gröschner, A. (2020). Praxissemester im Lehramtsstudium in Deutschland: Wirkungen auf Studierende. Springer VS.

van Geel, M., Keuning, T., Visscher, A. J., & Fox, J.-P. (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. *American Educational Research Journal*, *53*(2), 360–394. https://doi.org/10.3102/0002831216637346

van Geel, M., Keuning, T., Visscher, A., & Fox, J.-P. (2017). Changes in educators' data literacy during a data-based decision making intervention. *Teaching and Teacher Education*, *64*, 187–198. https://doi.org/10.1016/j.tate.2017.02.015

van Ophuysen, S., Behrmann, L., Bloh, B., Homt, M., & Schmidt, J. (2017). Die universitäre Vorbereitung angehender Lehrkräfte auf Forschendes Lernen im schulischen Berufsalltag. *Journal for Educational Research Online, 9*(2), 276–305. https://doi.org/10.25656/01:14952

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.

Wessels, I., Rueß, J., Jenßen, L., Gess, C., & Deicke, W. (2018). Beyond cognition: Experts' views on affective-motivational research dispositions in the social sciences. *Frontiers in Psychology, 9*, 1300. https://doi.org/10.3389/fpsyg.2018.01300

Wiesner, C., & Schreiner, C. (2019). Implementation, Transfer, Progression und Transformation: Vom Wandel von Routinen zur Entwicklung von Identität. Von Interventionen zu Innovationen, die bewegen. Bausteine für ein Modell zur Schulentwicklung durch Evidenz (en). *Praxistransfer Schul- und Unterrichtsentwicklung*, 79–140.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher Education*, *52*(4), 635–663. https://doi.org/10.1007/s10734-004-7263-y

Wirtz, M. A., & Böcker, M. (2017). Differential item functioning (DIF). In M. A. Wirtz (Ed.), *Dorsch – Lexikon der Psychologie* (18. Aufl.). Hogrefe.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACERConQuest Version 2: Generalised item response modelling software.* Australian Council for Educational Research.

Zeuch, N., Förster, N., & Souvignier, E. (2017). Assessing teachers' competencies to read and interpret graphs from learning progress assessment: Results from tests and interviews. *Learning Disabilities Research & Practice*, *32*(1), 61–70. https://doi.org/10.1111/ldrp.12126

Zlatkin-Troitschanskaia, O., Pant, H. A., Kuhn, C., Toepper, M., & Lautenbach, C. (2016). Messung akademisch vermittelter Kompetenzen von Studierenden und Hochschulabsolventen. Ein Überblick zum nationalen und internationalen Forschungsstand (Vol. 1). Springer.

Corresponding authors

Jana Groß Ophoff
Institute for Secondary Education, University College of Teacher Education Vorarlberg, Austria
E-mail: jana.grossophoff@ph-vorarlberg.ac.at

Christina Egger
Institute for Didactics, Teaching and School Development, University College of Teacher Education Stefan Zweig Salzburg, Austria
E-mail: christina.egger@phsalzburg.at

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-3

VALIDATION OF THE TEACHER DECISION-MAKING INVENTORY (TDMI): MEASURING DATA-BASED AND INTUITIVE DIMENSIONS IN TEACHERS' DECISION PROCESS

KRISTIN VANLOMMEL,
ELKE PEPERMANS

Abstract

Teacher decision making has a great impact on the quality of education in schools, yet we know little about how teachers make decisions in practice. It is assumed that teachers use both intuition and data in the different steps of the decision process. No reliable, valid scales are available to research both dimensions during the different steps of teachers' decision process (problem definition, data collection, sense making, and evaluation of alternatives). Building on the integrated framework we constructed in earlier research, the main aim of this study was to develop and validate a Teacher Decision-Making Inventory (TDMI). One hundred and one teachers in adult education participated voluntarily in a web-based survey. Based on the good EFA factor loadings, the CFA fit indices, and the internal consistency (Cronbach's alpha), we conclude that the TDMI is a valid psychometric tool that can be used to assess the intuitive and data-driven dimensions of teachers' decisions in large-scale quantitative research.

Keywords

teachers' decision-making, evidence-informed, intuition, data, validation, professional judgment, dual process

Introduction

Decision making is an important topic in education, since teachers' decisions greatly influence pupils' trajectories, especially when the stakes are high (e.g., passing or failing, moving on to the next educational track; Bonvin, 2003). Therefore, it is important for teachers to make wise, professional decisions. However, we know little about how teachers actually make decisions (Earl & Katz, 2006; Eurydice, 2011; Harteis et al., 2008). This leads to important questions: what are professional decisions, and how can we better understand decision making in practice?

In education, research that studies teacher judgment has shifted from a personal knowledge perspective based on expertise within the teaching profession towards an emphasis on data-based decision making (DBDM). Following disappointing findings with regard to the accuracy of teacher judgment (Urhahne & Wijnia, 2020), policymakers and researchers expected educational decision making to become more data informed (Mandinach, 2006). Data use models describe optimal teacher judgment as based on a systematic inquiry cycle: problem definition, data collection, analyses, and interpretation to evaluate alternatives before a decision is made (Datnow et al., 2007; Schildkamp & Lai, 2013; Strayhorn, 2009).

In the broader field of decision theory, many scholars have agreed that human judgment is guided by both data and intuition, which may influence the different steps of the decision process to a greater or lesser extent (Blackwell et al., 2006; Evans, 2008; Kahneman & Frederick, 2002). Although educational research mostly studies teacher judgment from either a data use or a teacher knowledge perspective, it seems appropriate to assume that both dimensions will influence teacher judgment in practice (Evans, 2008; Klein, 2008; Tversky & Kahneman, 1981). The question is not whether teachers make intuitive or data-based decisions. It is more interesting to grasp the extent to which teachers use data or intuition in the different steps of the decision process.

In the past, studies have indicated that teacher judgment shows much variability at the level of the individual teacher (Kaiser et al., 2013). In earlier research (Vanlommel et al., 2017, 2018, 2020), we therefore used a qualitative case study design to explore how teachers differ in the extent to which they use data or intuition before the final decision is made. Based on the level of data and intuition use by teachers, we identified four different approaches to decision making: (a) rational (high on data, low on intuition); (b) intuitive (high on intuition, low on data); (c) professional (a combination of both); and (d) arbitrary (a restricted decision process involving little use of data or intuition).

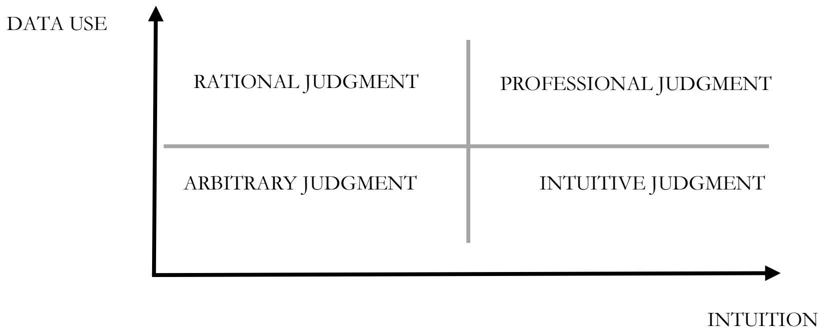


Figure 1 (Vanlommel, 2018)
Teachers' approaches to decision making

When we wanted to take the next step, to quantitatively study the use of data and intuition in the decision process, we encountered two major problems. First, there was no framework available that integrated data and intuition in the different steps of the decision process, and thus no instruments to study teacher decision making on a large scale. Second, there was conceptual haziness about intuition in the context of teacher judgment. We needed a transparent definition that disentangled the confusion stemming from a lack of insight and that permitted empirical research on this topic in education. We tackled both obstacles in previous research (see Vanlommel, 2018) and we build on those insights to develop and validate a Teacher Decision-Making Inventory (TDMI) that allows large-scale research in order to advance the field.

Theoretical Framework

The decision process

Both data-based and intuitive processes are considered to be valuable parts of teacher judgment that each have their own merits and pitfalls; thus, they need to be combined to make the best professional decisions possible. For example, intuitive recognition can allow teachers to recognize a problem quickly at an early stage, and expert knowledge is important to understand what data mean in a specific context. At the same time, research has shown that intuition may be vulnerable to different sources of bias (Burgess et al., 2009; Kahneman & Frederick, 2002). For one thing, teachers may mainly pay attention to indicators that confirm what they already believe and often ignore data that indicate the contrary (Goldstein & Hogarth, 1997; Harteis et al., 2008; Klein, 2008; Tversky & Kahneman, 1981). This may lead to self-fulfilling prophecies that perpetuate social or economic disparities. In order to prevent the pitfalls of confirmation bias, data use is crucial for questioning assumptions. To make good decisions, teachers are required to use both data and intuition in the different steps of the decision process. Information deriving from one source can complement information from another (Earl & Katz, 2006; Kahneman & Klein, 2009); the complexity of conclusions related to pupil competence also requires a detailed and balanced view drawn from more than one data source (Cohen et al., 2017.)

Professional decision making from an integrated perspective

In the field of decision making, theories on dual-process approaches to decision making indicate that data-based and intuitive processes both influence human judgment (Hogarth, 2001; Klein, 2008). In the field of education, theories that approach decision making as a dual process influenced by the use of both data and intuition are scarce. In earlier research (Vanlommel et al., 2017), we combined theories of DBDM that are commonly used within education (Datnow & Hubbard, 2015; Mandinach & Jimerson, 2016; Schildkamp et al., 2016) with the theory of naturalistic decision making that studies intuition as expertise (Klein, 2008). The recognition-primed decision model describes how experts can use their professional knowledge of subject and context to make accurate decisions, based on their expertise (Klein, 2008). We will elaborate on both dimensions in the next paragraphs and integrate them in the theoretical framework that will be used to develop our questionnaire.

What is DBDM and how can it contribute to teacher judgment?

In a movement away from the era in which research primarily studied how teachers' intuitive knowledge influenced the outcomes of teacher judgment, the initial body of data use research mainly conceptualized data as quantitative indicators of pupils' cognitive output (Hubbard et al., 2014). This was based on the assumption that the quality of educational decisions would increase to the extent that they were based on objective measures, such as standardized tests.

More recently, scholars have criticized this narrow view because it inhibits a full understanding of pupil competences and it has led to undesirable practices such as 'teaching to the test' (Brown, 2017; Ehren & Swanborn, 2012). Therefore, broadening the concept of data to include all indicators that inform some aspect of schooling has been advocated (Schildkamp & Lai, 2013). These definitions of data include quantitative measures, such as results from (standardized) tests or attendance rates, but also qualitative indicators, such as observations in the classroom or conversations with colleagues, pupils, or parents. To differentiate the formal use of data from incidental, spontaneous gathering of indicators, data collection needs to be initiated based on a clear goal or question and to follow an inquiry cycle (Earl & Louis, 2013; Schildkamp & Lai, 2013). Data are collected systematically and deliberately (Bromme et al., 2014).

DBDM can then be defined as a systematic process in which (1) a problem or question is diagnosed using at least one type of data collected deliberately and systematically, (2) data are collected systematically with the aim of exploring the question or problem, (3) data are interpreted by objective criteria, and (4) evaluative arguments are based on data use in steps 1–3 (Coburn & Turner, 2012).

What is intuition and how can it contribute to professional teacher judgment?

Theories of naturalistic decision making focus on the value of expert intuition, originating from early research on master chess players who were able to make accurate decisions because they recognized cues and complex patterns (Chase & Simon, 1973). This led to the definition of intuition as recognition, and was elaborated further in the recognition-primed decision model (Klein, 2008). Klein (2008) described how subject-matter experts are able to make good decisions in complex contexts because they recognize cues and patterns based on the expert knowledge stored in their memory, without a deliberate and systematic search. Applied to teacher judgment, this means that teachers are able to recognize a problem spontaneously, without using data in their diagnosis, and to make a decision without a deliberate and systematic collection and analysis of data (Kahneman, 2003; Klein, 2008). Teachers hold patterns in their memory, based on learning and experience, that draw attention to cues without a deliberate search for answers to a question. The spontaneous recognition of elements in a given situation triggers expectancies for the future based on similar cases in the past, and thereby informs decision making without deliberate analyses. In our study, we define intuition as a personal knowledge base that consists of patterns and mental models teachers have acquired through learning and experience, enabling them to recognize cues and solutions spontaneously without deliberate attention or a systematic approach.

Intuitive processes of decision making refer to (1) spontaneous recognition of a problem without further diagnosis, (2) automatic collection of information without a deliberate or systematic approach, (3) interpretation based on personal criteria, and (4) evaluative arguments based on evidence collected through intuitive processes in steps 1-3.

To an integrated framework: Data use and intuition integrated in the different steps of the decision process

In step 1, a problem or goal is defined when the actual situation is weighed against personal or objective standards for the desired situation (Mintzberg & Westley, 2001; Schildkamp et al., 2016).

A decision process may be initiated when a teacher recognizes a problem spontaneously without deliberately weighing the actual state of affairs against the standards. For example, a teacher spontaneously notices that a pupil is staring out the window during daily work. The teacher feels this might be a problem and keeps this information in mind. This intuitive problem recognition might or might not be followed by problem diagnosis. For example, the teacher can start observing this pupil using an observation protocol, focusing on pre-planned indicators.

Once teachers have defined the problem, this is expected to trigger a wider search for more data (Evans, 2008; Schildkamp et al., 2016). In step 2 (data collection), a data search may or may not be guided by the problem or question defined in step 1 or by a clear plan (e.g., Mandinach et al., 2006). For example, when the teacher defines the problem as a student's possible learning disorder in step 1, that teacher can develop a plan: what data do I need and how do I collect the data in order to gain fine-grained insight into the problem? Intuitive data collection might start from the same problem definition but is not guided by a plan. During teachers' daily practice, their attention is spontaneously drawn by elements that (mostly) confirm or (seldom) question their initial problem recognition. Independent of the rational or intuitive nature of teachers' data collection, in step 3, data need to be analyzed and interpreted before they can inform teachers' decision making (Bertrand & Marsh, 2015). In this sense-making process, it has been suggested that although data use models to prescribe optimal procedures for coming to valid conclusions (Bosker et al., 2007), in practice teachers might take mental shortcuts (heuristics) to reach quicker and easier conclusions (Evans, 2006; Kahneman, 2008; Klein, 2008). False inferences are often explained in terms of confirmation bias, when teachers frame the data to fit their existing beliefs (Harteis et al., 2008; Kahneman & Frederick, 2002). Therefore, it is important to look at the criteria used when teachers make inferences. While DBDM refers to the use of objective, pre-defined criteria, teachers might also trust in personal criteria to make sense of data (Vanlommel & Schildkamp, 2019).

In the fourth step, after teachers have run through steps 1-3, an important question concerns the extent to which they take data and intuition into account when they evaluate alternatives and make a decision. Information deriving from data-based and intuitive processes may coincide and thus strengthen teacher judgment, or it may provide contrasting viewpoints. In that case, an important question to investigate is how teachers use data and intuition to reach their final decision.

Even decision processes that are predominantly led by data use processes may result in intuitive judgment when information deriving from one intuitive cue overrules all other evidence. Research has shown that the decisive criteria applied by teachers are often based on subjective beliefs about good learning and teaching (Allal, 2013; Rubie-Davies, 2010; Zanting et al., 2001). For example, despite test results, reports, or conversations with colleagues, a teacher may rely on their personal trust or distrust in the student's motivation.

We approach professional decision making as the combination of both dimensions in the different steps of the decision process. Figure 2 provides a static visual overview of what is, in practice, a complex, iterative process.

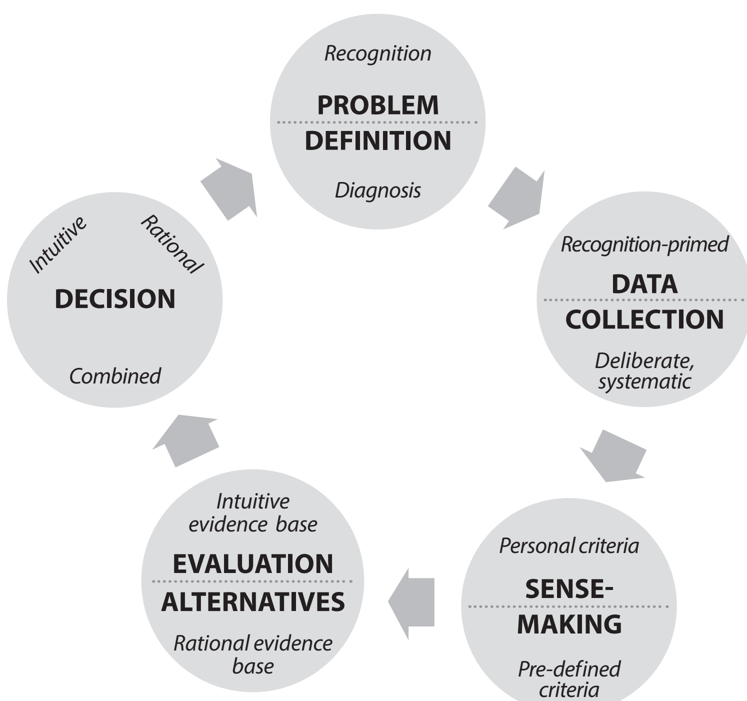


Figure 2
Theoretical model for the Teacher Decision-Making Inventory, based on Vanlommel et al. (2018)

Aim

Given the integrated framework we developed and tested in earlier research (Vanlommel et al., 2020), the main aim of this study is to develop and validate a questionnaire based on that theoretical framework that supports investigation of teachers' decision making in practice on a larger scale. This will contribute to our understanding of teacher decision making in practice and help strengthen the knowledge base on decision making in education.

Method

For the development and validation of the instrument, we used the five steps described by Hinkin (1998): (1) items were constructed based on a theoretical model, (2) the survey was administered to the target group, (3) the number of items was reduced by means of exploratory factor analysis (EFA), (4) the structure was confirmed by means of a confirmatory factor analyses (CFA), (5) the internal consistency of the final scales was measured.

Instrument

In this section we describe the step-by-step process of constructing and validating the instrument as suggested by Benson and Clark (1982). As a first step in phase 1 (planning), we formulated the purpose: we wanted to construct an instrument that allowed us to investigate decision-making in practice, starting from a dual process approach. Phase 1 was largely conducted in our previous research (Vanlommel, 2018) as it consisted of (a) a broad literature review and (b) semi-structured interviews. The literature review showed that no existing instruments were readily available and offered us guiding frameworks to study decision making (see Theoretical Framework). The semi-structured interviews offered us a rich qualitative insight in on how educators make decisions in practice. Both the literature review and interview results were used as input for phase 2 (Construction). In this phase, we constructed the instrument starting from in-depth, semi-structured interviews we had conducted in earlier research. In that research, we followed 32 individual cases (the decision process related to an individual student) during an academic year, using a case-study design with repeated interviews. This provided us with rich and dense descriptions we used to develop the questionnaire (inductive scale development; Hinkin, 1998). These items were derived from teachers' descriptions of what they did in the different steps of the decision process, how they used data, and how they used their intuition. For example, a teacher statement such as "In order to gain more understanding of the problem, I will observe my pupils during my daily practice and collect information on the fly" was translated into an item for phase 2 (data collection). This was followed by deductive scale development. We used the existing literature described above (on data use and the recognition-primed decision model) to develop a table of specifications (see Table 1) and complement the items.

Table 1
Table of Specifications for Scale

Dimension	Intuitive	Data-driven
Phase		
Problem definition	(6 items)	(3 items)
Collection	(3 items)	(4 items)
Sense-making	(6 items)	(4 items)
Evaluation	(4 items)	(3 items)

This was followed by a content validation session. The items were typed as they would appear in the final instrument without being allocated to the different scales. All items that referred to deliberate and systematic processes of collection and analyses, starting from a pre-defined goal or question, were to be allocated to the data-driven dimension. All items referring to recognition—information gathering without a deliberate or systematic goal or plan—were to be allocated to the intuitive dimension.

The two authors and one colleague (researcher) independently matched the items with the scales. All items were allocated to the same scale except for two. The first was "When I make sense of data, I discuss this with colleagues." It was not clear to what extent this referred to data-driven or intuitive processes. During our collegial consultation, we changed the item to "When I make sense of data, I use shared criteria discussed with a colleague." The second item was "When I evaluate alternative decisions, I tend to rely most on what's in the student's best interest." This was not clear and was therefore changed to "most on *my feeling about* what is in the student's best interest." After this construction of the instrument, Hinkin (1998) stressed the necessity of a qualitative pre-test to establish construct validity. We ran a pre-test with four teachers in adult education and two peers who were fellow researchers (Cohen et al., 2017). While participants filled out the instrument, a think-aloud protocol was used to strengthen the cognitive validity: did teachers interpret the items in the same way that we as researchers intended (Field, 2009)? An interview was administered after the survey had been completed, to assess each individual item's suitability, face validity, and readability (Burgess et al., 1998).

After adjustments were made to the instrument's items based on the comments, the instrument was converted into a survey of 33 items with a 4-point Likert scale response format (ranging from "1=not important at all" to "4=extremely important"). Thus the higher the score, the more important the item for the respondent. A Likert scale is often used to measure respondent's attitudes by asking the extent to which they agree or disagree with a particular question or statement (Benson & Clark, 1982). The survey was structured around the 4 steps of the decision process (problem definition, data collection, sense making, evaluation of alternatives) for each of the two dimensions (data-based and intuitive). The Teacher Decision-Making Inventory (TDMI) can be found in Appendix 1.

Data analyses

All statistical analyses were carried out in open source R software, using the lavaan package (Rosseel, 2012).

Exploratory Factor Analyses

The factor structure was tested by carrying out exploratory factor analyses (EFA) for each of the steps of the decision process. Given our hypothesis that data-based and intuitive elements of teacher decision making mutually influence each other, we used oblique rotation, because it accounts for the expected relationship between the different factors (Loehlin, 2004).

Three elements were taken into account when defining the likely number of factors:

(1) applying the Kaiser criteria by calculating the number of factors with eigenvalues > 1;
(2) visually inspecting the scree plot;
(3) checking factor loadings and seeing whether there was a sound theoretical explanation.

After we had defined the number of factors, we inspected the loadings and deleted the items with unsatisfactory loadings (Field, 2009). In the final instrument, we kept only those items with high factor loadings on their own factor (≥ 0.30) and no/low loadings on the other factor. If an item did load on two factors, the difference between the two loadings should be > 0.15.

Confirmatory Analyses of the Structure of the TDMI

We performed confirmatory factor analyses (CFA) to test if the factorial structure was consistent with the theoretical model we developed for the instrument. We used the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). CFI values ≥ 0.90 and an RMSEA value ≤ 0.05 were taken as indications that the data showed a relatively good fit with the model (De Maeyer & Kavadias, 2007).

Reliability Analysis

The internal consistency of the instrument was measured by calculating its Cronbach's alpha coefficient. According to De Maeyer and Kavadias (2007), a scale with a Cronbach's alpha coefficient (α) in the range of 0.60 to 0.80 has reasonably good internal consistency, in the range of 0.80 to 0.90 has good internal consistency, and $\alpha > 0.90$ shows excellent internal consistency.

The scale has poor internal consistency if $\alpha < 0.60$ and unacceptable internal consistency if $\alpha < 0.50$.

Participants

The web-based survey was administered to 101 teachers: 84 women (84%) and 17 men (16%) in adult education in Flanders (Belgium). The population of adult educators in Flanders consists of 665 teachers: 552 women (84%) and 113 men (16%). Although it is a small simple size for this pilot in validating the questionnaire, the sample is a good representation of the population in the distribution over men and women. Teacher participation was voluntary and all participants signed an informed consent form. They were informed about the purpose of the study, that they could decide to end their cooperation at any time, and that results could not be traced back to a single teacher's responses. Anonymity and preservation of the privacy of each participant was guaranteed.

Results

Exploratory Factor Analyses (EFA)

First, we conducted a data-driven approach. The factor structure was initially tested by carrying out exploratory factor analyses (EFA) with oblique rotation for each step of the decision process. The likely number of factors was found using the Kaiser criteria (eigenvalues > 1) and scree plot analyses (factors before the first inflection point). The analysis resulted in a two-factor solution for all four steps. The first factor referred to the intuitive dimensions of teacher decision making; the second factor referred to the data-based dimension of teacher decision making. Subsequently, factor loadings were checked and items were included if loadings were ≥ 0.30.

Table 2 shows the results of the factor analyses for the different phases of the decision process.

Table 2

Factor loadings (P = problem definition, D = data collection, S = sense making, E = evaluation of alternatives)

	Factor Loadings	
Items	Factor 1	Factor 2
The following aspects of the student's learning lead me to identify problems in relation to promotion…		
P1. Concentration-related behaviors that catch my attention	0.48	
P2. Motivation-related behaviors that catch my attention	0.54	
P3. Behaviors related to the student's interest in learning that catch my attention	0.61	
P4. Data that I analyze in the student tracking system		0.38
P5. Deficits related to literacy that I spontaneously recognize	0.37	
P6. Characteristics related to social status that I spontaneously recognize	0.59	
P7. Behaviors related to work ethic that catch my attention	0.47	
P8. Information from a regular meeting with a colleague		0.66
P9. Information from a team meeting		0.87
When I need more information in relation to the problem, I…		
D1. Observe the student using an observation protocol		0.49
D2. Search for information in the literature		0.45
D3. Read the notes I make during my daily practice D4. Administer a targeted tests or assignment (e.g., to measure literacy)		0.81 0.32
D5. Retrieve information from memory of similar cases in the past	0.49	
D6. Feel what my intuition tells me	0.74	
When I make sense of data, I…		
S1. Take into account the effort a student makes	0.51	
S2. Take into account the student's socio-economic situation	0.72	
S3. Take into account the student's first language	0.61	
S4. Take into account the student's well-being	0.71	
S5. Take into account the student's social behavior	0.48	
S6. Adjust my evaluative criteria to meet the student's individual needs	0.51	
S9. Use shared criteria discussed with a colleague		0.82
S10. Use fixed criteria that apply for the school		0.41
S11. Weigh this result against earlier results		0.40
S12. Use criteria discussed with the students		0.30
When I evaluate alternative decisions, I tend to rely most on…		
E1. Information on the student's well-being gathered on the fly	0.85	
E2. Information on the student's social background gathered on the fly	0.71	
E3. Information on the student's motivation gathered on the fly	0.60	
E4. Results of the student's self-evaluation		0.53
E5. Information on the requirements of the future track		0.91
E6. My feeling about what is in the student's best interest	0.56	
E7. Test results		0.31

Confirmatory Factor Analyses (CFA)

Subsequently, we tested whether the two-factor structure fit within our theoretical model build around the four phases of the decision process. We subjected this model to confirmatory factor analyses to confirm our model, but the initial model did not fit the data well (CFI = 0.77; RMSEA = 0.76). We carefully studied modification indices, looking for a better fitting model, with theoretical considerations also being taken into account. Given the observed cross-loading of the item "conversation with colleagues" (evaluation of alternatives) with "data collection" and "sense making," we deleted this item. Further, error covariances were included.

Our final model is shown in Figure 3. Based on the goodness-of-fit indices, we concluded that our data show a good fit with the model (CFI = 0.90; RMSEA = 0.05).

Reliability: internal consistency

Subsequently, the internal consistency was calculated. The different subscales appear to show reasonably good to good internal consistency: problem recognition (α = 0.72), problem analysis (α = 0.74); intuitive collection (α = 0.69), data-based collection (α = 0.60); subjective interpretation (α = 0.84), objective interpretation (α = 0.72); intuitive evaluation (α = 0.82), and data-based evaluation (α = 0.70).

Discussion and Conclusion

The aim of the present study was to fill an important gap in the research: the lack of a validated instrument that allows the investigation of teacher decision making on a large scale, adopting a dual-process approach. Moreover, fine-grained insight into how teachers use data or intuition in the different steps of the decision process is scarce. In our study, we developed and validated a Teacher Decision-Making Inventory (TDMI) that measures two dimensions (data use and intuition) in the four different steps of the decision process: (1) problem definition, (2) data collection, (3) sense making, and (4) evaluation of alternatives. This questionnaire was built in two phases. The starting point was our theoretical model derived from earlier research (Vanlommel et al., 2020). Based on the rich and dense descriptions from teachers in our qualitative research, we developed items for each of the steps in the model. In a second phase, this survey was pre-tested in practice and discussed in the research team before it was administered.

Exploratory factor analyses identified two dimensions: data-based and intuitive, with good, unique factor loadings. Ideally, we would have split the sample in two, using the first half for exploratory factor analyses to identify the initial structure and using the other half for confirmatory factor analyses.

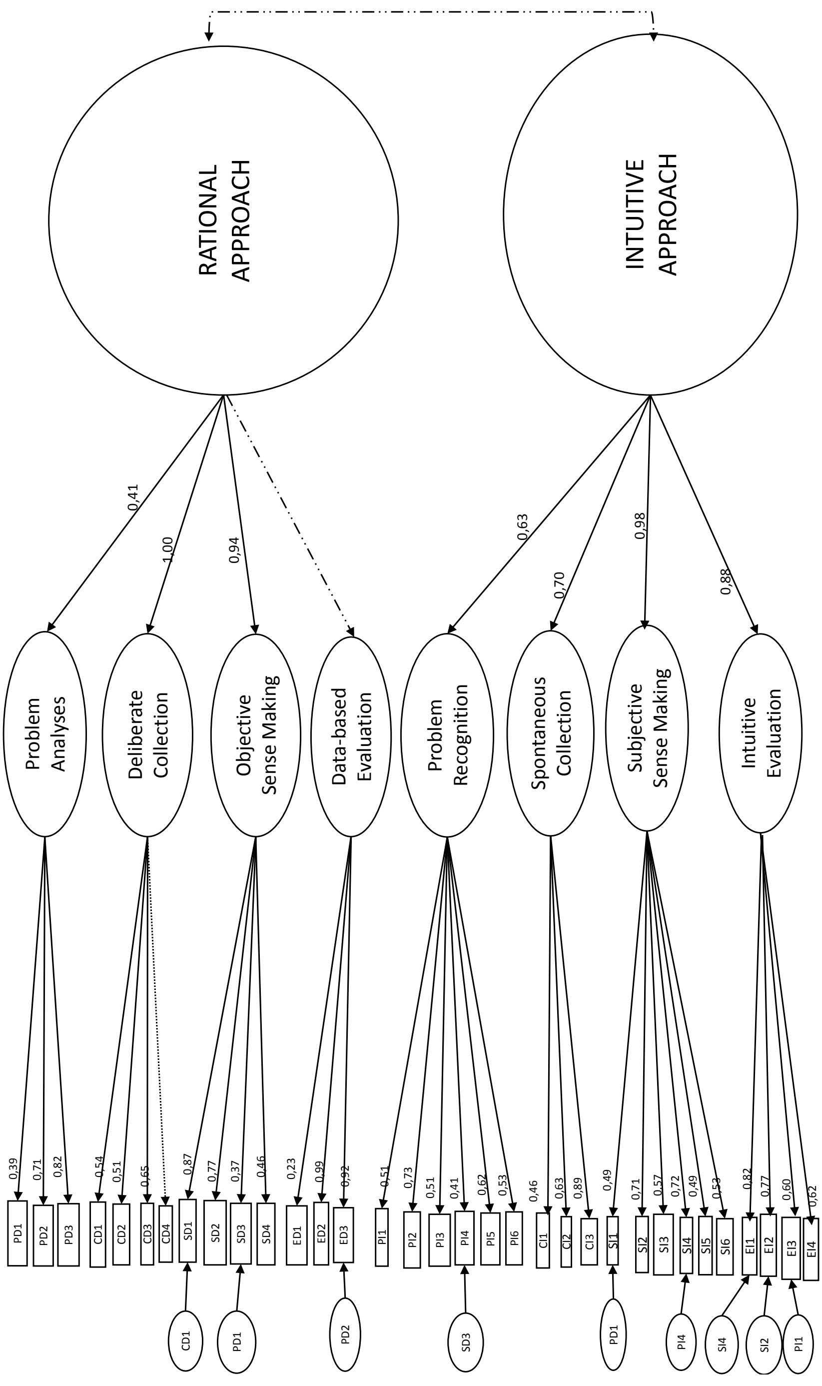


Figure 3
Findings of the confirmatory factor analyses (CFA) for the Teacher Decision-Making Inventory (TDMI). All factors with full arrows significant at the 0.05 level.

Our sample was, however, not large enough to follow this approach. This research can be considered as a pilot study in which the first, important steps were taken to develop and validate a teacher decision-making inventory. The scales measuring data-driven and intuitive collection of information need extra consideration.

Overall, our confirmatory factor analyses showed a good fit for our model, while reliability analyses showed reasonably good to good internal consistency of the scales. It can thus be concluded that the TDMI is both grounded in theory and a good psychometric tool that can be used to assess how data-based or intuitive a teacher's approach is in making decisions. This is an important step for research, policy, and practice to understand and support professional decision making in education.

Our starting point was that professional decision making begins with a wise combination of data and intuition, collected, analyzed, and weighed through an extensive decision process. The main question is not the extent to which teachers use either data or intuition; the crux of the matter is the extent to which teachers critically question problems they recognize, consciously search for answers, combine information, weigh alternatives, and conduct a decision process deliberately and skillfully. Our validated survey is a valuable step towards exploring, explaining, and strengthening this professional decision making in practice.

There are, of course, also limitations to this study. For one thing, we had a small sample size. Planning, constructing, and validating instruments requires large amounts of time, large funding, and large sample sizes (Benson & Clark, 1982). Therefore, validation should be seen as a continual process. We feel that our small scale pilot study delivers a valuable and important starting point for the next step in validating the TDMI. Future research is needed, on a larger scale, in different contexts such as different educational levels and in different educational cultures or political structures. For one thing, the educational system of Flanders (Belgium) is characterized by high decision-making autonomy and low accountability: there is, for example, no binding obligation to use the results of standardized test for streaming or tracking. Other decision-making processes or data may appear in other systems.

In order to contextualize and standardize the questions for all teachers to some extent, we also focused on a tough promotion decision. Given the high stakes related to promotion to a subsequent educational level/track or retention, we expected teachers to go through the decision process thoroughly and use a wide range of evidence before reaching a decision. Further research could test and further develop our survey for other decisions (e.g., student placement in groups, evaluation of creative projects, curriculum redesign, and so forth). In order to enhance the validity and reliability of teacher decisions, it is important to gain a fine-grained understanding of how teachers use data or intuition in different contexts, for different decisions. These insights can help practitioners, researchers, and policymakers to develop targeted support and training that strengthens professional decision making in education.

We do hope that our research will be used as a starting point for further exploration and validation of decision-making processes in practice. Investigating, understanding, and enhancing the quality of teacher judgment is important as it highly influences student's educational trajectories, fairness, and equity.

References

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice, 20*(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364

Benson, J., & Clark, F. (1982). A guide for instrument development and validation. *American Journal of Occupational Therapy, 36*(12), 789–800. https://doi.org/10.5014/ajot.36.12.789

Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal, 52*(5), 861–893. https://doi.org/10.3102/0002831215599251

Blackwell, R. D., Miniard, P. W., & Engel, J. F. (2006). *Consumer behavior.* Harcourt College Publishers.

Bonvin, P. (2003). The role of teacher attitudes and judgement in decision-making: The case of grade retention. *European Educational Research Journal, 2*(2), 277–294. https://doi.org/10.2304/eerj.2003.2.2.6

Bosker, R. J., Branderhorst, E. M., & Visscher, A. J. (2007). Improving the utilisation of management information systems in secondary schools. *School Effectiveness and School Improvement, 18*(4), 451–467. https://doi.org/10.1080/09243450701712577

Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft, 17*(4), 3–54. https://doi.org/10.1007/s11618-014-0514-5

Brown, C. (2017). Research learning communities: How the RLC approach enables teachers to use research to improve their practice and the benefits for students that occur as a result. *Research for All, 1*(2), 387–405.

Burgess, P. W., Alderman, N., Evans, J. O. N., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*(6), 547–558. https://doi.org/10.1017/S1355617798466037

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2009). *What parents want: school preferences and school choice.* CMPO.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. *American Journal of Education, 118*(2), 99–111. https://doi.org/10.1086/663272

Cohen, L., Manion, L., & Morrison, K. (2017). Action research. In L. Cohen, L. Manion, & K. Morrison (Eds.), *Research methods in education* (8th ed., pp. 440–456). Routledge.

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*(4), 1–26.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data*. University of Southern California, Center on Educational Governance.

De Maeyer, S., & Kavadias, D. (2007). *Openleerpakket beschrijvende statistiek*. Academia Press.

Earl, L. M., & Katz, S. (Eds.). (2006). *Leading schools in a data-rich world: Harnessing data for school improvement*. Corwin Press.

Earl, L., & Louis, K. S. (2013). Data use: Where to from here? In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 193–204). Springer.

Ehren, M. C., & Swanborn, M. S. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement, 23*(2), 257–280. https://doi.org/10.1080/09243453.2011.652127

Eurydice. (2011). *Grade retention during compulsory education in Europe: regulations and statistics*. OECD.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Field, A. (2009). *Discovering statistics using SPSS*. SAGE Publications Limited.

Harteis, C., Koch, T., & Morgenthaler, B. (2008). How intuition contributes to high performance: An educational perspective. *US-China Education Review, 5*(1), 68–80.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Hogarth, R. M. (2001). *Educating intuition*. University of Chicago Press.

Hubbard, L., Datnow, A., & Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation, 42*, 54–62. https://doi.org/10.1016/j.stueduc.2013.10.003

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755

Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction, 28*, 73–84. https://doi.org/10.1016/j.learninstruc.2013.06.001

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis* (4th ed.). Lawrence Erlbaum.

Mandinach, E. B., & Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education, 60*, 452–457. https://doi.org/10.1016/j.tate.2016.07.009

Mandinach, E. B., Honey, M., & Light, D. (2006, April 9). *A theoretical framework for data-driven decision making*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mintzberg, H., & Westley, F. (2001). It's not what you think. *MIT Sloan Management Review, 42*(3), 89–93.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology, 80*(1), 121–135. https://doi.org/10.1348/000709909X466334

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 177–191). Springer.

Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School Effectiveness and School Improvement, 27*(2), 228–254. https://doi.org/10.1080/09243453.2015.1056192

Strayhorn, T. L. (2009). Accessing and analyzing national databases. In T. Kowalski & T. J. Lasley (Eds.), *Handbook of data-based decision making in education* (pp. 121– 138). Routledge.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1007/978-1-4613-2391-4_2

Urhahne, D., & Wijnia, L. (2020). A review on the accuracy of teacher judgments. *Educational Research Review*, 32, 100374. https://doi.org/10.1016/j.edurev.2020.100374

Vanlommel, K. (2018). *Opening the black box of teacher judgement: The interplay of rational and intuitive processes* [Dissertation]. University of Antwerp. https://www.researchgate.net/publication/324909244_Opening_the_black_box_of_teacher_judgement_the_interplay_of_rational_and_intuitive_processes_Doctoral_thesis

Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal, 56*(3), 792–821. https://doi.org/10.3102/0002831218803891

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven*? International Journal of Educational Research, 83*, 75–83. https://doi.org/10.1016/j.ijer.2017.02.013

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education, 71*(1), 108–119.

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2020). Sorting pupils into their next educational track: How strongly do teachers rely on data-based or intuitive processes when they make the transition decision? *Studies in Educational Evaluation*, 69, 100374. https://doi.org/10.1016/j.stueduc.2020.100865

Zanting, A., Verloop, N., & Vermunt, J. D. (2001). Student teachers' beliefs about mentoring and learning to teach during teaching practice. *British Journal of Educational Psychology, 71*(1), 57–80. https://doi.org/10.1348/000709901158398

Corresponding authors
Kristin Vanlommel
University of Applied Sciences Utrecht, the Netherlands
E-mail: kristin.vanlommel@hu.nl

Elke Pepermans
Institute for Adult Education Antwerp, Belgium
E-mail: elke.pepermans@cvoantwerpen.be

Appendix 1

Teacher Decision-Making Inventory (TDMI)

This inventory is designed to explore how you make a tough transition decision in relation to promotion to the next educational level/track. We ask you to picture a tough case, in which the decision is not straightforward. The questions investigate the different phases of the decision process during the year. How were you able to define the problem in relation to promotion at the start of the year? How did you collect more information during the year? How did you interpret test results for this student and on what evidence base would you rely most at the end of the year?
You can answer these questions on a scale ranging from (1) not important at all to (4) very important.

Statement	Assessment			
The following aspects of the student's learning lead me to identify problems in relation to promotion…				
Concentration-related behaviors that catch my attention	1	2	3	4
Motivation-related behaviors that catch my attention	1	2	3	4
Behaviors related to the student's interest in learning that catch my attention	1	2	3	4
Data that I analyze in the student tracking system	1	2	3	4
Deficits related to literacy that I spontaneously recognize	1	2	3	4
Characteristics related to social status that I spontaneously recognize	1	2	3	4
Behaviors related to work ethic that catch my attention	1	2	3	4
Information from a regular meeting with a colleague	1	2	3	4
Information from a team meeting	1	2	3	4
When I need more information in relation to the problem, I…				
Observe the student using an observation protocol	1	2	3	4
Search for information in the literature	1	2	3	4
Read the notes I make during my daily practice	1	2	3	4
Administer a targeted tests or assignment (e.g., to measure literacy)	1	2	3	4
Retrieve information from memory of similar cases in the past	1	2	3	4
Feel what my intuition tells me	1	2	3	4
When I make sense of data, I…				
Take into account the effort the student makes	1	2	3	4
Take into account the student's socio-economic situation	1	2	3	4
Take into account the student's first language	1	2	3	4
Take into account the student's well-being	1	2	3	4
Take into account the student's social behavior	1	2	3	4
Adjust my evaluative criteria to meet the student's individual needs	1	2	3	4
Use shared criteria discussed with a colleague	1	2	3	4

Use fixed criteria that apply for the school	1	2	3	4
Weigh this result against earlier results	1	2	3	4
Use criteria discussed with the students	1	2	3	4
When I evaluate alternative decisions, I tend to rely most on…				
Information on the student's well-being gathered on the fly	1	2	3	4
Information on the student's social background gathered on the fly	1	2	3	4
Information on the student's motivation gathered on the fly	1	2	3	4
Results of the student's self-evaluation	1	2	3	4
Information on the requirements of the future track	1	2	3	4
My feeling about what is in the student's best interest	1	2	3	4
Test results	1	2	3	4

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-4

SENSEMAKING UNRAVELED: HOW TEACHERS PROCESS SCHOOL PERFORMANCE FEEDBACK DATA

GILA GUTWIRTH,
EVELYN GOFFIN,
JAN VANHOOF

Abstract

The present study investigates how Flemish middle school mathematics teachers make sense of school performance feedback data from low-stakes, external standardized tests. We take an in-depth look into the interpretive steps they take, based on a conceptual model that integrates intuitive and rational aspects of individual and collective sensemaking and empirical data collected in semi-structured interviews. We describe the nature of these sensemaking processes and consider the impact of influencing factors. Our findings demonstrate that the mere availability of school performance feedback data does not spontaneously spark sensemaking, nor does it necessarily lead to improvements in instructional practice. Teachers' sensemaking of school performance feedback data appears to be a largely intuitive process, grounded in external attributions and absent of triangulation. Challenges regarding expertise and lack of inquiry-based attitude and commitment result in superficial and often incorrect interpretations of the data that tend to remain uncorrected as teachers barely engage in collaborative professional dialogue about the data.

Keywords

sensemaking, data-driven decision making, school performance feedback, intuition, rationality

Problem Statement

In recent decades, extensive decentralization and deregulation have resulted in a growing degree of autonomy for schools worldwide in terms of shaping their school policies. Policymakers assume that schools possess sufficient policy-making capacity to implement high-quality policies and to investigate and monitor their internal quality in a systematic way. Moreover, both in educational research and from a societal point of view, there is an increasing emphasis on informed school development that is based on objective, reliable, and valid data and not just on intuition and experience (Lai et al., 2014; Van Gasse et al., 2017). Educational professionals might use a host of data to inform policy and practice, including formal data such as school performance feedback or student achievement results from assessments, informal data such as classroom observations, and research evidence or even big data (Schildkamp, 2019). However, research shows that sustainable school development cannot be achieved by merely collecting and providing data (Vanlommel & Schildkamp, 2019). What raw data mean—what data mean in relation to pre-defined goals and how the data might serve to inform decisions and actions that effectively address school and student needs—is seldom self-explanatory: transforming data into information and subsequently into actionable knowledge requires recipients to *make sense* of the data within their own specific setting (Schildkamp et al., 2019).

Sensemaking is regarded as a crucial phase in the systematic, goal-oriented, and iterative process that is data-based decision making (DBDM) (Schildkamp et al., 2013; Verhaeghe et al., 2010). It involves actively analyzing data, forming interpretations, and making inferences (Coburn & Turner, 2011; Mandinach & Gummer, 2016; Schildkamp, 2019; Spillane & Miele, 2007; Vanlommel & Schildkamp, 2019). This interpretive process is neither straightforward nor exclusively rational (Bertrand & Marsh, 2015; Vanlommel & Schildkamp, 2019). Moreover, since individual sensemakers each have their own specific frame of reference and prior experiences that form personal, subjective lenses, the same data can come to hold different meanings for different educational professionals. Research therefore attests to the importance of collective sensemaking in DBDM in schools (Mandinach & Schildkamp, 2021). Collective sensemaking is a process of co-construction that takes shape in social interactions. With regard to teachers' team-level sensemaking, Bolhuis et al. (2016) referred to the positive impact of cognitive conflict: the tension created by divergent knowledge and assumptions between participants and by discrepancies between known and new information. Provided that cognitive conflict is embedded in a constructive collaboration based on openness and trust, it can lead to possible adjustments of instruction and learning, but this requires a profound professional dialogue among teachers.

However, research shows that this is the exception rather than the rule, resulting in superficial short-term solutions and quick fixes (Mausethagen et al., 2019).

The present study is a response to researchers' calls for more insight into the sensemaking process in DBDM (Mandinach & Schildkamp, 2021; Schildkamp, 2019). Our primary research aims are to investigate the interpretive steps teachers take when they make sense of data and to tap into the way teachers' individual sensemaking is embedded in collective sensemaking processes within their schools. In addition, we want to explore the impact of a number of explanatory variables on individual and collective sensemaking. A wide range of factors have been found to impact DBDM, including factors at the user level, the organization level, and the level of the data (systems) themselves (Bolhuis et al., 2016; Schildkamp et al., 2019; Van Gasse et al., 2015; Vanhoof et al., 2011; Verhaeghe et al., 2010). Personal characteristics such as data users' attitudes, self-efficacy, and data literacy have been discussed (Bolhuis et al., 2016; Prenger & Schildkamp, 2018; Van Gasse et al., 2015, 2017; Verhaeghe et al., 2010), as has the impact of collaboration on teachers' individual data use (Van Gasse et al., 2017) and contextual expectations such as the stakes associated with assessment or accountability structures in educational systems (Datnow & Park, 2018; Van Gasse et al., 2014; Vanlommel & Schildkamp, 2019). The relational factors that may affect data use in general and the sensemaking phase in particular have been less extensively explored.

With a qualitative, interview-based inquiry into teachers' individual and collective sensemaking, we want to address these knowledge gaps. We will do so by exploring a phase of sensemaking in relation to school performance feedback data—formal data, often achievement-based, that is confidentially provided to schools by an external party for self-evaluation (Schildkamp & Teddlie, 2008; Visscher & Coe, 2003). Specifically, we zoom in on teachers' use of school performance feedback data from low-stakes standardized testing aimed at internal quality assurance in Flanders (the northern, Dutch-speaking region of Belgium). Many studies addressing aspects of individual and/or collective sensemaking were set in high-accountability educational systems (e.g., Datnow et al., 2012) or (other) high-stakes decision-making contexts (e.g., Vanlommel & Schildkamp, 2019). However, expectations regarding DBDM are also becoming more salient in contexts and systems traditionally marked by a lower degree of accountability, such as the research context of Flanders. Performing our research in a low-accountability setting will allow us to unravel sensemaking processes from a school improvement logic that is minimally conflated with external (accountability) expectations.

In summary, with this qualitative study, we intend to examine how teachers make sense of school performance feedback data of standardized tests, and how this sensemaking process unfolds within the complexity of teachers' own school (team) context. In addition, we investigate whether certain explanatory variables have an impact on individual and collective sensemaking. This translates into the following research questions:

1) How do teachers individually and collectively make sense of school performance feedback data?
2) Which factors promote or hinder teachers' sensemaking of school performance feedback data?

Theoretical Framework

We intend to investigate what happens in terms of sensemaking when teachers are confronted with school performance feedback data; specifically, we examine how they engage with personalized feedback reports that provide output results. For the purposes of this investigation, we conceptualize sensemaking by teachers within the cycle of improvement-oriented data use as *a continuous process in which teachers, from their own frame of reference and (school) context, individually and in interaction, notice and interpret information from school performance feedback data in a manner that will enable them to transform their schools' results into decisions and actions aimed at improving instructional practice and student achievement.*

In the following paragraphs, we unpack and substantiate the different components of our proposed conceptualization. We start by generally situating the sensemaking construct and a number of its central tenets, grounded in insights from cognitive and social psychology as well as literature on organizational change and knowledge management. Next, we distinguish between individual and collective sensemaking. In terms of individual sensemaking, we will look at the interpretive steps teachers take when engaging with the data, i.e., how they notice specific elements or cues and subsequently interpret these cues by framing them and by forming a judgment (Coburn & Turner, 2011; Spillane & Miele, 2007). In order to account for the fact that (individual) sensemaking involves intuitive as well as rational processes, and in line with prior sensemaking research, we will employ a dual-processing perspective (Kahneman & Frederick, 2005; Vanlommel et al., 2017). In terms of collective sensemaking, we explore the nature and affordances of professional dialogue (Gergen et al., 2004; Tsoukas, 2009). After all, discussing sensemaking in the context of DBDM in schools needs to take into account the fact that data use is rarely an isolated activity (Schildkamp, 2019). Moreover, individual data processing is shaped by the context and social environment of the sensemaker (Coburn & Turner, 2011; Spillane, 2012). Finally, our selection of influencing factors to explore is focused primarily on the relational factors that might shape sensemaking and sensemaking outcomes in school teams, as the object of sensemaking under scrutiny in this study is formed by school performance feedback data. Our selection is based on the premise that collective inquiry is supported by the presence of sufficient human capital (pertaining to knowledge and expertise in participants) and social capital (pertaining to interaction between participants) (Christman et al., 2016).

Defining Sensemaking

The sensemaking construct has roots in literature on organizational change, crisis situations, information processing and workplace learning. Organizational psychologist Karl E. Weick (1995), historically regarded as one of the most influential sensemaking theorists, defines sensemaking as a social and continuous process in which people, from their own identities, retrospectively give meaning to cues and uncertainties in their environment, and proposes that (true) sensemaking leads to changes in beliefs or action. Broadly speaking, sensemaking thus has "outcomes" in the sense that it leads to some type of change in thought or behavior; it has cognitive, interpretive properties, as well as social, discursive properties; and it has a temporal dimension (Maitlis & Christianson, 2014).

Many theorists have elaborated on different aspects of sensemaking. Klein et al. (2007) discussed how sensemaking can be aimed at abstract or functional understanding of a situation or a novel set of data. In a DBDM discourse, we could connect those ideas of abstract vs. functional understanding to types of data use, such as to the distinction between conceptual and instrumental data use. Klein et al. (2007) regarded sensemaking primarily as a cognitive process, and discussed individual judgment, interpretation, and the role of internal mental models or the explanatory frames people use to interpret cues. Other authors, such as Cook and Gregory (2019), focused more on the discursive aspects of sensemaking, stating that sensemaking predominantly manifests itself through conversations and stories. Maitlis (2005) confirmed this social constructivist nature of the sensemaking process. In contrast, Klein et al. (2007) defined sensemaking as a cognitive process and focused on individual judgment, interpretation, and internal mental models. Hill and Levenhagen (1995) connected both views by stating that sensemaking consists of developing and formulating an individual vision or mental model that subsequently may serve to create support for one's views (sensegiving). Finally, regarding the temporal aspect of sensemaking, and in contrast to the original Weickian definition, authors such as Gephart et al. (2010) emphasized the prospective nature of sensemaking. They suggested that sensemaking is an ongoing and shared process in which meaning is produced, negotiated, and maintained through verbal and nonverbal communication, adding that sensemaking uses the past to give meaning to future actions in the present. Similarly, Weick et al. (2005) stressed the necessity of sensemaking in organizations in order to achieve long-term goals.

Interpretive steps in individual sensemaking, from a dual-processing perspective

Noticing

While reading school performance feedback reports, each individual teacher will—from their personal prior knowledge, expertise, previous (work) experiences, their beliefs about high quality instruction and about specific students and class groups, their identity and emotions—quickly and unconsciously notice certain cues and relate them to information already stored in memory (Klein et al., 2007; Kudesia, 2017; Maitlis et al., 2013; Weick et al., 2005). Intuitively noticing, recognizing, and selectively paying attention to information in school performance feedback data occurs automatically and reactively. These selective perceptions focus attention and thought, causing other information to be ignored or less relevant information to be amplified. Since interpretations resulting from this fast, unconscious "System 1" thinking (Kahneman, 2011) are based on incomplete information and information intuitively selected from data—whether scientifically collected or not—they are not always accurate (Shleifer, 2012). Additionally, this partly explains differences in interpretations between teachers (Spillane et al., 2002). Noticing more complex or unexpected cues in a feedback report, however, requires mental effort: the activation of teachers' cognitive abilities through rational "System 2" thinking (Kahneman, 2011). Teachers will notice additional cues if they review and reflect on school performance feedback data more consciously and systematically (Cook & Gregory, 2019).

Interpreting

Framing

Several theories and models attempt to describe the process of individual sensemaking, including the internal conceptual changes that it entails and the cognitive mechanisms associated with it (Zhang & Soergel, 2016; Zhang & Soergel, 2019). According to the data-frame theory of sensemaking (Klein et al., 2007), everything revolves around the connection of data with cognitive frames or, put differently, categorizing cues or stimuli from data and connecting them with individuals' pre-existing and internalized cognitive frames.

Intuitively noticing and recognizing specific information and specific cues in data (such as school performance feedback data) may feed into a variety of cognitive activities: elaborating an existing frame based on experience and advanced understanding, questioning a frame when the selected information is inconsistent with it, maintaining a frame by dismissing anomalous information, comparing alternative frames and searching for connections within the selected information, shaping a new frame if existing frames are insufficient, and searching for a frame by looking in the available information for more cues that might have been previously ignored. According to Attfield and Baber (2017), multiple frames are activated simultaneously when one processes data in order to form a personal account. These cognitive frames contain both general knowledge and more specific and situational knowledge and representations (Attfield et al., 2018).

Calabretta et al. (2017) argued that framing occurs both intuitively and rationally. Intuitive framing entails both the fast and unconscious activation of all cognitive frames related to the cues noticed in the data as well as unconsciously seeking holistic connections between these cognitive frames (Dane & Pratt, 2007). Rational framing, on the other hand, is an explicit, analytical reasoning process that requires more time and involves structuring limited information according to logical, substantive, and personal or context-specific criteria before being able to arrive at judgment (Calabretta et al., 2017).

Judgment

Research by Vanlommel et al. (2018) showed that teachers use data less rationally and objectively than one might expect. They often judge student outcomes intuitively, based on perceptions, personal criteria, and various non-cognitive indicators. Because they rarely seek any other data sources or consider alternative explanations, inferences based on feelings and personal beliefs arise, which can in turn lead to unsound interpretations of student outcomes.

Kahneman and Frederick (2005) argued that intuition and rationality are closely intertwined and influence each other. Dual process models discuss how the fast and intuitive system on the one hand and the more thoughtful and rational system on the other interplay (Whittaker, 2018). Whereas some researchers have postulated that intuition precedes rationality by serving as an input to deliberate and rational thought processes (Salas et al., 2010), Calabretta et al. (2017) emphasized the integration of both systems, considering intuition not subordinate to rationality, but rather complementary. They suggested alternating rational judgment (step-by-step, thoughtful cognitive evaluation) with intuitive judgment (unconscious, rapid, and affectively charged evaluation) and subsequently evaluating the product of this process rationally, and they argued that allowing and even encouraging a balanced integration of both systems would result in effective strategic decision-making. Cook and Gregory (2019) also emphasized the interplay between cognition, emotion, and judgment.

Kahneman and Tversky (1979) argued that judgment is always made in comparison to a status quo. This reference point will thus determine the risk a teacher is willing to take with respect to a possible adjustment of their own instructional practice. We propose that individual judgment of school performance feedback will always be grounded in the broader context of accountability and/or development (i.e., the summative and/or formative purposes for educational testing and the stakes involved), the prevailing school culture, and interrelationships within the team, as well as the specific background of the students and class groups who participated in the tests.

Professional dialogue and its affordances for collective sensemaking

Sensemaking does not only occur internally, but also in a dynamic process of co-construction in which individuals' selective perception and intersubjective interpretation is embedded in the environment through verbal and nonverbal communication (Cecez-Kecmanovic, 2004; Kudesia, 2017). Like individual sensemaking, collective sensemaking occurs both intuitively through the use of frames and heuristics, and rationally and consciously (Avby, 2015; Cook & Gregory, 2019). Both noticing and interpreting information takes place within a particular context in which personal, professional, organizational, and social influences interact. Cognitive, affective, and political aspects all influence which and how much information is shared by whom and with whom (Cook & Gregory, 2019).

The interaction between individual and collective sensemaking is a non-linear and iterative process that can ultimately lead to shared sensemaking, in this case sensemaking of school performance feedback data and, ideally, further improvement of educational quality. In this regard, Tsoukas (2009) and Gergen et al. (2004) emphasized the importance of in-depth professional dialogue. This is a conversation between professionals that is characterized by an open exchange of ideas, assumptions, and experiences and by the explication of tacit knowledge (Cook & Gregory, 2019; Nonaka & Takeuchi, 1995). Collective sensemaking, as in data discussions, broadens the interpretive lens through which data are viewed, stimulates debate between participants, and has the potential to create new knowledge, both on an individual level and on a shared level (Coburn & Turner, 2011; Datnow et al., 2012; Spillane, 2012). Voicing individual interpretations and inferences in data interactions helps to expose assumptions and ambiguities (Bertrand & Marsh, 2015; Christman et al., 2016). Successful and productive professional dialogue tends to benefit from participants' adoption of a non-judgmental, curious attitude and their (readiness to engage in) active listening (Gergen et al., 2004).

Relational Factors as Proposed Predictors

Expertise

Teacher expertise influences in-depth sensemaking of school performance feedback data. The literature emphasizes the expertise of veteran colleagues in this regard. Amidu et al. (2019), Chudnoff (2019), and Sinclair (2010) referred to intuitive expertise in order to elaborate on this. According to these scholars, more experienced colleagues possess more elaborate cognitive structures or frames to assess information quickly and intuitively without prior deliberation or an explicitly rational approach. Experts will discuss a problem of practice based on underlying principles rather than superficial features (Chi et al., 1981). Moreover, their well-developed metacognitive skills allow them to continuously monitor and evaluate their own reasoning and to switch between intuition and rationality without much effort (Amidu et al., 2019).

Based on a comparative study of professional dialogue in two departments, Horn and Little (2010) found that individual knowledge, skills, and experience contribute to the depth of professional dialogue. According to this research, "normalizing, specifying, revising, and generalizing" problems relating to concrete instructional practices of new teachers fosters adaptive expertise within a department. This generic expertise can subsequently be deployed to quickly solve similar problems, through extrapolation and the integration of data literacy with subject-specific, pedagogical-didactic, and instructional expertise (Mandinach & Gummer, 2016; Shleifer, 2012). Thus, in-depth professional dialogue is only possible if teachers and departments have sufficient individual and collective expertise. Expertise serves as both input and output of in-depth professional dialogue.

Inquiry-based attitude

In terms of making sense of (educational) data, the notion of expertise is closely related to that of data literacy: the capacity to identify problems, transform data into actionable knowledge, and evaluate outcomes (Beck & Nunnaley, 2021). However, Krüger (2010) stated that not every teacher needs to be data literate—teachers need to deploy an inquiry habit of mind to be able to use and handle data effectively. Amels et al. (2019) added that data literacy only has a small impact on teacher capacity for change with respect to instructional practices. They propose that inquiry-based working is much more important. Characteristics of an inquiry-based attitude include curiosity, critical (self) reflection, asking questions, willingness to change perspectives without judgment, openness, honesty, willingness to share with others, and a focus on data, accuracy, and thorough understanding (Krüger, 2018; Uiterwijk-Luijk et al., 2017). An inquiry-based and problem-solving school culture ensures that teachers—in the sensemaking phase—shift from external attributions to the acknowledgement of their own contribution (in interaction with their colleagues) and, in the process, also question their own practice.

Trust

Van Gasse et al. (2017) emphasized the importance of meaningful and authentic data use interactions between teachers. This requires a climate of psychological safety, in which teachers are confident and willing to be vulnerable. They dare to express doubts, raise problems, admit mistakes, and hold each other accountable for errors (Edmondson, 1999). Sufficient trust or confidence that taking interpersonal risks will not have negative (relational) consequences fosters cognitive and emotional commitment and mobilization of expertise (Edmondson & Lei, 2014).

Tsoukas (2009) argued that the productivity of professional dialogue is determined by the degree of relational commitment: co-workers taking collective responsibility for sensemaking of information and for reciprocal relationships within the team. They are open to being influenced and, through self-reflection, consciously scrutinize their usual ways of thinking and acting. After all, the meaning attributed to individual verbal and nonverbal expressions in a professional dialogue depends to a large extent on the response of the recipient and on the prevailing (school) culture. Additionally, thoughts and feelings are influenced by the (perceived) presence of others (Gergen et al., 2004). Roesch-Marsh (2018) also noted that facilitative relationships are indispensable for achieving deep professional dialogue.

Commitment

(Collective) sensemaking is not only a question of capacity (expertise and inquiry-based attitude) and confidence (trust) but also one of commitment: being willing to engage. Committed teachers consciously and voluntarily take individual and collective responsibility for student learning by participating in in-depth professional dialogue (Cameron & Lovett, 2015). They apply their subject-specific and pedagogical-didactic expertise to further develop their instructional practices within their own context (Sammons et al., 2007).

A study by Fransson and Frelin (2016) found that highly committed teachers exhibit a strong sense of professionalism. Challenging situations and complex problems encourage them to search for potential solutions. They feel responsible for their students' well-being and learning as well as for their own professional development. This commitment results in in-depth sensemaking of school performance feedback data and a positive impact on learning gain and student achievement (Day & Gu, 2007).

Methods

In order to gain in-depth insight into the complexity of sensemaking of school performance feedback data from standardized tests, we conducted semi-structured interviews aimed at exploring and describing current sensemaking practices (Alase, 2017). Based on teachers' perceptions and personal experiences, we aimed to better understand how they make sense of school performance feedback data and which factors promote or hinder teachers' sensemaking of school performance feedback data. The following sections discuss the research context and the more technical methodological choices and the approach we set forward to achieve this aim.

Research context

In the absence of central examinations, Flemish secondary schools have few standardized instruments at their disposal to measure student achievement or learning gain. They can, however, voluntarily participate in the Flemish national assessments, if they have been randomly selected to be part of the representative reference sample, or proactively decide to administer freely available parallel tests from these national assessments. The Flemish national assessments and parallel tests cover a range of learning subjects and measure the extent to which attainment targets are met: a set of formal learning objectives formulated by the government for the end of certain grades. The representative sample needed in these national assessments to conduct a valid assessment typically ranges from 10 to 20% of eligible Flemish schools, depending on the research design. On a yearly basis, parallel tests are typically used by (under) 10% of Flemish secondary schools. Overall, Flemish schools do not have a strong tradition of using externally generated output information (Van Gasse et al., 2015).

After participating in a national assessment or administering parallel tests, schools receive a school performance feedback report that gives statistical information about the proportion of students that reached the attainment targets, as well as value-added information based on a comparative analysis with schools that are similar in a number of input and context characteristics. These results are presented in graphical representations with an extensive reading guide, but do not include personalized recommendations. The reports of the parallel tests also include individual student results, displayed by attainment level. School performance feedback reports are strictly confidential and schools may only use the data for internal quality assurance. There are no stakes involved for schools in these tests.

Participants

For the present study, we recruited teachers from secondary schools that had recently (i.e., in the past school year) received school performance feedback after taking parallel tests of their own accord or after having agreed to participate in a national assessment. In the interest of homogeneity, and because the national assessment of mathematics in the second year had recently taken place, we focused on middle school mathematics teachers. No other (stratification) variables were taken into account when selecting participants. We started by recruiting participants in parallel test schools but moved to national assessment schools when responses proved insufficient.

In total, 11 teachers were interviewed, all of whom gave their informed consent to participate in the study. The majority of these teachers (9 out of 11) were female, and participant ages ranged from 24 to 55 years old. Two participants held master's degrees; the rest held bachelor's degrees. Participants had on average 12 years of experience working in education, with one participant being a first-year teacher, and the most veteran participant having 34 years of educational experience.

Interviews and procedure

All semi-structured interviews were conducted between early and mid-March 2020. The teachers were interviewed with the school performance feedback report that they had previously received on hand. A combination of interview questions and a think-aloud section enabled us to study the sensemaking process in depth, both from participants' retrospective accounts and perceptions, as well as from our own observations during the interviews (Eccles & Arsal, 2017). The interviews were audio recorded and transcribed verbatim.

To enhance construct validity, a pilot interview was conducted with a fictitious school feedback report. To ensure content validity, the interview guide was grounded in the concepts identified in the theoretical framework, with concrete open-ended questions attached to each concept, aimed at gauging participants' thoughts, experiences and perceptions. Examples of such questions include "What was the first thing you noticed when you went through this feedback report?" (intuitive noticing); "How did your department colleagues explain the results, and do you agree with them?" (professional dialogue); and "Do you feel that school results can be freely discussed within your team?" (trust). Additionally, in order to enrich our understanding of how the different steps of the sensemaking process take shape, questions were included regarding participants' sensemaking of the data as it occurred *during* the interviews. The think-aloud method offered a way to explore participants' thought processes.

Analysis

In order to get a general overview of the main research findings and to facilitate coding, the most salient results for each participant were summarized under each group of questions in the interview guide (Alase, 2017). Next, the 11 interviews were coded deductively in NVivo 12, using a coding tree based on the theoretical framework. To ensure the reliability and validity of the coding, the conceptual foundations of the main codes were written out (see Table 1). Some codes were further divided, adding sub-codes afterwards, in order to facilitate the analyses. In addition to the codes as listed in Table 1, there was also a code that focused specifically on the collective aspects of sense making. This code was added to the data if elements of noticing, framing, or judging also involved social interaction or forms of professional dialogue. The distinction between individual and collective sense making (cf. research question 1) was thus brought into the analyses. In order to analyze differences and similarities, the data were analyzed both horizontally and vertically (Cohen et al., 2011; Donche, 2015).

Table 1
(Non-exhaustive) Description of Key Codes Used

Codes	Conceptual foundations
Sensemaking	
Intuitive noticing	The teacher mentions things that immediately struck them in the feedback reports without conscious thought. The teacher mentions cues or things that stood out, which they immediately recognized from prior experiences, personal prior knowledge, beliefs about certain students or class groups, etc. The teacher states that they are not aware of what they noticed immediately during the initial reading of the report.
Rational noticing	The teacher mentions things that stand out when they review the feedback reports and think about them intentionally, consciously, and systematically (based on guiding questions).
Intuitive framing	The teacher mentions their first impressions of or the ideas they formed about certain graphs or tables at first glance, from intuitive expertise, prior experiences, personal beliefs, knowledge about certain students or class groups, etc.
Rational framing	The teacher mentions their impressions of or their ideas about certain graphs or tables after having consciously and systematically thought about them (based on guiding questions), having triangulated them with other data sources, having reflected on them further or having explicitly related them to other parts of the feedback reports through an analytical reasoning process.

Intuitive judgment	The teacher mentions their initial judgment, based on prior experiences (with particular students or class groups), feelings, perceptions, assumptions, etc. The teacher states that they are not aware of having formed a judgment during the initial reading of the report.
Rational judgment	The teacher mentions additional potential causes for the results presented in the feedback reports, identified through a step-by-step and conscious cognitive evaluation of the data.
Influencing factors	
Trust	The teacher mentions things that indicate a climate of psychological safety and trust in each other's abilities, such as being open about mistakes, daring to express doubts, raising problems and seeking solutions together regarding instruction and student outcomes, giving each other feedback, etc.
Participant expertise	The teacher understands the concepts used in the feedback report and can interpret the graphs and tables correctly. The teacher is aware of misinterpretations and makes adjustments when needed.
Expertise within the department	The teacher mentions that there is sufficient expertise within the department and is able to illustrate this with examples.
Inquiry-based attitude	The teacher mentions elements that indicate an inquiry-based attitude such as questioning one's own practice, raising questions for reflection, etc. (Distinction was made between "Participant level" and "Department level")
Commitment	The teacher mentions things that indicate participation and taking responsibility for student learning based on student outcomes. (Distinction was made between "Participant level" and "Department level")

Findings

In order to answer the first research question, we describe participants' experiences, approaches, underlying thoughts, and feelings while making sense of school performance feedback data. We consider our participating mathematics teachers' recollections of their initial sensemaking upon reception of the feedback reports, as well as their sensemaking process as it took place when discussing the reports during the interview. We do this successively for the different steps of individual sensemaking and for collective sensemaking.

Individual sensemaking of school performance feedback data

Our findings with regard to individual sensemaking will be presented according to the theoretical distinction we made between noticing, framing, and judging. We discuss the intuitive or rational manifestations of these steps. Table 2 provides an overview of how individual sensemaking took shape for each of the participants.

Table 2
Coding for Initial Individual Sensemaking

Participant	Noticing	Interpreting	
		Framing	Judging
1	I	I	I
2	C	C	I
3	I	I	I
4	C	C	I
5	I	I	I
6	C	I	I
7	I	I	I
8	I	C	C
9	I	I	I
10	I	C	I
11	I	I	I

Note: I = intuitive, R = rational, C = combination

Intuitive and rational noticing

The initial reading of the feedback report appears to have been a predominantly intuitive process for all participants. All participants also stated that, in that initial intuitive phase, they overlooked some (important) aspects of the feedback reports. Intuitively, they had particularly remarked how many of their students were not meeting the attainment targets, which for many participants corresponded to their expectations for certain class groups or individual students.

> Out of a certain curiosity, you spot the things you want to know. ... I'm pretty sure I overlooked a lot of important things. After all, there's a lot of text to read, a lot of graphs to interpret. You focus on things that immediately stand out, for example, the red color. – Participant 9

Two participants stated that they were not (or were no longer) aware of what they had intuitively noticed when they first went through the feedback reports.

Only three participants indicated that after an initial intuitive reading, they had reread the reports with a more rational, step-by-step approach, comparing students and subject domains. In such cases, we speak of "combined" noticing.

When we asked participants to review the feedback reports during the interviews, three of them stated that they did not notice anything new or additional. Other participants indicated that they had taken a closer look at specific parts of the report because time had now been explicitly made available for it and because they had been explicitly requested to do so.

A number of them stated that they had now looked at the individual student results more thoroughly and step-by-step, thus rationally, in consideration of a potential adjustment of their own instructional practice or to inform discussions of individual students' educational progress at teacher council meetings.

Interpreting: Intuitive and rational framing

Almost all participants indicated that they had not or had only barely questioned their initial intuitive framing of the results. The teachers stated that they knew their students well, that they knew what to expect from their students, and that they based their framing of the results on this knowledge. They mainly focused on their own class groups and intuitively compared their own students' achievement to that of other class groups or study options, to the statistically expected scores, or to the national average. For instance, they said they immediately observed how some attainment targets were achieved by only a limited number of students. Some participants got this information from the tables; others found the visual representations a clearer way to get a general idea of the results from a broader perspective.

We prompted participants to think rationally about their initial sensemaking of the results, and had them reflect on their thought process, approach, and underlying thoughts and on potential questions this had raised for them, possibly also when triangulating the results with information from other sources. Only three participants indicated that they had thought about the feedback after their initial reading and sensemaking. As revisiting the feedback reports and looking for additional data for triangulation at a later stage during interpretation indicate a conscious and systematic process of sensemaking, we labelled this as "rational framing." Only one participant had compared the school performance feedback data with results from their own classroom assessments. This means that only four of the 11 participants engaged in combined (intuitive and rational) framing during the interpretation phase. We labelled two of these participants' noticing process as "combined".

During the interviews, we asked participants to review the feedback reports, in an attempt to encourage purposeful, systematic, and rational thinking. However, for more than half of the participants, this did not yield any additional insights. We also note that, even after having been stimulated to adopt a more rational approach, some participants continued to have difficulties forming an understanding of the data and indicated that they were unable to make any inferences based on the data. A large number of participants had difficulties with correctly interpreting the population scores and correctly comparing them to the achievement of their own students. We note that for at least eight of the participants, their interpretation of the information at hand was severely compromised.

One participant realized during the interview that their own initial framing of the information in the tables contradicted the graphs that essentially conveyed the same information. They tried to correct this by rationally reflecting on their earlier intuitive framing.

> Yeah, well, in that case something is off, right [laughs]. ... From the table I would infer that we did not do well and from the graph I would infer that we did. ... Here you are not compared to the rest, but you are compared to the attainment targets and here you are compared to the other participants, right. ... That's how I interpret the difference. I don't know if that's correct [laughs]. – Participant 7

Interpreting: Intuitive and rational judgment

In our conceptual logic, framing is followed by judging. Our analyses of the interview data show that, with one exception, all participants initially only made fast and intuitive judgments about the school performance feedback data and based these judgments on individual perceptions and personal criteria.

> Among our students, we have tremendous diversity. Many different home situations and native languages. Our students are also not motivated... Most of the time, they stop processing the subject matter instantly when the bell rings. ... Also, most of them do not show up for refresher classes and re-sit tests. – Participant 9

Two participants indicated that they did not remember how they arrived at judgments during their initial reading of the feedback reports. Other participants' initial judgments appear to have been largely intuitive. Teachers almost exclusively took into account input and context factors: student characteristics on the one hand, more specifically study attitude and motivation, specific educational needs, language proficiency, and mathematical knowledge and skills, and on the other hand the timing of and practicalities associated with test administration. Potential causes for the results that they put forward intuitively included the absence of support for students with special educational needs, the fact that some domains had not or had only just recently been covered at the time of the test and the fact that only final answers were scored instead of taking into account the solution strategy.

During the interviews, we asked the participants to reflect further on their judgments. Several of the initially and intuitively mentioned causes were explored further during the rational judgment phase—but in this case by participants who said they had not yet thought about these elements during their intuitive judgment phase. Additional explanations for performing well or not well were also provided. These included school-specific organizational features such as the availability of tutoring hours for mathematics and external context factors such as insufficient parental support, students' excessive use of social media, and overloaded curricula. Process factors and curriculum-related factors were also mentioned, such as a lack of classroom management and strong collaboration within the department on the one hand, and on the other hand the difficulty or abstract nature of certain domains, automaticity development in primary education, and the low difficulty level of test questions for certain attainment targets.

Collective sensemaking of school performance feedback data

In this section we discuss how the teachers made sense of the school performance feedback data together with their colleagues. It is striking that all participants, with one exception, indicated that the feedback reports were only distributed to the 7th and 8th grade mathematics teachers (involved in the test administration) and not to other staff members. Moreover, for four of the 11 participants, this only happened in the run-up to this interview. In addition, all participants indicated that the school leader or (internal quality assurance) coordinator did not link any specific instructions or reflection questions to the distribution of the feedback reports, other than that the teachers were expected to go over the reports and discuss them within the department. However, this only occurred in one school and this discussion remained short and superficial in nature. Two participants from another school indicated that during a recent department meeting the "discussion" was limited to a short announcement that the results were poor. A few schools had taken spontaneous initiatives to be able to interpret the results better, such as participating in a workshop or consulting with teachers from other schools to discuss the data together. However, according to these teachers, this had contributed only minimally to a better understanding of the terminology used in the reports and of the tables and graphs. They did not feel it had led to an in-depth professional dialogue or any additional rational sensemaking of the school performance feedback data.

In all other schools, the feedback reports had not been discussed as a team at all. Participants attributed this to factors at the policy level, for instance to the fact that they had been obliged to participate and therefore did not feel the need to discuss the results with their colleagues, a lack of sufficiently explicit expectations from the school leader, changes in the school leadership, and absence of the school leader due to illness. They also referred to factors at the teacher level and to the results themselves. Almost all participants indicated that their school currently had other priorities. The fact that most schools achieve average results, at least in the participants' interpretation, does not encourage collective discussion either. Most participants presumed that there would have been a collective formal or informal discussion if their results had been very disappointing.

In summary, therefore, we can conclude that collective sensemaking did not occur in most schools. In those cases where it did, the participants indicated that there was hardly any interaction and no meaningful exchange of individual interpretations from different perspectives. As a result, no new insights were created, so we cannot speak of true collective sensemaking.

Factors influencing sensemaking processes

In order to answer the second research question, we probed participants' perceptions of factors that influence sensemaking of school performance feedback data. Since the feedback reports had only been discussed during a department meeting in one school, we will look at factors influencing sensemaking of student outcomes and data use in general, without distinguishing between the individual and the collective sensemaking process. Table 3 provides an overview of influencing factors as mentioned by the participants.

Table 3
Coding for Influencing Factors

Participant	Trust	Expertise		Inquiry-based attitude		Commitment	
		Participant	Department	Participant	Department	Participant	Department
1	yes	yes	yes	no	no	yes	yes
2	yes	no	no	no	yes	yes	no
3	no	/	yes	no	no	no	no
4	yes	yes	yes	yes	no	yes	no
5	yes	no	no	no	yes	no	no
6	yes	yes	yes	/	yes	no	yes
7	yes	yes	yes	yes	yes	yes	yes
8	yes	yes	yes	yes	yes	no	/
9	yes	yes	yes	yes	no	yes	no
10	yes	/	yes	yes	yes	yes	no
11	yes	no	no	no	yes	no	no

Note: / = insufficient data

Trust

With one exception, all participants indicated that individual opinions, thoughts, and feelings regarding student outcomes could be discussed freely, both formally and informally, with other team members at their schools. These participants sensed an atmosphere of openness and trust (in each other's abilities). Teachers felt broadly supported and felt there to be ample and open communication and advice with the aim to improve instructional practice and, consequently, student achievement. Moreover, in several departments, sensitive issues tended to be discussed personally with the colleague in question, which further strengthened teachers' sense of psychological safety. However, although there appeared to be a strong sense of trust overall, this hardly (if at all) resulted in collective sensemaking of the school performance feedback data.

Expertise

Just over half of the participants appeared to have sufficient personal expertise to interpret the school performance feedback data correctly or to be able to correct intuitive misinterpretations by way of an in-depth, step-by-step, rational approach. Some participants explicitly stated during the interview that they did not understand the concepts or visualizations in the reports. They attributed this to the form and content of the feedback reports and to their own limited statistical literacy, indicating their training did not prepare them for this. Even during the interviews, several participants misinterpreted concepts or visualizations. These misinterpretations caused them to (un) consciously form incorrect inferences, particularly regarding the school level results in the data.

Almost all participants indicated they were quite confident that there was sufficient expertise within their team to interpret the school performance feedback data correctly. They often linked this to formal training and stated that teachers with a master's degree are more familiar with handling data and interpreting statistical analyses. Remarkably, one teacher who held a master's degree indicated that they found the feedback reports difficult to process and that they were convinced that there was insufficient expertise within their department to interpret these data correctly.

> I've taken a master's myself and I think it's difficult too so never mind someone who's never dealt with data. Um, give this to four out of five of my colleagues and they won't understand a thing. ... They should make those reports a little simpler. – Participant 11

While a lack of expertise had a negative impact on the accuracy of interpretations, it did not appear to have a fundamental impact on the sensemaking process per se. Even when sufficient expertise is present, (collective) sensemaking remained (extremely) limited and superficial in nature.

Inquiry-based attitude

Five out of the 11 participants indicated that they reflected on the data individually, seeking suggestions on how to improve their instructional practice. However, three of them stated that, much to their regret, they were unable to derive from the data what adjustments or improvements would be necessary, which resulted in their self-reflection remaining only general and superficial in nature. They specifically felt the lack of a report per test item or an overview of those units that students fail most, which could have informed and fostered their sensemaking process. They would have also appreciated some hands-on tips for how to improve instructional practice and general conclusions for the school level.

> I would really add conclusions for the school itself. General conclusions. ... and what they need to work on because I can't find that anywhere. ... So that people can also get to work on something concrete, even if they don't understand the report, so they still have something. – Participant 11

In the one school where the feedback reports had been discussed, one participant indicated that no follow-up questions had been asked and that the data were not looked at from different perspectives. In contrast, their colleague from the same school stated that the data had been reflected upon very briefly during the first discussion of the reports. However, because of the positive interpretation of the results and because the team had other priorities, this had not led to any adjustments of instructional practices. Both participants did indicate spontaneously that they had gained more insight into the school performance feedback data thanks to the reflection questions asked during the interview, because this forced them, so to speak, to adopt an inquiry-based attitude and to question their own practice.

Finally, one participant indicated that, in general, their department lacked an inquiry-based attitude, especially when seeking potential instructional adjustments. This participant added that they personally regularly question their own practice—just not based on the data at hand, since the data were in line with their expectations.

Overall, our findings suggest that teachers' inquiry-based attitudes may be an important predictor of the depth of their individual and collective sensemaking of school performance feedback data. More specifically, we found that the lack of an inquiry-based attitude stopped teachers in their sensemaking efforts when moving from the phase of basic reading of the data to making more complex reflections.

Commitment

Without being prompted, five out of the 11 participants explicitly stated that they had initially only superficially read the feedback reports (because this interview was scheduled), had scanned through the introduction, and had mainly searched for a general trend and their overall position compared to other schools. One of these participants had not read the school feedback report at all and two other participants had not looked at the graphs prior to the interview. Some schools only took the test because the school leadership had required it.

> Some people don't care about these kinds of inquiries and are only focused on their teaching. ... And I think that if it was not imposed upon us—that we have to engage in this—that it would fade into the background for everyone. ... This is something that would then slide into my drawer of "I'll do it one day," but only when all my other work is finished. ... But, really, like studying documents and things like that, that's just where they lose me. – Participant 5

For all of these participants, this lack of commitment resulted in incomplete, often partially inaccurate, and superficial sensemaking of the school performance feedback data.

Only a few participants indicated that they had individually looked at the reports again afterwards, which had led to corrections of initial sensemaking or additional and more specific interpretations of the data. The participants from the school who had discussed the reports collectively indicated that during the team meeting, questions were asked about the content of the feedback reports and that everyone or nearly everyone participated, which they regarded as a sign of commitment. Four participants, two from the same school, indicated that their department was highly committed anyway when it comes to discussing student outcomes. Yet eight of the 11 participants mentioned to a greater or lesser extent that they felt little personal commitment or involvement regarding the data. They also expected this to be the case among their colleagues in the department. They attributed this, among other things, to the (as they interpreted them) fairly good results, to a lack of commitment in general, and to a lack of commitment regarding educational research in particular. This lack of commitment thus appears to be a fundamental explanatory factor for the lack of deep and accurate (collective) sensemaking.

Conclusion and Discussion

Our aim in this study was to investigate how Flemish teachers make sense of school performance feedback data from standardized tests, to describe the steps they do or do not take in this sensemaking process, and to understand how these steps take shape within the complex context of secondary education. In addition, we wanted to investigate the potential influence of a selected number of explanatory variables on the sensemaking process.

How do individual and collective sensemaking processes take shape?

In general, we can conclude that the mere availability of school performance feedback data from standardized tests does not automatically give rise to sensemaking of these data. Within the Flemish educational system, which gives schools great autonomy in shaping school policy and internal quality assurance, does not mandate central examinations or other forms of standardized testing in secondary education, and grants educators absolute authority over decisions regarding student educational progress, this is a striking finding. We would expect schools that voluntarily choose to take standardized tests, from a school development perspective, to also intensively make use of this information. However, this study shows that hardly any collective sensemaking occurs. None of the school teams involved in the study engaged in in-depth professional dialogue to make sense of the reported outcomes. There is also food for thought in the observation that almost half of the participating teachers were only sent the feedback reports because they would be interviewed about it. Moreover, in all schools, communication from the school leadership had been limited to a request to discuss the feedback reports, without there being any specific expectations tied to this request and without any initiatives to manage, support, or follow-up the sensemaking process.

When we take a closer look at the different steps of the sensemaking process and consider the relationship between intuition and rationality, our findings are consistent with those of prior studies (Datnow et al., 2012; Vanlommel et al., 2017, 2018, 2019). Teachers rely heavily on their intuition during the sensemaking process. What is more, during initial sensemaking, none of the participants employed a purely rational approach to make sense of the school performance feedback data. Their selective attention was focused on things they recognized and understood. According to Chudnoff (2019) and Sinclair (2010), this is not necessarily a problem, provided that teachers have sufficient intuitive expertise to interpret cues correctly. However, our findings that only one teacher triangulated the data with other sources, that the participants hardly reflected on the test results, and that concepts and visualizations were often misunderstood or misinterpreted, called this idea of intuitive expertise into question. In the course of this study, it also turned out that teachers' first impressions often needed to be adjusted and, in some cases, were downright inaccurate compared to what could be objectively determined on the basis of the feedback report. Surprisingly, most teachers were not even aware of this, and it ultimately led to inaccurate or questionable framing. Furthermore, at no point during the judgment phase was any reference made to other data sources in order to support initial and often also intuitive judgments, and attributions for the schools' results were exclusively external.

Since all teachers indicated that they had only superficially read the feedback reports and had looked for general trends and things they intuitively recognized, we may wonder whether their selective attention was unconsciously steered by specific assumptions and feelings about students and class groups. This would be in line with earlier empirical findings in research on sensemaking in educational settings (Bertrand & Marsh, 2015; Vanlommel et al., 2017, 2019) and with research stating that intuition is not subordinate but complementary to rationality (Calabretta et al., 2017). We can probably also speak of confirmation bias here (Kahneman & Klein, 2009; Vanlommel et al., 2017).

However, in line with Mandinach & Gummer (2016), we argue that sensemaking is an essential part of purposeful and efficient data-driven decision-making processes. If rational data use is indeed this scarce during all the steps taken in initial individual sensemaking, and if individual intuitive expertise is not used to achieve a dynamic interaction process of co-creation and collective sensemaking, the resulting decision-making processes and actions will not adequately address existing gaps and will not have the intended or desired effect on the schools' internal quality assurance (Schildkamp, 2019) and decisions about individual students' learning processes (Vanlommel et al., 2017).

Which factors influence sensemaking?

To answer the second research question, we investigated the impact of trust, expertise, inquiry-based attitude, and commitment on sensemaking of school performance feedback data from standardized tests. We have no indication that a lack of trust is related to the lack of collective sensemaking in schools with respect to school performance feedback data. With one exception, all teachers sensed an atmosphere of openness and trust in their schools and felt their school culture was based on collaboration and knowledge sharing in a safe climate. Likewise, according to our findings, teacher expertise does not appear to have a fundamental impact on the occurrence of sensemaking. In every secondary school, at least some team members will have the necessary and appropriate expertise to correctly interpret the data, which is sufficient according to Krüger (2010). Our findings also show that even in schools where sufficient expertise was present, this had little impact on the depth and rational nature of sensemaking or on the time spent on sensemaking.

A lack of inquiry-based attitude and particularly a lack of commitment within school teams do, however, appear to impede in-depth individual and collective sensemaking of school performance feedback data. A lack of commitment can be explained to a certain extent by teachers' lack of interest in (participating in) research in general. Our study suggests that another and even more important explanatory factor is teachers' feeling that the data presented in the reports are not relevant to their day-to-day instructional practice. Consequently, most teachers did not feel the need to adopt an inquiry-based attitude, and sensemaking remained limited to a general, superficial, and largely intuitive interpretation of the data. A lack of guidance and clear expectations from the school leadership also appeared to contribute to the lack of commitment we found. Since the school performance feedback is to be used for internal quality assurance, and since mathematics instruction is characterized by a cyclic approach and iterations throughout the different grades, we would have expected that the feedback reports would at least have been distributed to all teachers in the mathematics department. However, this was not the case, leaving teachers with insufficient information to engage in in-depth professional dialogue as a team. These findings are consistent with previous research regarding the link between relevance and instrumental data use among teachers on the one hand and the crucial role of the school leader on the other (Jimerson, 2014; Van Gasse et al., 2015). Finally, lack of commitment to adopting an inquiry-based attitude may explain some teachers' misunderstandings and misinterpretations. We found that teachers who had barely read the introductory sections of the feedback reports, for instance, misinterpreted central concepts and visualizations.

Discussion

In order to increase our understanding of and insight into the phase of sensemaking within the cycle of data use, we unraveled processes of sensemaking of school performance feedback data by secondary school teachers and integrated notions of intuition and rationality into the different steps they undertook in this process. In doing so, we addressed a knowledge gap, since sensemaking of school performance feedback data had not been previously studied in-depth (Schildkamp, 2019). Thus, one of the main scientific contributions of this study is the way we interpreted and conceptualized teachers' intuitive and rational sensemaking processes when making sense of data from standardized testing.

The distinctions we make between intuitive and rational processes pertain to timing, pace, depth, thought processes, and options. These are always dependent on the context in which the teacher operates (Coburn & Tunner, 2011; Abrams et al., 2020). When the teachers in our study intuitively made sense of school performance feedback data, this happened immediately during the initial reading of the feedback reports. Teachers quickly and superficially make sense of the data by looking only at the big picture, at what they recognize, expect, and understand. They are unaware of alternative explanations, and unequivocally choose one particular interpretation of the results. In contrast, rational sensemaking happens over time, when teachers revisit the feedback reports and think about them more deeply, possibly led by guiding questions. This thought process is slower and deliberate. By way of analytical and step-by-step reasoning, they broaden and deepen the sense they previously made of the data and consciously make a choice from among several alternative interpretations.

From a methodological point of view, this study was also an attempt to investigate an unexplored area in the literature. Since there was no ready-made theoretical framework available and we wanted to gain in-depth insight into sensemaking processes of school performance feedback data from standardized tests, we focused specifically on aspects of intuition and rationality throughout the different steps of the sensemaking process. Brock (2015) as well as Dane and Pratt (2007) argued that measuring intuition and even describing what happens during intuitive processes is complex because these processes are fast and often unconscious. We addressed this issue by having participants go through the different steps identified in the theoretical framework via the think-aloud method during the interviews. Nevertheless, our findings could be further enriched with observations. By observing teachers and departments and by discussing those observations with individual teachers, the validity of our findings regarding sensemaking of school performance feedback data from standardized tests can be further enhanced. Observations offer the opportunity to map the iterative and non-linear nature of the different steps, as well as the way intuition and rationality are intertwined, as described in research by Kahneman and Frederick (2005) and suggested by our own findings. In a later phase, quantitative research can also be carried out on a large scale in order to investigate the generalizability of our findings. In addition, we propose that future research should investigate how the different steps of the sensemaking process influence each other.

References

Abrams, L. M., Varier, D., & Mehdi, T. (2020). The intersection of school context and teachers' data use practice: Implications for an integrated approach to capacity building. *Studies in Educational Evaluation*, *69*, 1–13 . https://doi.org/10.1016/j.stueduc.2020.100868

Alase, A. (2017). The Interpretative Phenomenological Analysis (IPA): A guide to a good qualitative research approach. *International Journal of Education and Literacy Studies*, *5*(2), 9–19. https://doi.org/10.7575/aiac.ijels.v.5n.2p.9

Amels, J., Krüger, M., Suhre, C., & van Veen, K. (2019). Impact of inquiry-based working on the capacity to change in primary education. *Journal of Educational Change*, *20*(3), 351–374. https://doi.org/10.1007/s10833-019-09337-3

Amidu, A. R., Boyd, D., & Gobet, F. (2019). A study of the interplay between intuition and rationality in valuation decision making. *Journal of Property Research*, *36*(4), 387–418. https://doi.org/10.1080/09599916.2019.1687572

Attfield, S., & Baber, C. (2017, June 20). *Elaborating the frames of data-frame theory*. 13th International Conference on Naturalistic Decision Making, Bath, United Kingdom.

Attfield, S., Fields, B., & Baber, C. (2018). A resources model for distributed sensemaking. *Cognition, Technology and Work*, *20*(4), 651–664. https://doi.org/10.1007/s10111-018-0529-4

Avby, G. (2015). Professional practice as processes of muddling through: A study of learning and sense making in social work. *Vocations and Learning*, *8*(1), 95–113. https://doi.org/10.1007/s12186-014-9123-y

Beck, J. S., & Nunnaley, D. (2021). A continuum of data literacy for teaching. *Studies in Educational Evaluation*, *69*, 1–8. https://doi.org/10.1016/j.stueduc.2020.100871

Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal*, *52*(5), 861–893. https://doi.org/10.3102/0002831215599251

Bolhuis, E., Schildkamp, K., & Voogt, J. (2016). Data-based decision making in teams: Enablers and barriers. *Educational Research and Evaluation*, *22*(3–4), 213–233. https://doi.org/10.1080/13803611.2016.1247728

Brock, R. (2015). Intuition and insight: Two concepts that illuminate the tacit in science education. *Studies in Science Education*, *51*(2), 127–167. https://doi.org/10.1080/03057267.2015.1049843

Calabretta, G., Gemser, G., & Wijnberg, N. M. (2017). The interplay between intuition and rationality in strategic decision making: A paradox perspective. *Organization Studies*, *38*(3–4), 365–401. https://doi.org/10.1177/0170840616655483

Cameron, M., & Lovett, S. (2015). Sustaining the commitment and realising the potential of highly promising teachers. *Teachers and Teaching: Theory and Practice*, *21*(2), 150–163. https://doi.org/10.1080/13540602.2014.928132

Cecez-Kecmanovic, D. (2004). A sensemaking model of knowledge in organisations: A way of understanding knowledge management and the role of information technologies. *Knowledge Management Research & Practice*, *2*(3), 155–168. https://doi.org/10.1057/palgrave.kmrp.8500040

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Christman, J. B., Ebby, C., & Edmunds, K. (2016). Data use practices for improved mathematics teaching and learning: The importance of productive dissonance and recurring feedback cycles. *Teachers College Record*, *118*(11), 1–32.

Chudnoff, E. (2019). In search of intuition. *Australasian Journal of Philosophy*, *98*(3), 465–480. https://doi.org/10.1080/00048402.2019.1658121

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Cohen, L., Martin, S., McCulloch, G., O'Sullivan, C., Manion, L., Morrison, K., & Bell, R. (2011). *Research methods in education* (7th ed.). Routledge.

Cook, L., & Gregory, M. (2019). Making sense of sensemaking: Conceptualising how child and family social workers process assessment information. *Child Care in Practice*, *26*(2), 182–195. https://doi.org/10.1080/13575279.2019.1685458

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of Management Review*, *32*(1), 33–54. https://doi.org/10.5465/AMR.2007.23463682

Datnow, A., & Park, V. (2018). Opening or closing doors for students? Equity and data use in schools. *Journal of Educational Change*, *19*(2), 131–152. https://doi.org/10.1007/s10833-018-9323-6

Datnow, A., Park, V., & Kennedy-Lewis, B. (2012). High school teachers' use of data to inform instruction. *Journal of Education for Students Placed at Risk*, *17*(4), 247–265. https://doi.org/10.1080/10824669.2012.718944

Day, C., & Gu, Q. (2007). Variations in the conditions for teachers' professional learning and development: Sustaining commitment and effectiveness over a career. *Oxford Review of Education*, *33*(4), 423–443. https://doi.org/10.1080/03054980701450746

Donche, V. (2015). *Inleiding in methoden en technieken. Een open leerpakket [Introduction to methods and techniques. An open learning package]*. University of Antwerp.

Eccles, D. W., & Arsal, G. (2017). The think aloud method: What is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531. https://doi.org/10.1080/2159676X.2017.1331501

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

Fransson, G., & Frelin, A. (2016). Highly committed teachers: What makes them tick? A study of sustained commitment. *Teachers and Teaching: Theory and Practice*, *22*(8), 896–912. https://doi.org/10.1080/13540602.2016.1201469

Gephart, R. P., Topal, C., & Zhang, Z. (2010). Future-oriented sensemaking: Temporalities and institutional legitimation. In T. Hernes & S. Maitlis (Eds.), *Process, sensemaking and organizing* (pp. 275–312). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199594566.003.0013

Gergen, K. J., Gergen, M. M., & Barrett, F. J. (2004). Dialogue: Life and death of the organization. In D. Grant, C. Oswick, & L. Putnam (Eds.), *The Sage handbook of organizational discourse* (pp. 39–60). Sage. https://doi.org/10.4135/9781848608122.n2

Hill, R. C., & Levenhagen, M. (1995). Metaphors and mental models: Sensemaking and sensegiving in innovative and entrepreneurial activities. *Journal of Management*, *21*(6), 1057–1074. https://doi.org/10.1177/014920639502100603

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, *47*(1), 181–217. https://doi.org/10.3102/0002831209345158

Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation*, *42*, 5–14. https://doi.org/10.1016/j.stueduc.2013.10.010

Kahneman, D. (2011). *Thinking, fast and slow*. Business Contact.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & B. Morrison (Eds.). *Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making* (pp. 113–155). Lawrence Erlbaum Associates.

Krüger, M. L. (2018). *Onderzoeksmatig leidinggeven: Handreiking voor schoolleiders en bestuurders [Inquiry-based leadership: Guide for school leaders and administrators]*. ACCO.

Krüger, M. L. (2010). *De invloed van schoolleiderschap op het onderzoeksmatig handelen van leraren in veranderingsprocessen [The impact of school leadership on teachers' inquiry-based work in change processes].* Hogeschool van Amsterdam.

Kudesia, R. S. (2017). Organizational sensemaking. In *Oxford Research Encyclopedia of Psychology* (pp. 1–42). Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.78

Lai, M. K., Wilson, A., McNaughton, S., & Hsiao, S. (2014). Improving achievement in secondary schools: Impact of a literacy project on reading comprehension and secondary school qualifications. *Reading Research Quarterly*, *49*(3), 305–334. https://doi.org/10.1002/rrq.73

Maitlis, S. (2005). The social processes of organizational sensemaking. *Academy of Management Journal*, *48*(1), 21–49. https://doi.org/10.5465/AMJ.2005.15993111

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *The Academy of Management Annals*, *8*(1), 57–125. https://doi.org/10.1080/19416520.2014.873177

Maitlis, S., Vogus, T. J., & Lawrence, T. B. (2013). Sensemaking and emotion in organizations. *Organizational Psychology Review*, *3*(3), 222–247. https://doi.org/10.1177/2041386613489062

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, *69*, 1–10. https://doi.org/10.1016/j.stueduc.2020.100842

Mausethagen, S., Prøitz, T. S., & Skedsmo, G. (2019). School leadership in data use practices: Collegial and consensus-oriented. *Educational Research*, *61*(1), 70–86. https://doi.org/10.1080/00131881.2018.1561201

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company.* Oxford University Press.

Prenger, R., & Schildkamp, K. (2018). Data-based decision making for teacher and student learning: A psychological perspective on the role of the teacher. *Educational Psychology*, *38*(6), 734–752. https://doi.org/10.1080/01443410.2018.1426834

Roesch-Marsh, A. (2018). Professional relationships and decision making in social work: Lessons from a Scottish case study of secure accommodation decision making. *Qualitative Social Work*, *17*(3), 405–422. https://doi.org/10.1177/1473325016680285

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management*, *36*(4), 941–973. https://doi.org/10.1177/0149206309350084

Sammons, P., Day, C., Kington, A., Gu, Q., Stobart, G., & Smees, R. (2007). Exploring variations in teachers' work, lives and their effects on pupils: Key findings and implications from a longitudinal mixed-method study. *British Educational Research Journal*, *33*(5), 681–701. https://doi.org/10.1080/01411920701582264

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, *61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Schildkamp, K., Lai, M. K., & Earl, L. (Eds.) (2013). *Data-based decision making in education. Challenges and Opportunities.* Springer.

Schildkamp, K., Smit, M., & Blossing, U. (2019). Professional development in the use of data: From data to knowledge in data teams. *Scandinavian Journal of Educational Research*, *63*(3), 393–411. https://doi.org/10.1080/00313831.2017.1376350

Schildkamp, K., & Teddlie, C. (2008). School performance feedback systems in the USA and in The Netherlands: A comparison. *Educational Research and Evaluation*, *14*(3), 255–282. https://doi.org/10.1080/13803610802048874

Shleifer, A. (2012). Psychologists at the gate: A review of Daniel Kahneman's Thinking, Fast and Slow. *Journal of Economic Literature*, *50*(4), 1080-91. http://doi.org/10.1257/jel.50.4.1080

Sinclair, M. (2010). Misconceptions about intuition. *Psychological Inquiry*, *21*(4), 378–386. https://doi.org/10.1080/1047840X.2010.523874

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*(2), 113–141. https://doi.org/10.1086/663283

Spillane, J. P., & Miele, D. B. (2007). Evidence in practice: A framing of the terrain. *Teachers College Record,* 109(13), 46-73.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431. https://doi.org/10.3102/00346543072003387

Tsoukas, H. (2009). A dialogical approach to the creation of new knowledge in organizations. *Organization Science*, *20*(6), 941–957. https://doi.org/10.1287/orsc.1090.0435

Uiterwijk-Luijk, L., Krüger, M., Zijlstra, B., & Volman, M. (2017). The relationship between psychological factors and inquiry-based working by primary school teachers. *Educational Studies*, *43*(2), 147–164. https://doi.org/10.1080/03055698.2016.1248901

Van Gasse, R., Vanhoof, J., Mahieu, P., & Van Petegem, P. (2015). *Informatiegebruik door schoolleiders en leerkrachten [Information use by school leaders and teachers].* Garant.

Van Gasse, R., Vanhoof, J., & de Vos, W. (2014). Informatiegebruik in verantwoordingsgerichte en schoolontwikkelingsgerichte onderwijssystemen: Een Nederlands-Vlaams perspectief [Information use in accountability oriented and school development oriented education systems: A Dutch-Flemish perspective]. *Pedagogiek*, *34*(2), 84–106.

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2017). The impact of collaboration on teachers' individual data use. *School Effectiveness and School Improvement*, *28*(3), 489–504. https://doi.org/10.1080/09243453.2017.1321555

Vanhoof, J., Verhaeghe, G., Verhaeghe, J. P., Valcke, M., & Van Petegem, P. (2011). The influence of competences and support on school performance feedback use. *Educational Studies*, *37*(2), 141–154. https://doi.org/10.1080/03055698.2010.482771

Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal*, *56*(3), 792–821. https://doi.org/10.3102/0002831218803891

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education*, *71*, 108–119. https://doi.org/10.1016/j.tate.2017.12.011

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, *83*, 75–83. https://doi.org/10.1016/j.ijer.2017.02.013

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

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

Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Whittaker, A. (2018). How do child-protection practitioners make decisions in real-life situations? Lessons from the psychology of decision-making. *British Journal of Social Work*, *48*(7), 1967–1984.

Zhang, P., & Soergel, D. (2016). Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, *42*(1), 59–78. https://doi.org/10.1177/0165551515615834

Zhang, P., & Soergel, D. (2019). Cognitive mechanisms in sensemaking: A qualitative user study. *Journal of the Association for Information Science and Technology*, *71*(2), 158–171. https://doi.org/10.1002/asi.24221

Corresponding authors
Gila Gutwirth
Department of Training and Education Sciences, Faculty of Social Sciences, University of Antwerp, Belgium
E-mail: g.gutwirth@yahoo.be

Evelyn Goffin
Department of Training and Education Sciences, Faculty of Social Sciences, University of Antwerp, Belgium
E-mail: evelyn.goffin@uantwerpen.be

Jan Vanhoof
Department of Training and Education Sciences, Faculty of Social Sciences, University of Antwerp, Belgium
E-mail: jan.vanhoof@uantwerpen.be

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-5

STUDENT GUIDANCE DECISIONS AT TEAM MEETINGS: DO TEACHERS USE DATA FOR RATIONAL DECISION MAKING?

ROOS VAN GASSE,
MARTINE MOL

Abstract

In the past decade, the belief has grown that student guidance decisions can benefit from systematic data use. Systematic data use can be considered as completing the circle of inquiry (from data discussion to interpretation, to analysis, diagnosis, and action) with a reasonable depth. However, little is known about how teachers use data to inform student guidance decisions. This qualitative study analyzed the field notes of 17 teachers' meetings that were intended to formulate student guidance decisions in secondary education. The results showed that data were used only sporadically and often not in a systematic way. Moreover, the depth of inquiry in formulating diagnoses on poor student functioning was low. These results indicate a need to raise awareness among teachers and policymakers on the stepwise and self-questioning process that data use should be in order to be effective.

Keywords

student guidance, data use, teacher meetings, secondary education

Introduction

Teachers make decisions every day, often related to student guidance. “What can I do to improve this student’s learning process?” is a question all teachers are familiar with. How teachers deal with this question is vital for student learning, but also for their learning trajectories and for what education can bring about for them.

This idea and the quest to close the gap for disadvantaged students in education initiated the conviction that teacher decisions on student guidance can benefit from an adequate use of data (i.e., all the information that can inform teachers about student functioning) (Wayman et al., 2013). Data use is a systematic process with a sequence of subphases around the discussion and interpretation of data, the definition and analysis of potential causes (i.e., diagnosing phase), and the formulation of improvement actions (Schildkamp et al., 2016; Van Gasse et al., 2017). The systematic use of data such as test scores or classroom observations is considered a means of preventing teachers from hasty decisions on pupil guidance because such use challenges cognitive biases and preconceptions. That the systematic use of data has the power to prevent teachers from such biases has been supported by empirical evidence. In the past decade, research has shown that adequately using data can result in better instructional decisions and eventually in increased student achievement (Campbell & Levin, 2008; Carlson et al., 2011).

Two aspects are vital to effective data use. The first aspect relates to the sequential activities needed to arrive at an inquiry cycle that challenges existing assumptions. The literature has distinguished phases of (1) *data discussion and interpretation* in which data is contextualized and transformed into information, (2) *diagnosis of potential causes* in which hypotheses are challenged and investigated, and (3) *formulating improvement actions* in which appropriate actions for the defined problems are designed. Research has shown that such inquiry cycles interrupt the human tendency to jump to conclusions without identifying causes based on data (Schildkamp et al., 2016). The phases provide guidance for teachers to thoughtfully investigate and reflect on classroom and school practices. Therefore, data use processes that have gone through the inquiry cycle can result in concrete improvements, such as improvements in the classroom (Van Gasse et al., 2017). However, the literature on data use reveals that teachers generally do not have the capability to systematically collect and use data appropriately (Datnow & Hubbard, 2016). Teachers experience problems with interpreting data or investigating potential causes, or with determining which improvement actions are appropriate within a certain situation (Bertrand & Marsh, 2015). As such, the data use cycle is often hampered and the full potential of data use cannot be reached.

Following the inquiry cycle for data use does not in itself ensure improvements based on data. Even if these sequences are correctly followed, great differences remain regarding the depth of investigation that is achieved across different teacher teams (Schildkamp et al., 2016). The reflective stance that teachers take in creating knowledge based on data can in turn produce a range of different knowledge outcomes and is thus critical in this regard (Hubers et al., 2016). Data users who do not question their own role in defining problems or causes will not reach the depth of investigation needed to achieve far-reaching improvements through data use. Therefore, the depth of inquiry is the second aspect that is crucial to effective data use (Schildkamp et al., 2016).

However, although knowledge is available on what effective data use practices look like, there is limited evidence on how teachers systematically use data to discuss student progress and achievement at formal team meetings and on whether data use reaches its full potential in this context. Generally, the literature outlines a rather pessimistic situation in data use (Van Gasse et al., 2017). How teachers use data for guiding students successfully through their trajectories has not yet been extensively investigated. Moreover, in-depth knowledge on how the decision processes in team meetings on this topic involve appropriate data use remains underexplored. The sequence of the data use cycle and the depth of investigation in data use practices are key. Nevertheless, integrating these perspectives for an in-depth examination of data use practices has been done in only one intervention-based study that was not specifically related to guiding student trajectories (i.e., Schildkamp et al., 2016). Therefore, when considering what effective decisions on student trajectories can produce in terms of what education can bring about for students, knowledge is needed on how teachers use data in formative team meetings on student progress and more specifically on how they run through the data use cycle and to which degree of examination depth. The following research questions guide this study:

1. To what extent do formal teacher teams incorporate the data use cycle?
2. What is the depth of inquiry of data use processes in formal teacher teams?

Context of this study

Educational decisions in this study are situated in the context of formal student guidance meetings in Flemish secondary schools (12- to 18-year-old students). These class-group level team meetings take place two or three times a year (i.e., once in September, once in December, and in lower grades once in April). In the first meeting at the beginning of the school year, teachers discuss the student dossiers. The second meeting (i.e., the one that was observed for this study) is for discussing student progress during the year. This meeting informs the team meeting at the end of the school year that serves to advise student trajectories. The teams under study are temporary and interdisciplinary (Vangrieken et al., 2013), and are collectively responsible for the learning of a group of students. In the guidance meetings, the teams discuss the student learning progress and functioning to improve student guidance. The meetings involve all teachers who teach a certain subject in the student group, supplemented by a student counsellor.

The study took place in Flanders. Flemish schools have a lot of freedom to design student guidance and the Flemish government does not collect specific data to support this (e.g., learning monitoring systems or central tests). Schools themselves are responsible for insight into whether they reach the Flemish standards at the end of secondary education (De Volder, 2012). Thus, governmental expectations toward data use are rather implicit and the responsibility for using data and support for data use lie with individual schools and teachers. Therefore, data is broadly conceptualized and includes all data related to student functioning. This data can be both qualitative (i.e., observations) and quantitative (i.e., class tests).

Theoretical Framework

Data use and data

Data use is not simply about data. It refers to a sequence of activities in which data are transformed into knowledge for making rational decisions (Coburn & Turner, 2011; Marsh, 2012). Therefore, data use is a less straightforward activity than it seems. It is a complex inquiry process in which current situations are fully analyzed and improved.

Effective data use is a reflective process that follows a certain sequence (Ciampa & Gallagher, 2016; Marsh & Farrell, 2015). This sequential process supports teachers in the translation from data into meaningful decisions (Marsh et al., 2015). The tendency to jump from data to improvement actions without in-depth consideration of potential causes and alternatives is interrupted by explicitly installing the different inquiry phases in data use (Hubers et al., 2017; Schildkamp et al., 2016). Therefore, the presence or absence of the phases is an indicator of the quality of the data use process and is essential to expand and refine the knowledge as to how teachers use data to decide on student guidance decisions.

The first step in the data use sequence is to define the educational problem. In this phase, teachers formulate the problem independent of data. Data use generally does not start with data, but with problems teachers encounter (Schildkamp et al., 2016). After collecting (or selecting) data, the subsequent phase of the data use cycle is discussing and interpreting data in relation to the educational problem. Interpreting data correctly transforms data (which are independently meaningless) into information. In the third step, the problem diagnosis, a deliberation of potential causes and explanations is carried out. This implies that knowledge is created based on the available information. The final phase involves the formulation of educational decisions (e.g., designing improvement actions that can be implemented in the classroom) (Verhaeghe et al., 2010).

The data use cycle may seem linear and straightforward. However, the literature has shown that data use cycles are often interrupted or that teachers return to previous phases (Marsh & Farrell, 2015; Schildkamp et al., 2016). Non-linear sequences in data use are generally not seen as problematic and are part of the inquiry process. Nevertheless, completing the full cycle has proven essential for solving the presented educational problems (Gelderblom et al., 2016). Despite this knowledge, apart from intervention studies very limited evidence is available on teachers going through the data use cycle at formal meetings.

Depth of analysis

Completing the full data use cycle is essential for the quality of data use in schools. However, running through all the phases of this sequence does not automatically imply data use trajectories of high quality. Also between similar teacher teams that follow the circle of inquiry, large differences occur in the quality of the inquiry processes and the associated results (Hubers et al., 2016; Schildkamp et al., 2016). Schildkamp et al. (2016) found differences between teams regarding the depth of inquiry throughout the circle of inquiry. And the research by Hubers et al. (2016) showed that identical sequences in data use can result in different knowledge creation in teams. Thus, the success of data use strongly depends on what happens during the different phases of the circle of inquiry.

Differences in how teams evolve during the data use cycle can be explained by user characteristics (e.g., self-efficacy, attitude), school characteristics (e.g., school culture), and context characteristics (e.g., accountability context) (Datnow & Hubbard, 2016; Ehren & Swanborn, 2012; Van Gasse et al., 2017). However, even between similar teams, schools, and (data use) contexts, large diversity in data use processes can occur (Schildkamp et al., 2016). Research has shown that the effectiveness of data use processes depends on the way in which new knowledge is created, or in the diagnosing phase. In this phase, it is necessary to combine different types of knowledge to define causes for the presented educational problem (Gummer & Mandinach, 2015). Therefore, this phase is crucial for the quality of data use and introduces great differentiation between data users.

Attribution and diagnosis

Research has shown that teachers generally formulate external causes for educational problems (e.g., problems related to student capacity or student home environment) (Schildkamp et al., 2016). It is usually only when these formulated causes turn out to be incorrect that teachers start searching for explanations that are related to the school or their own functioning and become able to solve the presented educational problem (Schildkamp et al., 2016). Therefore, the attribution of causes in the diagnosing phase can be considered as a potential indicator for the depth of the inquiry process in data use. Nevertheless, insights into this attribution of causes in data use processes are rather limited.

According to Weiner (2010), attribution theory states that formulated causes have a multidimensional character. The first is the causal locus. Causes can be attributed to internal factors (e.g., high competency can lead to success on a test) or external factors (e.g., success on a test can depend on its difficulty). The second dimension is the causal stability. Whereas one's competences can be considered a stable cause, the effort one invested in a task can be seen as unstable. The third dimension Weiner (2010) distinguished is causal control. Task difficulty cannot always be controlled; effort is controllable. Thus, formulated causes can be placed in a multidimensional space of causal locus, causal stability, and causal control.

The work of Schildkamp et al. (2016) is one of the first studies that used the idea of causal attribution to investigate the depth of inquiry in data use processes. In this regard, the study mainly focused on the causal locus in the diagnosing phase in data use in the sense that internal causes (e.g., teacher behavior) were distinguished from external causes (e.g., student prior knowledge level). In this study, we also examine the causal control that teachers perceive. Schildkamp et al. (2016) already noted that even when causes are externally attributed (e.g., student motivation), it is important that teachers reflect on their role in exercising control over it (e.g., how can I motivate my students?). Therefore, the perceived causal control not only comes to the surface in the discussion on causes (i.e., diagnosis), but also in the decisions on follow-up actions (i.e., action phase).

Attribution and formulating actions

Follow-up actions in data use can take different forms. Coburn and Turner (2011) distinguished four types of possible actions when it comes to data use to improve classroom practices. First, teachers may choose to adapt their instruction for (some) students. This implies that they change their behavior. However, this type of educational improvement is not self-evident. Research has provided some indications of changed teacher behavior (Ebbeler et al., 2017), but this appears to be less common in research out of the scope of intervention settings (Van Gasse et al., 2016). A second possible action is to provide students with supplemental materials on certain topics (Coburn & Turner, 2011). Examples can be additional or different exercises or supportive materials for certain lessons. Grouping students is a third strategy. When data show, for example, that some students need extra attention on some topics, teachers may choose to split the student group in smaller groups that go through the subject matter at their own manner or pace. The last action Coburn and Turner (2011) distinguished are actions on other dimensions of the classroom and school practices. Examples may include that the school is organized differently (e.g., more individualization of learning trajectories) as a result of data use practices.

Following attribution theory, it is likely that the concrete actions that are formulated after a data use sequence are related to the causal locus and the causal control that are perceived by teachers. It is, for example, less likely that teachers will change their instruction when they perceive that poor student results are due to the fact that the student chose a wrong educational track for their cognitive capacities (i.e., external locus – no perceived control) (Schildkamp et al., 2016). When teachers are convinced that they can exercise control over the situation, other actions may be formulated (Schildkamp et al., 2016). For example, when teachers are convinced that a student's poor results are due to a lack of motivation (external locus), and they are convinced that they can affect this motivation (causal control), a potential data use outcome may be that teachers agree on making changes in their instruction to motivate students (Baten et al., 2020). However, when teachers consider their students' motivation as an uncontrollable factor, this type of action will not be formulated (Weiner, 2010). Therefore, the depth of inquiry through causal attribution comes to surface in the perceived causes and it is also reflected in the actions that are formulated.

Method

Participants, design, and instrument

This qualitative study used observations to examine data use in formal team meetings in Flemish secondary schools. Data were collected within two secondary schools (ISCED 2 and 3) in Flanders. The two schools participated voluntarily. Within the schools, all team meetings were observed that (1) took place on two meeting days, (2) did not show overlap with other meetings in the meeting schedule, and (3) included teachers providing consent to the observation. In total, 17 team meetings of 17 classes were observed. The number of students in each class ranged between 3 and 18 with a median of 8. Across the team meetings, we collected data for 149 student guidance processes.

The teams were responsible for the learning of students (14- to 18-year-old students) in technical and vocational tracks. The team meetings took place among all teachers of the class (generally about 11, teaching different subjects to the class) and the student counsel. The meetings served in discussing the progress of students during the school year and for making appointments on student guidance decisions. In the sample schools, three minutes of discussion per student was scheduled. Both achievement and social-emotional behavior are possible subjects for discussion.

For the observations, the two researchers collaboratively developed an observation sheet based on the theory of data use. In this sheet, the different phases of the circle of inquiry were included: *discussion and interpretation*, *diagnosis*, and *action*. The observation process was twofold. For every student out of the student group was an indication of whether the different phases in the sequence were followed (by adding a 1 or a 0). Open observational notes were added, including which data were analyzed, what causes were formulated, etc. In this way, the observations combined a quantitative and qualitative observation approach. All 17 formal team meetings were observed systematically following this observation scheme by one independent researcher (the second author).

Analyses

The extent to which formal teams incorporate the data use cycle

To answer the first research question, the binary data of the observation sheets were analyzed. In a first step, the total number of students for whom data-based *discussion and interpretation*, *diagnosis*, and *action* took place at the formal team meetings was calculated per team meeting. Subsequently, the ratio of this number and the total student population that was subject of the team meeting was calculated. This resulted in an occurrence ratio of data *discussion and interpretation*, *diagnosis*, and *action* for each team meeting. As such, we investigated the extent to which formal teams incorporated the data use circle of inquiry in student guidance discussions.

Additionally, our analyses of the data use cycle took an in-depth approach. Based on the binary data of the coding scheme, we examined the extent to which the formal teams completed the data use sequence (i.e., formulating actions after discussion and interpretation and diagnosis). To this end, we counted the “complete” processes per team meeting. Subsequently, we investigated why some data use cycles were incomplete. We looked at the number of sequences that stopped or were not implemented as prescribed by theory (e.g., skipping the diagnosing phase).

The depth of inquiry in data use

To analyze the depth of inquiry, we analyzed the researcher's open observation notes in the observation schemes. In line with the theoretical framework, we focused on the diagnosis and action phases for this analysis. First, we coded the causes that were discussed in the formal team meetings and the actions that were decided on in different axial codes that were discussed by the two researchers. These codes were then clustered into the presented theoretical framework (see Table 1).

Table 1
Coding tree

	Cluster	Axial codes
Diagnosing phase	Internal locus – uncontrollable	–
	External locus – uncontrollable	Learning disorder Problematic home situation Medical problem
	External locus – controllable	Subject-specific learning problem Learning attitude problem Language problem Emotional problem
	Internal locus – controllable	–
Action phase	Adapting instruction	Adapted student guidance in class Adapted learning trajectory Catch-up lessons
	Materials	Remedial assignment Holiday assignment Extra test Extra exercises
	Grouping pupils	–
	Other dimensions	Follow-up meetings (parents / students) Behavioral contract

In a next step, the axial codes were clustered. For the diagnosing phase, the clustering was based on the theory of causal attribution. Each axial code was assessed on the dimensions *causal locus* and *causal control*. For the action phase, the clustering was based on the different actions that were distinguished by Coburn and Turner (2011): adapting instruction, adapting materials, grouping pupils, and other dimensions of the classroom and school practice.

This coding process resulted in an in-depth cross-case analysis of the diagnosing and action phases of the different formal teams.

Results

The data use cycle in teacher teams

Table 2 shows the extent to which the different phases in problem diagnosis were carried out within the different team meetings. The table provides the raw number of pupils for whom data were discussed and interpreted, for whom data were searched for causes (diagnosis), and for whom actions were formulated. The ratio of students subject to a certain phase of data use and the total number of students are expressed in percentages.

Table 2
Data use at formal student guidance meetings

Team	N students in class?	Discussion and interpretation (%)	Diagnosis (%)	Action (%)
1	8	3 (37.5)	0 (0)	1 (12.5)
2	9	5 (55.6)	5 (55.6)	5 (55.6)
3	10	2 (20.0)	1 (10.0)	1 (10.0)
4	5	2 (40.0)	1 (20.0)	2 (40.0)
5	8	7 (87.5)	5 (62.5)	4 (50.0)
6	8	1 (12.5)	1 (12.5)	1 (12.5)
7	6	4 (66.7)	4 (66.7)	1 (16.7)
8	5	5 (100)	4 (80.0)	2 (40.0)
9	8	0 (0)	0 (0)	0 (0)
10	4	4 (100)	3 (75.0)	3 (75.0)
11	16	4 (25.0)	3 (18.8)	3 (18.8)
12	10	6 (60.0)	5 (50.0)	6 (60.0)
13	18	13 (72.2)	6 (33.3)	13 (72.2)
14	5	2 (40.0)	2 (40.0)	0 (0)
15	5	4 (80.0)	4 (80.0)	4 (80.0)
16	9	4 (44.4)	4 (44.4)	4 (44.4)
17	5	1 (20.0)	0 (0)	1 (20.0)
Average		50.7%	38.2%	35.7%

On average, data were discussed and interpreted for about half of the students (51%). This means that for the other half of the students, no data were used across the meetings. The observations indicated that this share of students was generally not even the subject of discussion because teachers did not formulate a clear problem or situation sketch. In other words, their learning processes were not discussed at the team meetings. Diagnosing causes underlying the data happened for about 38% of the pupils who were discussed, whereas concrete actions were formulated for about 36% of the pupils. This implies that data discussion and interpretation do not end in concrete guidance actions for all students.

It is necessary to emphasize that there are great differences across the team meetings in the extent to which data are used for student guidance. As Table 2 shows, in one team (Team 9)no data discussion took place. In every other team, data discussions were initiated. However, the extent of pupils who were the subject of data discussion varied a lot across teams. The two classes in which the teams discussed data for all the students were small (4 and 5 students). Class size is not obviously related to the extent of data discussion; for some larger class groups, a high number of data discussions were initiated (e.g., Team 13). On the other hand, the extent of data use for some smaller pupil groups was rather limited (e.g., Team 6 and Team 17).

Table 3
Sequential interruptions in data use

	N	**Percentage**
Completed inquiry circles	38	56.7
Only discussion and interpretation	8	11.9
Discussion, interpretation, and diagnosis	10	14.9
No diagnosis	11	16.4
Total processes	**67**	

The observations indicated that teachers in the team meetings do generally not use data in a systematic way. Table 3 provides an overview of the interruptions in the data use sequence of the 67 student guidance processes in which data use was initiated (i.e., in which data were discussed). In 38 of those processes (57%), the data use sequence was completed. The 29 other data use processes (43%) were incomplete or showed imperfections from a theoretical perspective. Incomplete data use sequences stopped after data discussion (N = 8) or after data diagnosis (N = 10). This implies that student data (e.g., subject scores) were only discussed (in 8 cases) or that teachers talked about potential causes (in 10 cases) but that no discussion on appropriate actions took place. Next to this share of incomplete data use processes, there was a reasonable share of data use sequences in which the diagnosing phase was skipped (in 11 cases). In these processes, teachers formulated actions without deliberation on potential causes for the student's problem. Whereas the incomplete data use sequences came to the front in different teams, the data show that a large number of processes that skipped the diagnosing phase occurred within one team.

The Depth of Inquiry in Data Use

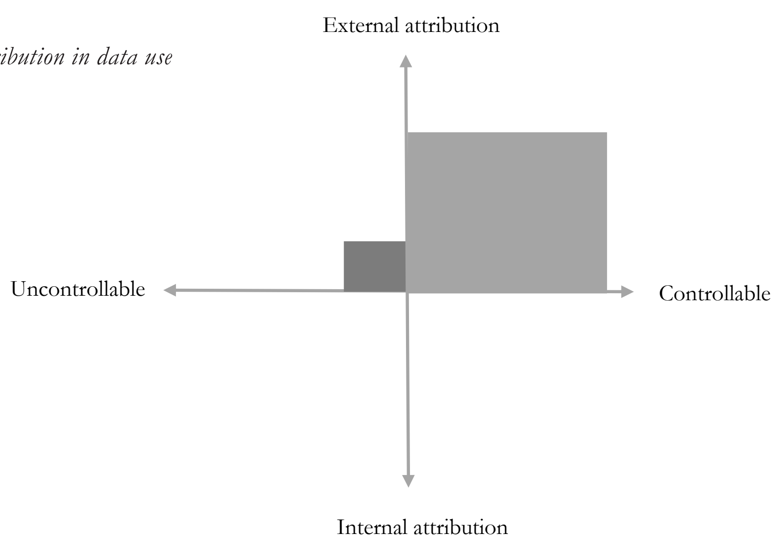
The diagnosing phase

The analysis regarding the depth of inquiry in the team meetings focused on the concept of *attribution*. Figure 1 presents a visualization of the number of coded fragments in the diagnosing phases. The figure makes clear that the vast number of discussed causes among the teachers were external and controllable. All causes were externally attributed. None of the discussed causes related to the teachers or their teaching. Teachers all pointed at student "problems" as causes for, for example, underachievement.

Within these external causes, the majority of the causes that were raised were controllable by teachers. Learning problems within certain subjects (e.g., math problems) and problematic learning attitudes were mentioned the most (in 52 of the 58 coded fragments). But next to that, emotional problems (in 5 coded fragments) and language issues (in 1 coded fragment) came to the surface. There were no clear differences in diagnostic processes based on the presented topic.

Some of the external causes that teachers perceived were uncontrollable. Most of the time these causes related to student characteristics or student learning environments. Student learning disorders (e.g., autism or ADHD) and their medical hindrances (e.g., chronic headaches) were the most important causes mentioned in this category (in 5 and 4 coded fragments respectively). The problematic home environments of students also came to the surface (in 2 coded fragments). Generally, the externally attributed causes that were uncontrollable by teachers were less present in the observational data.

Figure 1
Causal attribution in data use



The formulated actions

The formulated actions were distributed into four categories, based on the different actions that were distinguished by Coburn and Turner (2011): adapting instruction, materials, grouping pupils, and other dimensions of the classroom and school practice.

First, the most common actions that were formulated across the team meetings related to providing students with supplemental materials. This strategy came to the front in 40 of the 61 coded fragments on formulating actions. Teachers mainly suggested remedial assignments (in 15 of the coded fragments) and additional assignments during holidays (in 15 of the coded fragments). But next to that, additional exercises and additional tests were formulated as actions during the team meetings (in 8 and 2 coded fragments respectively).

Besides providing students with materials, teachers suggested some other actions in the team meetings. This *other* category appeared to be the second biggest in the data set (11 of 61 coded fragments). Mainly when it came to problematic student behaviors, alternative actions were suggested. Examples included follow-up meetings with students and parents (in 5 and 3 coded fragments respectively); in some cases there was agreement on behavior contracts with students (i.e., in 3 coded fragments).

Adapting instruction came to the front in the observational data but was not a common action after the data use processes. Of the coded fragments, 10 related to this category. For some students, teachers agreed upon adapted guidance in class (in 6 of the coded fragments). In two cases, the team decided on an adapted school trajectory (e.g., a combination of school and work); in two other cases, the teachers agreed to provide them with additional lessons.

The analysis of the observation notes made clear that subgrouping pupils was not a strategy in the different team meetings. This strategy was not suggested in any of the 17 meetings.

Discussion

What education can bring about for students depends on how their learning processes are guided and therefore on the educational decisions teachers make. Systematic data use offers the potential for teachers to improve student guidance. In this regard, it is essential that data use follows a sequence of discussion and interpretation, diagnosis, and action. It is also vital that teachers take a self-reflective stance in the examination of causes and formulation of improvement actions (Schildkamp et al., 2016). This study investigated teacher use of data at formal student guidance meetings. The aim of this in-depth qualitative study was twofold. The observations of 17 formal student guidance meetings in Flanders provided insight into the extent to which the data use circle of inquiry was followed and the depth of inquiry in these data use processes.

A first finding was that of all 149 students that were the subject of discussion at the team meetings, only about half of them (51%) were discussed at the team meetings in a way that was supported by data. The main reason for teachers to start discussing and interpreting data was the consideration that there was a problem with student functioning. Concerning the other half of the students, who did not show clear problems, there was almost no discussion. Teachers only discussed students with obvious learning problems; they did not discuss, for example, talented students who needed more challenges or deepening in the course materials. Given that the observed team meetings are the only formal occasions for the teams to make arrangements on student learning trajectories, this finding implies that for about half of the students, data-based reflection on student guidance among teachers was missing. A high responsibility remains with individual teachers to properly detect and follow up needs for gifted students. This may also have consequences for the advisory function of the team toward students for their study progress. The question arises of whether the team meetings can result in clear and data-informed picture of all student capacities.

Two elements are important to highlight to explain these findings. The fact that teachers all teach different subjects may imply that they consider their teaching task as an individual practice and responsibility. In this perception, teachers may not raise issues they consider as specific for their teaching. As a result, only problems that exceed individual teaching practices come to the front at formal team meetings. Such a lack of connectedness for teaching and learning in interdisciplinary teams is not uncommon in educational research, and in data use research in particular (Van Gasse et al., 2017). The downfall, however, can be that teachers do not have the full picture of student functioning because teachers do not share information when none of the teachers have the feeling that there are serious problems. As such, some issues in regular or gifted student guidance might be overlooked in the early stages and no optimal data-based guidance is provided.

The second explanation can be that teachers often only initiate data use processes when problems are perceived (i.e., problem-based data use) (Ansyari et al., 2020). Such data use following intuition is effective in installing data use practices when teachers are not familiar with it. Working on problems that teachers recognize and acknowledge can be a stimulus for using data in schools (Schildkamp et al., 2016). However, when data use only follows the "intuitive" problems of teachers, student guidance processes still strongly depend on teachers' pre-existing assumptions. Again, students will only be "guided" when teachers (or parents, student counsellors, or students themselves) perceive problems, not when students do not differentiate in behavior or achievement from the modal student. Therefore, relying only on problem-based data use cannot avoid and may even facilitate some of the cognitive biases (e.g., attention bias) that data use is expected to counter. There is also a risk that teachers' advice for student trajectories that follows later on is based on intuition rather than on knowledge that is built within the team meetings.

A second finding in this study is that slightly more than half (i.e., 57%) of the observed student guidance discussions in which data were discussed and interpreted followed the subsequent phases of the data use circle of inquiry. Our analysis showed a considerable number of data use processes that stopped after the discussion and interpretation or diagnosing phase. Further, a reasonable share of discussions were observed in which the diagnosing phase was skipped. In other words, in these student guidance processes teachers did not search for possible causes before coming up with improvement actions. The need for diagnosis lies in challenging existing (intuitive) assumptions. Although some experienced teachers may choose appropriate improvement actions based on their expertise, skipping this phase results in a higher risk for cognitive biases in decision making (e.g., confirmation bias or attention bias). Bearing in mind that data was used in only half of the student guidance processes, the data use cycle was completed in about a quarter of these processes.

Interruptions in the data use cycle and skipping phases have also been observed in other studies in data use (e.g., Marsh & Farrell, 2015). Because teachers in this study were not bound to the phases by means of an intervention design or a coach, it is likely that the teams in the study were not aware of what effective data use processes look like. In Flanders, data use is not a common activity, so the data use cycle is not widely known (Van Gasse et al., 2017). The first step in being successful in data use is knowing how the process should look. Only then can all the necessary other competencies that are needed for effective data use (i.e., data literacy) be used.

The last general finding in this study concerns teachers' causal attribution in the data use processes studied. This study showed that teachers only formulated external causes. The cause of student (learning) problems was always assigned to the students themselves (for example, their learning attitude or their subject-specific learning problems). Interactions between teacher practices and student functioning were never raised. As a result, actions that were formulated generally related to the level of the student (for example "providing additional exercises") and only exceptionally to the level of teacher activities. As such, the responsibility for improvement lay mainly with the students. In a limited number of guidance processes, actions were formulated at the level of the learning trajectory or the instructional strategy that was used. The depth of inquiry as conceptualized in this study was rather limited.

The idea that teachers tended not to question their own functioning at the team meetings is not quite surprising. Generally speaking, people tend to attribute successes to their personality and failures to contextual factors (Weiner, 2010). In the field of data use research, the depth of inquiry in data use processes was sometimes questionable because teachers predominantly identified external causes (Schildkamp et al., 2016). This can be due to the fact that processes that need to be questioned are directly related to teacher functioning. This is a sensitive matter and teachers may not like to discuss these issues with all of the colleagues of the formal team. Prior research has shown that the colleagues whom teachers consult for data use are people with whom they maintain friendship-relationships (Van Gasse et al., 2020).

The qualitative research design with observations enabled us to study the data use processes in depth. As such, we learned that only a limited number of student guidance processes are underpinned with data use processes as described in the literature. However, our study sample contained 17 formal teams from two schools. Therefore, the (data use) culture in these schools may have affected our data collection and results as the variation between teams within the schools may be smaller than between schools (i.e., multilevel problem). The research design did not account for this fact.

Another limitation of this study is that it digs deep into the processes of data use but did not shed light on the bigger picture of these processes. This study involves cross-sectional data. This implies that we did not gain insights into the whole guidance process of students during the school year. Additionally, it remains unclear what these data use processes at formal team meetings provide for teachers and students. Maybe despite the superficial processes, the fact that teachers hear stories of each other's practices and discuss improvement actions for students results in later reflections on strategies to cope with certain problems of pupils. As such, some aspects of data literacy (e.g., pedagogical knowledge) may be developed during the meetings. Another effect of data use at student guidance meetings is what it delivers for students. For example, were the actions that were formulated effective for students or how did they perceive them? Fully understanding data use processes requires some insights into the effects as well. Although these insights are often lacking in data use research, they are indispensable to assess and evaluate how effective data use processes should look (Ansyari et al., 2020).

Conclusion

This study has generated some important implications for future research and practice. The results show that achieving complete and in-depth cycles in data use is not an easy endeavor. Therefore, it is crucial to gain insight into how effective data use can be stimulated. A lot of efforts have been made to support data use activities in schools through intervention research. However, it is clear that whereas core teams may get engaged in data use, this engagement does not always flow through the entire school (Hubers et al., 2017). As we know that the relation between formal and informal networks may play an important role here (Van Gasse, 2019), it is crucial to gain insight into how data use can be facilitated apart from intervention settings. In line with what we stated earlier, effect measures should play a role. Assessing and evaluating "what works" in data use cannot take place without considering the long-term and short-term effects. Additionally, more information is needed to explain the current findings. How can decision making on pupils guarantee an appropriate balance between intuitive expertise and data-based decision making? And why does the current balance favor intuition over evidence?

Second, to ensure that teacher teams succeed in completing data use cycles, more insight is needed in how data literacy as a cluster of competencies can be influenced. Data literacy starts with knowing how to implement the data use cycle and being familiar with its challenges and pitfalls. Analytical and data interpretation skills are only a part of the puzzle (Gummer & Mandinach, 2015). Therefore, more insight is needed in how all these particular competencies can be learned by teachers. This can start with research into how data literacy can be measured; such measurements could lead to effect studies on data literacy interventions, such as collaborative inquiry, learning by case studies, etc.

This study is among the first to use systematic observations to expose how teachers use data at formal student guidance meetings. It reveals that the implementation of the data use cycle and the depth of inquiry during this implementation are questionable. Therefore, for Flemish policymakers and practitioners it is essential to raise awareness on the stepwise self-questioning process that data use should be. As efforts are currently being made in Flanders to increase teacher (statistical) data literacy skills, it is crucial to emphasize the importance of approaching data literacy as a broader cluster of competences. Competences to adequately use data for improvement processes start with being aware of the theory of action behind it. Other essential competences (e.g., goal setting, self-questioning, and reflection skills) are also needed to complete the process with a reasonable depth of inquiry. Only by responding to this whole cluster of competences will teachers be able to optimize pupil guidance processes on the basis of data.

References

Ansyari, M. F., Groot, W., & De Witte, K. (2020). Tracking the process of data use professional development interventions for instructional improvement: A systematic literature review. *Educational Research Review, 31*, 1–18. https://doi.org/10.1016/j.edurev.2020.100362

Baten, E., Vansteenkiste, M., De Muynck, G.-J., De Poortere, E., & Desoete, A. (2020). How can the blow of math difficulty on elementary school children's motivational, cognitive, and affective experiences be dampened? The critical role of autonomy-supportive instructions. *Journal of Educational Psychology, 112*(8), 1490 –1505. https://doi.org/10.1037/edu0000444

Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal, 52*(5), 861–893. https://doi.org/10.3102/0002831215599251

Campbell, C., & Levin, B. (2008). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability, 21*(1), 47–65. https://doi.org/10.1007/s11092-008-9063-x

Carlson, D., Borman, G., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Education Evaluation and Policy Analysis, 33*(3), 378–398. https://doi.org/10.3102/0162373711412765

Ciampa, K., & Gallagher, T. L. (2016). Teacher collaborative inquiry in the context of literacy education: Examining the effects on teacher self-efficacy, instructional and assessment practices. *Teachers and Teaching, 22*(7), 858–878. https://doi.org/10.1080/13540602.2016.1185821

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective, 9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Datnow, A., & Hubbard, L. (2015). Teachers' use of data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*(4), 1–26.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change, 17*(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2

De Volder, I. (2012). *Externe schoolevaluaties in Europa*. Garant.

Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2017). The effects of a data use intervention on educators' satisfaction and data literacy. *Educational Assessment, Evaluation and Accountability, 29*(1), 83–105. https://doi.org/10.1007/s11092-016-9251-z

Ehren, M. C. M., & Swanborn, M. S. L. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement, 23*(2), 257–280. https://doi.org/10.1080/09243453.2011.652127

Gelderblom, G., Schildkamp, K., Pieters, J., & Ehren, M. (2016). Data-based decision making for instructional improvement in primary education. *International Journal of Educational Research, 80*, 1–14. https://doi.org/10.1016/j.ijer.2016.07.004

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record, 117*(4), 1–22.

Hubers, M. D., Moolenaar, N. M., Schildkamp, K., Daly, A., Handelzalts, A., & Pieters, J. M. (2017). Share and succeed: The development of knowledge sharing and brokerage in data teams' network structures. *Research Papers in Education*, 1–23. https://doi.org/10.1080/02671522.2017.1286682

Hubers, M. D., Poortman, C. L., Schildkamp, K., Pieters, J. M., & Handelzalts, A. (2016). Opening the black box: Knowledge creation in data teams. *Journal of Professional Capital and Community, 1*(1), 41–68. https://doi.org/10.1108/JPCC-07-2015-0003

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers' College Record, 114*, 1–48.

Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record, 117*(4), 1–40.

Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership, 43*(2), 269–289. https://doi.org/10.1177/1741143214537229

Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School Effectiveness and School Improvement, 27*(2), 228–254. https://doi.org/10.1080/09243453.2015.1056192

Van Gasse, R. (2019). The effect of formal team meetings on teachers' informal data use interactions. *Frontline Learning Research, 7*(2), 40–56. https://doi.org/10.14786/flr.v7i2.443

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2016). Teacher collaboration on the use of pupil learning outcome data: A rich environment for professional learning? *Teaching and Teacher Education, 60*, 387–397. https://doi.org/10.1016/j.tate.2016.07.004

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2017). Unravelling data use in teacher teams: How network patterns and interactive learning activities change across different data use phases. *Teaching and Teacher Education, 67*, 550–560. https://doi.org/10.1016/j.tate.2017.08.002

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2020). Teacher interactions in taking action upon pupil learning outcome data: A matter of attitude and self-efficacy? *Teaching and Teacher Education, 89*, 1–9. https://doi.org/10.1016/j.tate.2019.102989

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2013). Beyond individualism and isolation in schools: What types of teacher teams are there in schools? Or can we merely speak about groups? *Frontline Learning Research, 1*(2), 86–98. https://doi.org/10.14786/flr.v1i2.23

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

Wayman J., Spikes D., Volonnino M. (2013) Implementation of a Data Initiative in the NCLB Era. In: K. Schildkamp, M. Lai, & L. Earl (Eds.), *Data-based Decision Making in Education: Challenges and Opportunities* (pp. 135–153). Springer. https://doi.org/10.1007/978-94-007-4816-3_8

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist, 45*(1), 28–36. https://doi.org/10.1080/00461520903433596

Corresponding authors

Roos Van Gasse
Department of Training and Educational Sciences, University of Antwerp, Belgium
E-mail: roos.vangasse@uantwerpen.be

Martine Mol
Antwerp School of Education, University of Antwerp, Belgium
E-mail: martine.mol@uantwerpen.be

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-6

ASSESSMENT DECISION MAKING IN VOCATIONAL EDUCATION AND TRAINING

HENNING FJØRTOFT,
ELIN BØ MORUD

Abstract

Assessment decision making is a highly contextual phenomenon. In this paper, we explore this topic in vocational education and training (VET). Thirty-eight teachers from five Norwegian upper secondary schools were interviewed before and after an 18-month research–practice partnership. To understand assessment decision making in VET, we draw on two bodies of knowledge: (a) research on teachers' decision making in assessment and (b) conceptualizations of teachers' professional capital. Four main findings emerged from the analysis: three assessment-related dilemmas and one professional capital-related dilemma. We then discuss how these aspects of practice affect assessment decision making and the implications for developing VET teachers' decisional capital.

Keywords

assessment, decision making, professional capital, vocational education, research–practice partnership

Introduction

Assessment decision making in teaching has long been considered a complex phenomenon. Teacher assessment literacy has been conceptualized in several ways, but typically combines a knowledge base of discipline- and pedagogy-related strands, the ability to make sound judgments about student learning processes and performances, and practical skills such as communicating assessment decisions (DeLuca & Braund, 2019; Pastore & Andrade, 2019; Willis et al., 2013; Xu & Brown, 2016). However, assessment decision making is a highly contextual phenomenon. Xu and Brown (2016) emphasized the need to integrate sociocultural phenomena, such as policy, values, and social norms, into teachers' assessment literacy. Similarly, Willis et al. (2013) framed assessment decisions as part of a "dynamic, context dependent social practice" (p. 242) in which teachers negotiate curriculum elements, such as learning goals, with cultural knowledge of classroom phenomena.

The objective of this interview study is to explore assessment decision making in vocational education and training (VET). To understand this practice, we draw on two bodies of knowledge: (a) research on teachers' decision making in assessment and (b) professional capital. We then discuss how these aspects of practice affect assessment decision making and the implications for developing VET teachers' decisional capital.

Theoretical Framework

Teachers' decision making in assessment

Decisions are involved in all aspects of assessment, from design (Bearman et al., 2016; Boschman et al., 2014) to instructional decision making (Garner et al., 2017) to high-stakes assessment settings (Vanlommel & Schildkamp, 2019). Several approaches to teaching, such as diagnostic testing, assessment for learning, and data use, involve decision making as a key component (Van der Kleij et al., 2015). Teachers use a broad set of evidence for decision making in the classroom, such as digital tests, homework assignments, oral tests, paper-and-pencil tests, portfolios, practical tasks, presentations, and questionnaires (Kippers et al., 2018). For example, in grading situations, teachers use a combination of (a) deliberately and systematically and (b) nondeliberately and nonsystematically collected data to make inferences about student learning (Vanlommel & Schildkamp, 2019). A century of research on teachers' grading practices has shown not only that the meaning of grades has been hotly debated, but also that teacher assessment is able to capture multiple dimensions of student learning (Brookhart et al., 2016).

Although some have argued that teachers primarily exercise judgment on an individual level in assessment situations (Kain, 1996), it is now common to consider teacher assessment practice as a situated phenomenon shaped by the collective practices of a community (Allal, 2013). For example, there are considerable national differences in teachers' approaches to assessment due to policy and testing frameworks; at the same time, there are differences at the microlevel between individual teachers' views on issues such as teacher professional autonomy and judgment or student agency and metacognition (DeLuca et al., 2021). Allal (2013) argued that although teacher judgment is subject to error and bias, it is similar to clinical judgment in the medical professions in that teacher judgment establishes a relationship between everything the evaluator knows about a particular individual and a wide array of knowledge, including explicit and tacit professional knowledge as well as institutional norms and rules. Therefore, learning how to make sound assessment-related decisions is not a simple procedural task but one of "earning foundational ideas and building an integrated stance toward teacher as assessor through contextualized reflective learning" (DeLuca & Braund, 2019, p. 13)

Teachers' internal beliefs and values often clash with the pressures of external demands, creating tensions between the practicalities of classrooms and the rigorous application of measurement principles (McMillan, 2005). This suggests that teacher assessment decision making is affected in various ways by contextual factors such as policy and accountability frameworks, assessment practices (e.g., psychometric approaches, written essays, and performance assessments), and professional autonomy and judgment. Such contextual factors are likely to shape decision-making procedures for grading (e.g., the balance between analytical and holistic approaches to scoring, what counts as acceptable evidence of student learning, and the approaches to moderation used to ensure reliable results).

Tacit knowledge is important for skill development and has long been considered an important part of teachers' assessment literacy. Tacit knowledge is a crucial part of teachers' professional knowledge in feedback situations (Sadler, 1998), and scholars have argued that tacit knowledge is required if teachers are to provide students with meaningful knowledge of standards and criteria (O'Donovan et al., 2004). When making assessment decisions, teachers move back and forth between tacit and explicit knowledge (Wyatt–Smith et al., 2010). In short, tacit knowledge constitutes a key component of assessment decision making, alongside numerical cut-offs, exemplars, and verbal descriptions (Sadler, 1987).

Recently, researchers have conceptualized teachers' decision making as two distinct processes: a rational process using purposively collected data and an intuition-driven process in which teachers process cues almost effortlessly and base their decisions on intuitive expertise and "feelings of knowing" (Vanlommel et al., 2017, p. 82). Navigating complex dilemmas to make decisions is a key part of this process (Xu & Brown, 2016). If teachers' decision making relies on rational data-based processes and intuitive "knowing/feeling" processes (Vanlommel et al., 2017), then our current understanding of teachers' decision making must be further enriched by contextual studies exploring the interplay of knowing and feeling in assessment decision making. Therefore, we turn to the concept of professional capital as a way of framing our understanding of assessment decision making.

Professional capital

The concept of teachers' professional capital is a useful theoretical lens for understanding their assessment decision making. Professional capital is a conceptualization of teachers' professionalism that includes three aspects: human capital (individual talent), social capital (relational trust and collaborative capacity), and decisional capital (making good judgments with incomplete or conflicting evidence; Hargreaves & Fullan, 2012). Without decisional capital, defined as the "ability to make discretionary judgements" (Hargreaves & Fullan, 2012, p. 93), human capital and social capital are insufficient. However, because external factors, such as curricula, assessment policies, and governance structures, vary across boundaries, teachers' professional capital may take different forms in different contexts (Shirley, 2016).

In assessment contexts, decisional capital is important because the ability to make judgments is acquired by examining and comparing cases in structured and unstructured experience, practice, and reflection and is enhanced by drawing on colleagues' insights and experiences (Hargreaves & Fullan, 2012). Unlike procedurally formed decisions, decisional capital cannot be based on fixed rules or incontrovertible evidence. Therefore, decisional capital is inherently social in that it rests on the accumulated experience of other professionals.

The components of assessment literacy do not have the same meaning or the same importance across contexts, and teachers must sometimes navigate competing assessment demands within their classrooms (Pastore & Andrade, 2019). Given the dynamic and contextually sensitive nature of assessment practices (Willis et al., 2013), assessment decision making should be investigated across contextual and cultural borders so that our understanding of the phenomenon is enriched. For example, the competing narratives that teachers must manage in assessment decision making require teachers to navigate and adapt to complex situations (Bonner, 2016). Previous studies of teachers' intuitive decision making have been conducted in primary school settings (Vanlommel et al., 2017); research in upper secondary school settings is scarce.

Assessment affects student learning, emotional well-being, and future opportunities in many ways. Although few studies have focused on teacher perceptions of grading, surveys have shown that teachers often include noncognitive and nonachievement factors such as effort or participation (McMillan, 2019). In some cases, decision making and responsibilities are subsumed under rational-legal forms of authority, increasing standardization of work procedures, and managerialism (Evetts, 2009). These issues are likely to converge in assessment dilemmas in which teachers must make decisions that impact student learning, well-being, or further career opportunities.

Professional capital among VET teachers is at least as complex as that of teachers in conventional academic subjects. In many countries, VET has traditionally focused on the acquisition of tacit knowledge and practical skills through hands-on experience (Lave & Wenger, 1991; Sennett, 2008). However, in recent years, teachers have also been required to teach and assess students' basic, digital, and soft skills (Organisation for Economic Co-operation and Development [OECD], 2021). For example, career guidance is a social activity that requires collegiality, support, and trust from a range of stakeholders (Hearne & Neary, 2021). However, it has been suggested that VET teachers are less likely to engage in deep collaboration that requires high levels of interdependence (Bükki & Fehérvári, 2021). Therefore, efforts have been made to improve VET teacher collaboration, such as by using action research and professional learning community approaches (Andreasen & Duch, 2020).

Drawing on Willis et al. (2013), we focused on VET teachers' need to negotiate the intersection of locally generated understandings of national policies, the teachers' learned knowledge of a discipline or vocation (e.g., mathematics or construction techniques), and their personal beliefs about learning and assessment as developed through experience. The following research question guided this study: *What assessment dilemmas do VET teachers encounter, and how do these dilemmas affect the communities in which the teachers participate?*

Methods

Research design overview

In this qualitative study, we explored assessment decision making in the VET context. To examine the phenomenon, we followed a pragmatist approach. As part of a larger research–practice partnership in upper secondary schools in Norway, we conducted semi-structured group interviews before and after an 18-month research–practice partnership. The study was informed in part by the theoretical framework outlined in the literature review section and in part by the researchers' experiences participating in research–practice partnership projects in VET. The present study is a secondary analysis focusing on assessment decision making in schools offering VET programs.

We concentrated on a group of teachers in schools offering vocational programs and the teachers' decision making in assessment situations. We explored the teachers' perceptions and experiences with decision making as part of the research–practice partnership.

Study participants and data sources

The Norwegian curriculum and assessment system is described in the national curriculum document. The curriculum document privileges disciplinary learning objectives (e.g., mathematics, literacy, and various vocational topics) and a broader human development perspective on values and principles (e.g., human dignity, ethical awareness, and democracy and citizenship) in an equal manner. Teacher assessment decision making plays a crucial part in this system. However, there are few guidelines for decision making, and the system, therefore, relies heavily on teacher professionality.

This study is a secondary analysis of data collected as part of a larger project involving 19 upper secondary schools and 12 university researchers collaborating to solve urgent problems of practice identified by the schools (Fjørtoft & Sandvik, 2021). We selected data from five upper secondary schools offering one or more VET programs in the partnership. We excluded schools without VET programs and departments that offered tertiary vocational education (i.e., corresponding to higher engineering education). School leaders were asked to nominate interview participants based on a maximum variation strategy (i.e., both genders, varying levels of teaching experience, and a range of curriculum areas). Thirty-eight teachers from five Norwegian public upper secondary schools were selected by the principals. The teachers were interviewed during the 18-month research–practice partnership initiative (see Table 1 for details). The participants did not receive incentives for participating and were allowed to withdraw at any time during the interview process.

Table 1
Interview Participants

Interview/school	Participants: Teaching background, gender, and years of experience as a VET teacher						Length of interviews (min)
A	Agronomy; *male*; 5 years of experience (A1)	Welding; *male*; 15 years of experience (A2)	Construction techniques; *male*; 33 years of experience (A3)	Process technology; *male*; 15 years of experience (A4)			103
A	Healthcare; *female*; 9 years of experience (A5)	Nursing; *female*; 13 years of experience (A6)	Pre-school education and School counseling; *female*; 15 years of experience (A7)	Nutrition and health care; *female*; 35 years of experience (A8)			78
B	Information technology; *male*; 4 years of experience (B1)	Electrical engineering; *male*; 7 years of experience (B2)	Language; *female*; 12 years of experience (B3)				47
C	Construction techniques; *male*; 12 years of experience (C1)	Floral design; *female*; 29 years of experience (C2)	Gardening and horticulture; *male*; 28 years of experience (C3)	Railway engineering; *female*; 4 years of experience (C4)	Carpentry; *male*; 9 years of experience (C5)		51
D	Health care; *female*; 10 years of experience (D1)	Health care; *female*; 13 years of experience (D2)	Construction techniques; *male*; 4 years of experience (D3)	Process technology; *male*; 15 years of experience (D4)	Health care; *female*; 35 years of experience (D5)		57
E	Electrical engineering; *male*; 4 years of experience (E1)	Skin care; *female*; 8 years of experience (E2)					35
F	Social science; *male*; 25 years of experience (F1)	Agriculture; *male*; 25 years of experience (F2)	Agriculture; *female*; 5 years of experience (F3)	Landscaping; *female*; 5 years of experience (F4)	Gardening; *female*; 9 years of experience (F5)	Language; *female*; 8 years of experience (F6)	39

G	Industrial engineering; *male*; 15 years of experience (G1)	Language; *male*; 20 years of experience (G2)	Industrial engineering; *male*; 7 years of experience (G3)				51
G	Automation and Electrical engineering; *male*; 18 years of experience (G4)	Language; *female*; 12 years of experience (G5)	Automation and Electrical engineering; *male*; 7 years of experience (G6)				51
G	Social science and Language; *male*; 39 years of experience (G7)	Physical education and Language; *male*; 8 years of experience (G8)	Restaurant and Catering services; *female*; 20 years of experience (G9)				32
	The average length of experience is 14.7 years. The median length of experience is 12 years. 21 were male, and 17 were female.					Total: 544 min	

Researcher positionality

We participated in a team of researchers supporting the development of assessment literacy in upper secondary schools during a 5-year period. We were involved in data collection and the research–practice partnership with the schools. Each school received support for a minimum of 18 months. The data collection was approved by the Norwegian Centre for Research Data.

We were participant observers in at least one of the schools. We also met school leaders from all the schools at network meetings, and we frequently discussed teachers' and school leaders' efforts to improve assessment decision making. This allowed us to interpret the dataset in the context of teachers' daily practice.

Data collection

We conducted a secondary analysis of an existing set of interview data, including data collected by other researchers participating in the project. Four interviews were conducted by the research team before the research–practice partnership took place, and six were conducted after it concluded. The interviews lasted between 32 and 103 min (mean duration 51 min). The questions were open to allow for teachers to express their reflections and experiences, and included probing for assessment dilemmas and compromises, how expectations were communicated to students, emotional issues related to assessment (for teachers, students, and school leaders), and teachers' assessment identities and conceptions of assessment literacy. For example, we asked teachers what kinds of decisions they made when assessing, what emotions were related to assessment, and how they viewed themselves as assessors. We audio-recorded the interviews using digital devices, transcribed the data verbatim, and selected a sample consisting of 10 group interviews with 2–6 teachers per group (N = 38 teachers). We selected interviews with teachers working in the VET context.

Analysis

The first coding cycle was conducted using an inductive approach. We familiarized ourselves with the dataset, coded parts of the dataset, and discussed our coding approach collaboratively. This approach yielded a variety of codes related to the organization of the partnership and awareness of assessment-related issues. For example, teachers commented about resourcing and time allocation in the partnership. Teachers also discussed the process of becoming aware of the role of various assessment tools across situations or ensuring that students became aware of the role of self-assessment or peer assessment practices.

In the second coding cycle, we aggregated the codes by relating them to concepts drawn from literature regarding teachers' professional capital and assessment dilemmas or decision making. This body of literature is vast; therefore, we focused primarily on publications related to secondary education and VET. We followed the abductive approach, which meant that this process was reflexive, where the emerging codes were related to the literature and vice versa.

Choosing which inferences to follow is a key challenge in abductive analysis, as inferencing is a skill developed through acquiring ways of seeing and habits of thought (Tavory & Timmermans, 2014, pp. 38–39). Our inferencing relied on our familiarity with secondary schools (both authors have worked as teachers), with research–practice partnership and partnership activities (both authors have extensive experience coordinating and participating in research–practice partnership initiatives), and with assessment research. For example, we identified several instances of teachers discussing dilemmas that arise in their decision-making practices. This led to reviewing the literature on dilemmas in assessment decision making and the role of dilemmas in professional capital. Consequently, our positioning as researchers led us to reflect on our multiple roles as scholars and educators in the partnership and to scrutinize our epistemological assumptions and theoretical lenses.

We conducted the analysis individually and collaboratively; organized codes, categories, and samples from the dataset in spreadsheets; and cross-checked results in all stages until we reached agreement. This approach is similar to coding techniques where codes are considered open and fluid, and where the coding process is evolving and recursive; such approaches are considered interpretive and conceptual, reflecting the researchers' engagement with and interrogation of the data (Braun & Clarke, 2021). For example, we combined the initial codes *body language* and *specialist terms* into a category called *Communication* and then conceptualized the codes as tensions between tacit and explicit knowledge. We resolved interpretive conflicts during each stage of analysis.

Results and Interpretation

Four main findings emerged from the analysis: three assessment-related dilemmas and one professional capital-related dilemma. The assessment-related dilemmas consisted of (a) tensions between tacit and explicit forms of knowledge; (b) tensions between curriculum objectives, business standards, and student ipsative goals; and (c) students at risk of failing. The professional capital-related dilemma was related to the interplay of assessment decisions and teachers' broader professional capital.

Assessment-related dilemmas

Tensions between tacit and explicit knowledge

Several dilemmas were related to tensions between tacit and explicit knowledge in VET settings. In some cases, this tension was related to justifying the assessment of students' behavior and social skills. Often, there was a discrepancy between the knowledge that students reproduced in assessment situations and students' behavior in professional settings. For example, some students performed poorly in their written responses but better in practice vocational contexts. Teachers described the dilemma of assessing students who behave in ways that contradict "what they write on paper" (A6) as challenging and noted "huge contrasts if you have a learning objective in relation to vocational behavior" (A8). Furthermore, teachers noted that students' use of mobile phones or cursing was unacceptable in some situations. We interpret this as an indication of the discrepancy between explicit and tacit knowledge in vocational settings.

We also found that the teachers talked about dilemmas connected to communication skills as part of vocational practice. For example, in healthcare vocations, students are expected to be able to communicate their theoretical understanding to teachers and to communicate with patients (e.g., older individuals, young children with minority language backgrounds, or patients who are hard of hearing). Furthermore, students in service industries are expected to communicate well with a range of customers and cater to their well-being and individual needs. This requires considerable tacit knowledge and situational awareness. The teachers commented that some students mastered communicating with patients well without being able to communicate professionally using specialist language. This situation constituted an assessment dilemma for the teachers: "Some students can be incredibly skilled at communicating . . . So, the question is how to assess [communicative] skills if the theoretical content is thin" (A8).

Furthermore, communicating about expectations was a challenge for teachers, especially when students had a poor understanding of the standards. One teacher described the discrepancy between students' and teachers' understanding in emotional terms:

> Some might come to me saying, "I'm hoping for a 6" [the top grade]. And I've only had them for a month. And already I know that if we can get those students to achieve a 3 [midrange grade], then I'll be really happy. And it's hard to reach them. (A7)

In conclusion, there are considerable tensions between tacit and explicit knowledge in assessment decision making in VET. This tension affects and is affected by a range of other factors, such as communication, behavior, and standards.

Tensions between curriculum objectives, business standards, and student ipsative goals

Teachers reported experiencing a gap between the standards in the curriculum, expectations for vocational performance in businesses, and individual students' academic level. This led to teachers spending time uncovering students' existing knowledge and skills, inviting business representatives to the school to share their expectations, and emphasizing to the students the importance of meeting such expectations:

> We have to assess them based on what we are teaching and what we have been through. We can't just go on and on teaching if they don't know anything—if nothing sticks. So, we have to figure out where they are in terms of the student and in terms of our teaching. (C3)

> At Vg2 [the second year], I send them off for an apprenticeship period. They are learning construction work, and I have had people from the trade come to school several times. The business owners clarify what they envision in their employees. We are supposed to involve the private sector, right? So they come in and provide an idealized version. This provides a lot of guidance for how to behave for the students. (C3)

> VET […] is really an application process directed towards the business. That's when they get to show what they can do, so it's important that they are met in relation to the expectations there—becoming aware of being on time, being loyal, and doing what you're told. Those things are pretty specific. (C5)

However, this work was further complicated by the need to negotiate business standards and the needs of the student in some cases. Teachers sometimes asked students to focus on just a few curriculum objectives because the teachers knew that these objectives were more important in the business world. Formally speaking, this is in violation of the national curriculum, which states that students should master all domains in the curriculum. However, the teachers made such idiosyncratic decisions because they were familiar with the expectations in the business world. "I know what my colleagues expect from the students who leave school," a teacher (A6) commented, defending his choice to ignore certain parts of the curriculum in some cases.

Other teachers struggled to keep up with the dynamics of business standards. One teacher commented that he felt out of touch with the world of industrial practice after having worked as a VET teacher for more than a decade: "We've been in the school system for 10 to 15 years or more, so we find ourselves a little bit on the outside of society, so we've been isolated quite a while" (A4).

The teachers had developed a strategy for navigating this dilemma. For example, teachers mentioned using self-assessment and reflection exercises to prepare students for vocational standards. One teacher illustrated this practice with a perspective-taking activity in which the roles of the customer and the worker were reversed. The activity was coupled with reflection on professional standards and developing a sense of pride: "We're pushing professional pride a lot. What do you think the customer expects from a skilled worker?" (B2). This reversal of perspectives was intended to support students in understanding vocational performance from the perspective of clients and customers.

Students at risk of failing

Although national policies explicitly prohibit assessing student effort as part of the final grades, teachers felt that effort should be included as a mitigating factor for students at risk of failing. This seemed especially pronounced in borderline cases of passing or failing. A teacher (A3) said, "If they haven't shown up at school, or if I see that a student is unwilling or not trying, I would rather fail that student compared to others who show up at school and do their best." This practice is known as "pulling for students" (teachers want to give students the highest grades possible) and explains why nonachievement factors such as effort and improvement have been important in grading (Bonner, 2016). Several teachers reported experiencing emotional difficulties in making decisions in high-stakes situations carrying potentially grave consequences for the students: "I have strong feelings related to assessment ... A passing or nonpassing grade can have huge consequences for them" (A4). One teacher described losing sleep over grading and the process as "terribly painful" (A4). Another teacher reported similar experiences: "The worst time of the year is when you are assigning final grades because you are making decisions for the entire future of the students. So you're shaky all the time" (A8). One teacher discussed how decisions related to potentially failing at-risk students lead to increased teacher workload, deliberation with school leadership, and ultimately, negative consequences for students' potential opportunities:

> I should have failed a student in science. But the student wanted to become a truck driver, so I gave him a 2 [lowest passing grade] so that he passed. Failing him in science would have meant a mountain of work for me. I discussed this with the school administration. And I had to see the bigger picture, too. He is working as a truck driver now and has a trade certificate. (A2)

These findings confirm that noncognitive aspects, such as emotions and intuition, impact teachers' decision making in VET. Furthermore, the findings suggest that teachers may struggle with specific motivations related to supporting students by "pulling" and that teachers use evidence of students' success after graduation to support their decision to do so.

Professional capital–related dilemma

The fourth dilemma was related to teachers' professional capital and their ability to use decisional capital in developing their assessment decision making. Several teachers commented that assessment was a socially situated practice and that shared understandings were required to maintain high levels of consistency: "We don't assess alone. We assess together with other teachers who share teaching interdisciplinary responsibilities" (C1). "Feedback and grading and assessment and all that . . . It should be the same for all. There is a certain degree of discrepancy" (A2). However, some teachers resisted engaging in the community, a stance other teachers deemed unproductive. E2 said, "There are people in an organization who are not willing [to change] and who explicitly resist participating [in research–practice partnership activities]. They have a negative impact on group processes, frankly speaking."

The desire for improvement was evidenced in several statements and illustrates how teachers felt that the program was helpful in improving decision making. A teacher (G4) said, "You get a colleague who is more alert and forward-leaning. You don't lean back and say, 'The next 10 years I'm going to be a laidback teacher.' No, [I'm] forward-leaning and more focused on the student." Another teacher (D2) stated, "She [the principal] keeps saying that all the time. We are not stopping with [the project] now; we're going to continue what we're doing. I'm thinking there's a lot to be seen in relation to assessment."

However, in one school, teachers felt that the research–practice partnership program was too unstructured, and that clearer leadership was needed. A teacher (E2) stated, "Having a project where you are free to find your own line of inquiry is a good idea, but in our department, we would have benefited from a tighter style of leadership, providing more structure and supervision." The teachers at this school were less enthusiastic about the program and seemed to resist the opportunity to exercise professional autonomy. Thus, not all teachers chose to seize the opportunity to develop as practitioners.

Discussion

Teacher decision making is a rational and intuition-driven process (Vanlommel et al., 2017) and requires teachers to build an integrated stance through contextualized reflective learning (DeLuca & Braund, 2019). If teachers in conventional academic subjects struggle to reconcile different aspects of the assessment system (Bonner, 2016), integrating rational and intuitive processes is likely to be even more challenging for VET teachers. In particular, discrepancies between school curricula, business standards, and student needs constitute threats to the integrity of assessment practices in VET. The present study showed that VET teachers face a range of dilemmas in their assessment decision making. Some dilemmas are well-known from previous literature, such as negotiating tensions between tacit and explicit dimensions of learning, students at risk of failing, and problems related to professional collaboration between teachers in assessment-related questions. Other dilemmas are specific to the nature of VET, such as tensions between different sets of goals (i.e., curriculum, business, and student goals).

In the remainder of this discussion, we focus on three aspects of assessment decision making in VET: decision making in high-stakes situations; discrepancies between curriculum, business standards, and student goals; and negotiating vocational learning and human development. We chose these aspects because they illustrate assessment dilemmas in VET contexts and how these dilemmas affect teacher communities.

First, in high-stakes situations, assessment decision making requires teachers to exercise judgment, especially in situations in which students may suffer dire consequences. McMillan (2019) suggested that teachers consider factors that have negatively influenced student achievement (e.g., illness) when making decisions in borderline cases, as well as tipping the balance toward higher grades when students show a clear learning progression. Although teachers' tendency to "pull for students" (wanting to give the highest grades possible) inevitably leads to questions of how decision making can be operationalized, as well as the fairness of using nonachievement factors such as effort or improvement in decision-making situations, idiosyncrasies may paradoxically lead to enhanced validity (McMillan, 2019). This recommendation is likely important for VET contexts, given the complexity of assessing the entirety of vocational knowledge and skills, as well as other cognitive and affective aspects of vocational learning. Adding the business sector as a third stakeholder in such attempts would involve stakeholders in setting standards and building tacit knowledge of vocational standards through shared experiences.

Second, although curriculum standards are stable and change only during periods of reform, vocational standards are less explicit and more dynamic, as illustrated by the interplay between the supply and demand of goods and services. For example, the transition from combustion to electric engines represented a paradigm shift in the world of mechanics. Consequently, VET teachers must negotiate a static (but explicit) curriculum and a dynamic and implicit set of standards in the world of business and commerce. The teachers must also prepare students to navigate the dynamics of the same sector. Lack of access or exposure to vocational communities may cause teachers' tacit knowledge to weaken and standards to concurrently shift. Discussing the role of tacit knowledge in assessment situations, Sadler (1987) suggested (a) sharing experiences through moderation attempts and (b) inviting students into such shared experiences to improve their understanding of criteria and standards. This could be achieved through dialogue, observation, practice, and imitation processes (O'Donovan et al., 2004) by which "exposure to other people's imaginations and strategies extends and enriches the teacher's repertoire of tactical moves" (Sadler, 1998, p. 81).

Third, the VET teachers in this study work in the Norwegian education system, which strives for a balance between employability (ensuring skilled workers) and a "whole student" philosophy. Therefore, assessment decision making must always negotiate the development of vocational skills with human development. The processes outlined in our discussion of high-stakes situations might also enable teachers to navigate this dilemma. Furthermore, teachers must also consider student ipsative goals as part of the larger human development process, including students in shared experiences of understanding standards and assessment criteria or making sense of teacher feedback. This is especially important if teachers are to provide students with feedback on their progression toward curriculum standards as well as the development of their character.

Limitations

This study was based on a small sample size and focused primarily on teachers' self-reported practices and perceptions. Furthermore, the study was contextually limited to Norwegian schools, where teachers enjoy considerable professional autonomy. Therefore, because of the variety of structure and content in VET education internationally, the applicability of the findings to other contexts is limited. Further research, especially comparative and using different kinds of data, would expand the understanding of assessment decision making in VET represented in this study. In particular, extending the research to include the perceptions of students and businesses would enrich the perspectives.

Implications

Research on assessment literacy suggests that reflection on decision making and participation in community activities are two main ways for teacher learning to occur (Xu & Brown, 2016). Encountering dilemmas in assessment practice (e.g., being confronted with ambiguous or conflicting evidence regarding a student's learning outcomes) may trigger teachers to seek additional information and reduce uncertainty (Allal, 2013). Although this shows that assessment decision making can be developed by reflecting on contextually relevant dilemmas, it is unlikely that VET teachers' decision making is improved solely by focusing on dilemmas alone. Hargreaves and Fullan (2012) pointed to the need for collective responsibility and external accountability in the teaching profession. Therefore, by situating research–practice partnerships in the communities in which teachers work, and by reflecting on the specific dilemmas teachers encounter in their practice, teachers could develop their decisional capital using contextually relevant cases to build principles for sound judgment. This is especially relevant for VET, where tensions between tacit and explicit knowledge and different sets of standards in curricula and businesses shape teachers' decisions. However, the relation between collective responsibility and external accountability is also fraught in other areas of education. Therefore, dilemmas in assessment decision making should be considered as a threat to assessment integrity and as a potential source for teacher learning.

References

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364

Andreasen, K. E., & Duch, H. (2020). Teachers' professional development at VET – Reflections on professional learning communities (PLCs). In C. Nägele, B. E. Stalder, & N. Kersh (Eds.), *Trends in vocational education and training research*, *Vol. III*. Proceedings of the European Conference on Educational Research (ECER), Vocational Education and Training Network (VETNET) (pp. 11–18). https://doi.org/10.5281/zenodo.3997331

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, *21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217

Bonner, S. M. (2016). Teacher perceptions about assessment: Competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (1st ed., pp. 21–39). Routledge.

Boschman, F., McKenney, S., & Voogt, J. (2014). Understanding decision making in teachers' curriculum design approaches. *Educational Technology Research and Development*, *62*(4), 393–416. https://doi.org/10.1007/s11423-014-9341-x

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Bükki, E., & Fehérvári, A. (2021). How do teachers collaborate in Hungarian VET schools? A quantitative study of forms, perceptions of impact and related individual and organisational factors. *Empirical Research in Vocational Education and Training*, *13*(2), 1–26. https://doi.org/10.1186/s40461-020-00108-6

DeLuca, C., & Braund, H. (2019, June 25). *Preparing assessment literate teachers*. Oxford research encyclopedia of education. Retrieved 2021, October 31 from https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-487

DeLuca, C., Rickey, N., & Coombs, A. (2021). Exploring assessment across cultures: Teachers' approaches to assessment in the U.S., China, and Canada. *Cogent Education*, *8*(1), 1921903. https://doi.org/10.1080/2331186X.2021.1921903

Evetts, J. (2009). The management of professionalism: A contemporary paradox. In S. Gewirtz, P. Mahony, I. Hextall, & A. Cribb (Eds.), *Changing teacher professionalism international trends, challenges, and ways forward* (1st ed., pp. 19–30). Routledge.

Fjørtoft, H., & Sandvik, L. V. (2021). Leveraging situated strategies in research–practice partnerships: Participatory dialogue in a Norwegian school. *Studies in Educational Evaluation*, *70*, 101063. https://doi.org/10.1016/j.stueduc.2021.101063

Garner, B., Thorne, J. K., & Horn, I. S. (2017). Teachers interpreting data for instructional decisions: Where does equity come in? *Journal of Educational Administration*, *55*(4), 407–426. https://doi.org/10.1108/JEA-09-2016-0106

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school* (1st ed.). Teachers College Press.

Hearne, L., & Neary, S. (2021). Let's talk about career guidance in secondary schools! A consideration of the professional capital of school staff in Ireland and England. *International Journal for Educational and Vocational Guidance, 21*(2), 1–14. https://doi.org/10.1007/s10775-020-09424-5

Kain, D. L. (1996). Looking beneath the surface: Teacher collaboration through the lens of grading practices. *Teachers College Record*, *97*(4), 569–587. https://www.tcrecord.org/Content.asp?ContentId=1408

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199–213. https://doi.org/10.1016/j.tate.2018.06.015

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (1st ed.). Cambridge University Press.

McMillan, J. H. (2005). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 84–112). ASCD.

O'Donovan, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, *9*(3), 325–335. https://doi.org/10.1080/1356251042000216642

OECD. (2021). *Teachers and leaders in vocational education and training.* OECD Reviews of vocational education and training. OECD Publishing. https://doi.org/10.1787/59d4fbb1-en

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, *13*(2), 191–209. https://doi.org/10.1080/0305498870130207

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84. https://doi.org/10.1080/0969595980050104

Sennett, R. (2008). *The craftsman* (1st ed.). Yale University Press.

Shirley, D. (2016). Three forms of professional capital: Systemic, social movement, and activist. *Journal of Professional Capital and Community*, *1*(4), 302–320. https://doi.org/10.1108/JPCC-08-2016-0020

Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research.* University of Chicago Press.

Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, *22*(3), 324–343. https://doi.org/10.1080/0969594X.2014.999024

Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal*, *56*(3), 792–821. https://doi.org/10.3102/0002831218803891

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, *83*, 75–83. https://doi.org/10.1016/j.ijer.2017.02.013

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9

Wyatt–Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, *17*(1), 59–75. https://doi.org/10.1080/09695940903565610

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Corresponding authors
Henning Fjørtoft
Department of Teacher Education, NTNU Norwegian University of Science and Technology, Trondheim, Norway
E-mail: henning.fjortoft@ntnu.no

Elin Bø Morud
Department of Teacher Education, NTNU Norwegian University of Science and Technology, Trondheim, Norway
E-mail: elin.morud@ntnu.no

Studia paedagogica
vol. 26, n. 4, 2021
www.studiapaedagogica.cz
https://doi.org/10.5817/SP2021-4-7

EXPLORING LEADERSHIP INFLUENCE WITHIN DATA-INFORMED DECISION-MAKING PRACTICES IN AUSTRALIAN INDEPENDENT SCHOOLS

VENESSER FERNANDES

Abstract

There is increasingly strong pressure on schools to use data within their decision-making processes; the pressure comes not just from high-stakes testing but also from the subsequent comparative analysis conducted in the international, national, state, and local jurisdictions that represent the educational systems responsible for ensuring that students continue to receive quality education (Harris & Jones, 2017). There is paucity in empirical research within Australia on the practice of data use within schools; research is lacking on data interactions among school leaders in their workplace settings (Coburn & Turner, 2012). This study contributes toward this identified gap in Australian research literature on the practice of data-informed decision making (DIDM) in schools. Using a case-study approach at two K-12 independent schools in Victoria, Australia, the study sought to understand the "how" and "why" of DIDM systems that are currently in use within Australian independent schools in order to better understand what data-informed school improvement processes are being used in practice in this sector of Australian schooling. Based on the findings, we offer recommendations for developing improved system capabilities that make schools data literate and numerate and identify the important transformational role that senior and middle-level school leaders play in building up data-informed collaborative school cultures within their schools.

Keywords

school leaders, data-informed decision making, collaboration, continuous school improvement, organizational trust

Introduction

The last decade has seen increasing pressure on schools to use data within their decision-making processes. This pressure comes from high-stakes testing and from subsequent comparative analyses conducted at international, national, state, and local levels of jurisdiction. These jurisdictions represent the educational systems responsible for ensuring students are receiving quality education (Harris & Jones, 2017). Similar to the United States, in Australia "the use of data in educational decision-making is expected to span all layers of the education system—from the federal to the state, region, school and classroom levels" (Means et al., 2009, p. 7). Some studies have looked at the growing emphasis given to high-stakes testing (Harris & Hargreaves, 2015) as one of the reasons why more data use has been seen (Smeed et al., 2011). Research indicates that a key strategy directly contributing toward school improvement is effective data use within their decision-making processes (Argyris & Schön, 1996; Brown, 2015; Coburn & Turner, 2012; Earl & Katz, 2006; Earl & Timperley, 2009; Fernandes, 2016, 2019, 2021; Phillips, 2003; Robinson & Walker, 1999; Schildkamp et al., 2017; Schildkamp et al., 2019; Timperley, 2005a, 2005b, 2006; Vanlommel et al., 2018).

As Australian independent schools (N=1169 schools) are autonomous, their accountability frameworks and continuous school improvement processes vary, unlike the public or Catholic school sectors. Freedom in autonomous decision making is considered to be fundamental to independent schooling in Australia. This autonomy enables them to work in partnership with their school community to develop unique and custom-built schools that meet the specific needs of their students (ISCA, 2018). Independent schools are also compliant with external accountabilities set by state and federal education regulations and standards.

Australian independent schools have always been self-managing school systems. Within their regular systems of educational practice, reasonable levels of data literacy and data numeracy are expected to be used when making informed educational decisions. This role of data use in educational decision making has gained traction in recent years with the accessibility of information and mandated transparency of school data in Australia. As reported by the Independent Schools Queensland (2018):

> There is now more publicly available information on schools than ever before, particularly in the case of independent schools. This means their level of accountability to parents, Governments and the wider community has increased significantly. Transparency in terms of school outcomes is now firmly embedded into the education system… An important aspect of school transparency and accountability is the need for Governments to be assured

> their increasing expenditure on schooling is being used appropriately and is achieving outcome levels that are in line with national and state expectations. Governments collectively spend nearly $60 billion annually on schools; it is not unreasonable on behalf of taxpayers, that they should have data to indicate the effectiveness of such expenditure (pp. 1–2).

School data includes any relevant information about school stakeholders derived from qualitative and quantitative sources internal or external to the school representing a certain aspect of school improvement being studied. The researcher recorded 32 different kinds of datasets commonly found in schools in Victoria, Australia (Fernandes, 2019). More broadly, these datasets could be divided into four groups: demographic data, student learning data, perception data, and school processes data (Bernhardt, 2003), all providing insights into various aspects of Australian schools.

This paper discusses how two Australian K-12 independent schools make use of data-informed decision-making (DIDM) approaches to inform their school improvement processes. Second, it discusses how senior and middle-level leaders at these schools have engaged in understanding their data, made data-informed decisions, and moved their continuous school improvement processes forward. Third, this paper focuses on investigating how the practice of a DIDM approach was developed at these case-study schools. Two themes emerged from the findings: the level of dynamic over passive enhancement of school capabilities and the level of active versus passive leadership involvement in DIDM.

Literature Review

An international body of literature supports DIDM and its role in equipping school leaders and teachers in improving educational outcomes within their schools (Datnow & Hubbard, 2015, 2016; Datnow & Park, 2018; Earl & Katz, 2006; Fernandes, 2019, 2021; Lai et al., 2014; Lai & Schildkamp, 2013; Mandinach et al., 2015; Mandinach & Gummer, 2013; Poortman & Schildkamp, 2016; Schildkamp & Kuiper, 2010; Thoonen et al., 2011; Van Gasse et al., 2016; Vanhoof et al., 2012; Vanlommel et al., 2016; Wayman et al., 2012; Wayman & Jimmerson, 2014).

The divide between expectations of data-informed school improvement practices and the actual realities of school improvement practice (Honig & Coburn, 2008; Spillane, 2012) are significant in Australia. As Schildkamp et al. (2017) noted, "although data-based decision-making can lead to improved student achievement, data are often not used effectively in schools" (p. 242). A better understanding of DIDM would help schools in making better use of their data for school improvement (Fernandes & Henderson, 2020). A data-informed school has: shared leadership and responsibility; a mission identifying its goal and purpose as a learning community; a cadre of leaders, learners, inquirers, and worriers; identification of critical issues, essential questions, and focus problems; planning strategies to collect data and information; processes for implementation; resources and time to engage and complete tasks; feedback and recycling of experiences; reporting and sharing data and experience; and repetition and continuation of the process of data collection, use, and change (Hansen et al., 2003). Coburn and Turner (2012) identified a definite lack of empirical research on the practice of data use within schools:

> …in spite of all of the policy and reform activity focused on data use in education, empirical research on data use continues to be weak. In particular, we still have shockingly little research on what happens when individuals interact with data in their workplace settings (p. 99).

Addressing this gap in Australian research literature on DIDM, this paper seeks to understand the "how" and the "why" of DIDM systems in practice within two independent schools in Victoria, Australia. "Practice" within this context is defined as "the coordinated activities of individuals and groups in doing their 'real work' as informed by particular organizational or group contexts" (Cook & Brown, 1999, pp. 386–387). Schools that use data effectively practice data use regularly.

Park and Datnow (2009) suggested that data-driven decision-making is co-constructed by multiple actors at three different levels of interaction. First, leaders play a pivotal role at all levels in co-constructing the vision and implementation of data-driven decision-making through their framing of the purpose of data use and the creation of an ethos of learning and continuous improvement rather than one of blame. Next, leaders distribute decision-making authority, empowering different staff members to utilize their level of expertise. Third, the school system directs their resources on building staff capacity by modelling and knowledge brokering amongst their staff. Schildkamp et al. (2019) discussed how leadership within schools can enable or hinder the use of data within respective units or teams using data for informed decision-making. Schildkamp et al. (2019) observed that a transformational leadership approach leads to better data use which subsequently could lead to successful changes in teaching practices within schools. According to Schildkamp et al. (2019), this is done in five ways and through the use of communities of practice known as *data teams*. First, school leaders establish a vision, norms, and goals. Second, they provide individualized emotional support to staff. Third, they promote intellectual stimulation such as sharing knowledge and providing autonomy. Fourth, they create safe climates for data use focused on improvement over accountability. Lastly, they use networking to connect different parts of the school organization by creating an internal data use network with the school.

Research Design and Methods

This study used an explorative qualitative case-study design. The study investigated emerging themes around school leadership by studying the use of an evidence-based organizational change and development approach making use of DIDM within their continuous school improvement processes.

The participants of this study included the senior and middle-level school leaders at two K-12 independent schools in Victoria, Australia. Data collection methods included semi-structured interviews with participating school leaders (N=25), field observations of DIDM activities (N=18), and institutional document analysis (N=28). The researcher used a reflective journal recording four types of reflective notes for both research sites. Data collection took place over 2017–2018. Lincoln and Guba (1985) identified four characteristics for assessing trustworthiness in qualitative research: credibility, dependability, conformability, and transferability. Member-checking of interview transcripts was employed, in which the interview transcripts were checked by the participants for clarity of meaning and accuracy and to build credibility of the datasets. Dependability was ensured by triangulating the datasets. Conformability was established through thematic analysis by comparing approaches taken by participants in their use of DIDM within their respective schools. Interviewing multiple participants at two levels of leadership at each site provided a congruence of themes. Strengthening the trustworthiness of the findings. While the findings from this study cannot be generalized for all independent schools across Australia, the similarities between both schools, such as having a comparable range of socio-educational advantage, a larger proportion of the student population having English as their main language, and similar organizational structures, i.e. primary and secondary sections, shared some similarities with other schools across the Victorian independent sector where these findings might be relevant and transferable.

Case-study school 1 was an independent, single-sex school located in the inner eastern part of Victoria, with 173 employed teaching staff and 103 non-teaching staff. The school is multi-sited. It has 800 students enrolled with 82% native English speakers and 18% non-native English speakers. The school has 88% of its parents in the upper 50% of socio-educational advantage. Case-study school 2 was an independent co-educational school located in the outer eastern part of Victoria, with 73 employed teaching staff and 52 non-teaching staff. The school is single-sited. It has 390 students enrolled with 77% native English speakers and 23% non-native English speakers. The school has 68% of its parents located in the upper 50% of socio-educational advantage and 32% in the lower 50% of socio-educational advantage.

By using a constructivist approach, the researcher investigated the social construction around DIDM within these schools as the study sought to examine the "multiple realities associated with different groups and perspectives" (Maxwell, 2011, p. 10) around actual data use and data-informed practice. According to Ponelis (2015), using a case-study approach "is particularly appealing for applied disciplines since processes, problems, and programs can be studied to engender understanding that can improve practice" (p. 536). The researcher moved from a conventional to an alternative research approach engaging in deeper critical understandings around the practice of data use in schools. The researcher used different aspects of self-reflexivity during the study to uncover more nuanced understandings around data use as discussed below.

Moving from conventional to alternative research methods – Researching data use practices

Coburn and Turner (2012) discussed three unsuccessful conventional research approaches that have failed to highlight the issues around the "actual practice" of data use, which is relevant since data use remains underutilized and under-researched in Australian schools. First, a focus on aggregate outcomes and data use through school improvement initiatives undertaken by successful data use schools. Second, a focus on data-informed activities that schools engage in where emphasis is given to the data use interventions instead of the relational aspects involved within the practice of data use. Third, research that focuses on providing an optimistic approach to data use through "how-to guides" focusing on the importance of data use. These conventional research approaches do not provide insight into the educational processes through which outcomes are produced; into educational interventions developed and implemented; or into an alternative approach undertaken based on the identification of significant contextual realities that enable or disable the practices of data use.

Coburn and Turner (2012) suggested that research should pay attention to the practice of data use through deeper investigations into understandings around what actually happens when people engage with data during their ongoing everyday schoolwork and by making connections on how this engagement relates back to instructional change, where visible shifts are found in student achievement and organizational learning. Hence, instead of focusing on an optimistic approach to data use, preference in this study was given to a realistic approach, understanding how senior and middle-level school leadership engage in understanding their data, make better decisions based on inferences drawn from the data and shift their continuous school improvement processes in the right direction. Schildkamp et al. (2017) suggested how these realistic approaches could be enacted as "practice" within schools,

> …in order for data-based decision-making to lead to school improvement in terms of increased student achievement, it is crucial that data are also used for school development and instructional purposes. Therefore, we need to study the extent to which school staff use data for accountability, school development, and instructional purposes (p. 243).

Research into these practices investigates what really happens when people at different levels within a school organization use external and internal data in their regular day-to-day practice. Conventional research approaches in Australian literature on DIDM have not discussed the reality around the kinds of data interactions and sense making that people engage in at the system or school level and how these interactions impact practice. These approaches have also not identified ways in which data is interpreted and embedded into the redesign of continuous school improvement activities. The challenges and tensions faced in these practices and how schools have overcome the challenges or have delayed using DIDM have not been sufficiently discussed. The context-specific impact analysis of data-informed interventions requires further research into why interventions might work well in one setting and yet fail to do so in another setting.

Maxwell (2012) discussed an alternative research approach that can be used to address this gap, known as "causal explanation." This provided the researcher with opportunities to better understand the contextual realities within which participants engaged or disengaged with DIDM processes and their use in their schools. In justifying this research approach within qualitative research studies like the current study, Maxwell (2012) further stated that:

> This alternative approach to causation is compatible with the practice and "theory-in-use" of many qualitative researchers and enables qualitative researchers to credibly make and support causal claims… Adequate causal explanations in the social sciences depend on the in-depth understanding of meanings, contexts, and processes that qualitative research can provide (p. 655).

Likewise, Pawson (2006) asserted that "the nature of causality in social programs is such that any synthesis of evidence on whether they work will need to investigate how they work. This requires unearthing information on mechanisms, contexts, and outcomes" (p. 25). Findings from this Australian study investigated the practice of data use in schools by looking at relationships between school leaders, their theories-in-use, and their active engagement or passive disengagement with data as they engaged in sense-making of data within their respective school contexts.

Exercising deeper researcher reflexivity – embedding four ways of reflective thinking

As this was a qualitative explorative study, the researcher was interested in studying how meanings around the practice of a data-informed approach to decision-making were developed especially within the particular social, cultural, and relational context of these case-study schools. Reflexivity is the process of examining oneself as a researcher; the research process and the research relationships develop between the researcher and the research participants. Guba and Lincoln (2005) defined reflexivity as "a conscious experiencing of the self as both inquirer and respondent, as teacher and learner, as the one coming to know the self within the processes of research itself" (p. 183).

The researcher had previously worked in both teaching and leadership roles in schools and used their own insider understanding of schools to recognize causal explanations of the practice of DIDM processes being observed. Mann (2016) discussed the importance of a reflective journal and suggested that

> keeping a journal or diary is desirable if not essential in qualitative research… The journal is a vehicle to explore a dialogue between theory, experience, and identity. It helps make explicit, your assumptions and evaluate how this shapes the interview interaction (p. 19).

Mann (2016) explained how the journal provides space for "qualitative researchers to record dilemmas, concerns, and troubling ethical questions, as well as breakthroughs and realisations" (p. 19).

The researcher used journaling to observe the processes of DIDM by using their insider-outsider perspective to make causal explanations of data use at these schools. Blaxter et al. (2001) addressed the strong reflexive approach used through a research journal by discussing four levels of reflexivity covering different pieces of the research process as well as the construction of research knowledge. These include: observational notes used to describe events such as observations and interviews; methodological notes focusing on the actions and role of the researcher; theoretical notes used to describe the preliminary understandings from the data; and analytical memos used to bring together and draw inferences after reviewing all the datasets collected, the notes, the memos, and the theoretical literature, so that the researcher works toward synthesizing patterns and themes that are emerging from the data.

The researcher used field notes throughout the interviews and school observations. Field observations led the researcher to reflect on the research relationships developed through this study; the researcher examined their relationship with the research participants, and how the relationship dynamics had an effect on the responses to interview questions. This reflexive approach helped the researcher in examining their assumptions and preconceptions and considering how these could affect research decisions, particularly the selection of research participants, research methods, research questions, and theoretical literature, as well as the overall construction of research knowledge.

The researcher used methodological notes when revising the interview questions after collecting data from the first case-study school. Through the use of semi-structured interview questions, the researcher worked at developing a context of interactive meaning-making between researcher and research participants. Observations of school improvement activities and analysis of organizational documents were useful data collection methods requiring the researcher to exercise reflexivity from the beginning of data collection. The interpretation of these qualitative datasets required the researcher to engage in regular reflection on different aspects of the research context for each school. In exercising reflexivity, the researcher made the research process itself a focus of inquiry by laying open their preconceptions of DIDM and its links with school improvement processes and ensuring that these preconceptions did not influence the findings from this study. The researcher journaled a series of methodological notes as data was collected across the two case-study schools. For example, during the data-collection stages at case-study school 1, the researcher found the term "data-driven" had negative connotations associated with it, with research participants mostly aligning the purpose of their interview with the accountability aspect of DIDM, rather than the improvement aspect. This repeated observation with participants made the researcher reconsider the term through their reflections in their methodological notes, and revise it to "data-informed" instead of "data-driven" decision making at case-study school 2.

Theoretical notes helped the researcher focus on theoretical literature from a new lens looking at the nuanced meanings behind data-driven decision-making and DIDM in the field. The researcher was careful in how theoretical literature would be positioned and used during data analysis. In discussing the practice of data use in schools, Datnow (2017) suggested that "...educators play a critical role in shaping how and why data are used, what counts as data, and so on. DIDM is a more appropriate term for this practice, rather than data-driven decision-making, used most often in the field" (p. 11). The theoretical reflexivity exercised through theoretical notes subtly moved the focus of the data analysis from studying what data-driven decision-making is and the effects it has on school improvement processes to studying how a DIDM context affects school improvement processes. Schildkamp et al. (2019) described DIDM as a new way "in which data can never completely drive decisions. Instead, data can inform decisions, which, combined with the professional knowledge of educators, can contribute to achievement and learning in schools" (p. 2).

During data collection, the researcher used NVivo and Endnote software to analyze and develop analytical memos to synthesize emerging themes within the data. Reflexivity through analytical memos facilitated the researcher in making deeper connections between theory and data. They continued to reflect upon understanding how and why the practice of data use in decision making provided a more nuanced understanding to the role of evidence-based continuous school improvement practice. Through self-reflection, the researcher worked at ensuring that they had not assumed that meanings around this nuanced understanding of the practice of data use in decision making suggested that data use was fixed, static, stable, concrete, and ready for use in any school context. They found that analytical reflexivity allowed them to continue checking and establishing that they had not developed expectations that truth could be discovered by asking the right questions, made assumptions that their questions were always objective, or assumed that participant answers were straightforward, with clear and definitive meanings and singular realities. Instead, through on-going reflexivity and using a "causal explanation" research approach, the researcher worked at understanding how all meanings were interactively and culturally constructed and how each of these research participants as individual social actors were variously located within the social settings at their respective schools. The researcher sought to understand the positioning of leaders at each of these research sites by observing the way they were positioned by their leadership position, subject expertise, gender, age and other emerging ascriptive characteristics that came through the data. This reflexivity helped the researcher to move beyond the apparent and to better understand what was obscure in the processes being examined. A cross-comparative thematic analysis was used while analyzing data from which two main themes emerged.

Findings

First, the findings brought to light patterns of disengaged versus engaged DIDM practices found at these schools and their effect on the continuous school improvement processes. This finding directly informed the researcher of the influence that an enhancement of school system capabilities has when DIDM processes are applied. Second, the findings indicated that when looking at the role and actual involvement of senior and middle-level school leaders and their influence on a data-informed school culture, various kinds of active and passive forms of leadership dispositions emerged. The findings also suggested that the importance of contextual realities within independent schools should be considered in discussions around DIDM processes since these schools operate as independent, autonomous entities led by school leadership teams and school councils.

Enhancing school system capabilities:
Dynamic versus passive engagement in data-informed decision-making processes

The findings from the two schools explored how school personnel engaged with data during the course of their ongoing everyday work and made decisions that affected the organizational culture within these schools. The data analysis focused on investigating their ways of thinking around data use, their use of school-based data management systems, their leadership practices, and their system supports that some of the organizational processes used for enhancing school system capabilities.

In both cases, school datasets were used conservatively, with school leaders and teachers unable to see the scope of using data for deeper analysis on their school practice. School leaders engaged with data in two significant ways that impacted how decision-making processes around school improvement took place. These approaches included either active, mindful collection of data for school improvement or more traditional collection of data for the purpose of recording and archiving information. The former approach was progressive and dynamic; the latter was passive and limited. The findings indicated that while huge amounts of data were collected, not all of it was actively used. As one senior leader noted,

> *One of our strengths I think is that we have lots of data, we actually are almost like drowning in data, we collect a lot of different data, whether it's questionnaires or making observations. We have a supportive principal who is really interested in data, we have some staff with experience, we have also got some interested staff, so they may not have experience, but they're interested. But there is not much use of this data except for some datasets and those are used mostly for the reason why they were collected* [CSS 1.1].

As indicated, the purpose of data collection was not always clear, nor was it always easy to know the link between the collection of data and its influence within their current school improvement agenda. At times, an absence of policies around the purpose of data collection led to datasets being collected but having no direct influence on school improvement processes. As this leader noted:

> *One of our weaknesses is that we actually don't have enough data-literate people on staff. There are only a few systems for using data in place, we've talked about that at times, but feedback cycles are either minimal or not structured enough to make better use of the data* [CSS 1.1].

This indicated that the wider usage of data was restricted by a limited understanding of the interpretation of quantitative datasets, a lack of analytical expertise in the practice of data use for school evaluation, and a lack of expertise in the analysis and inferencing of multiple datasets. This suggested that data use did not include a data-informed approach with analytical interconnections drawn amongst school improvement elements. One or two expert data-literate or data-numerate staff members were on staff and were called upon to use their data interpretation skills. A lack of proper data use or re-use of previously analyzed datasets for trend-mapping was found. As a middle-level leader from case-study school 2 mentioned:

> *We used to get the VCE [Victorian Certificate Exam] data – we did have access to the printouts, but no-one really walked us through and explained how they worked, because there's a lot of scaling and different things. And if you didn't have that background knowledge, it was difficult. So, one of the teachers in the science department, this is something he really enjoys, and so he will sit down with you one-on-one or in a meeting style and run through with you how to best utilize that data. And since that has happened in the last three years, it has been really beneficial. I can really understand the data. I get where I could have done better, or actually, where I'm doing really well, and how I can support my next lot of Year 12 students, the next year* [CSS2.2].

Due to this lack of analytical understanding around data use, there was an avoidance by participants around its use in active decision-making processes; the approach for data use was *ad hoc*, with school leaders preferring to rely on their own professional judgment and experience or that of their peers to make decisions, even though they had school datasets.

Role of leadership: Proactive versus inactive involvement

In this study, two main types of leadership dispositions towards DIDM were identified. School leaders at both levels either proactively used DIDM or avoided the use of data and played down its importance within school improvement processes, demonstrating clear inactive involvement.

At case-study school 1, a new sophisticated data management system was put in place at the start of the school year with very little time and information given to leaders at either level to help understand how they would use this system to streamline the school improvement processes they were to lead. This weak change approach resulted in limited engagement with datasets available through the data management system and disengaged decision-making processes occurring as the school underwent a problematic roll-out of this change process. A number of senior and middle-level school leaders were not professionally developed in the necessary technological skills required for this change to occur smoothly. Some leaders were not kept fully informed within the communication loop on key decisions regarding continuous online student reporting, curriculum management and student welfare systems integrated within this data management system. The findings indicated a need for stronger relationships around data use developed between senior and middle-level leaders and between middle-level leaders and teachers at these schools. One of the leaders stated:

> *I think that even having the sort of understanding that we've got data and having a group of people get together to look at the data is a new idea. While there's some privacy issues around some datasets, we've talked about de-identifying that. However, everyone sits on their data and doesn't want to discuss it together. At [CSS1], it's very much like that. Even the idea of people talking to each other about this is a bit new. It hasn't happened yet... that's part of the cultural thing – that we've worked in silos for so long and they've not really been asked to do that* [CSS1.14].

A transformational leadership approach across both levels of leadership was noticeably absent or only developing. Such an approach would facilitate building a stronger organizational culture for school improvement. The findings suggested that developing stronger organizational trust through transparent systems of planning, organizing, and monitoring could have assisted in better integration of DIDM. Also, the strengthening of communication channels between both levels of leadership was critical. As one of the leaders suggested:

> *I guess what I would love to see is that we take this opportunity to say what could we do better as a professional learning community? Let's set up time for teams to work together, because with that we're going to be achieving so much. It'll be about the learning, and [the learning management system] will be part of that. It will be about what data do we have that shows us, you know, what interventions are required? What's tracking along really well? And then from there, okay, how will [the learning management system] help us with that? I think that brings the focus back to "Well we're here as a community, we're here for our learning and teaching, we see that as a priority and we're going to provide time for that"* [CSS 1.7].

As seen in the above quote, as with any change improvement, people need to have the necessary skills required for the assigned improvement task and time assigned to study how best to implement the change, to make the change, and then to reflect and improve upon the change.

At case-study school 2, senior leaders expressed concern about the insufficient time for reflection on the enactment of data-informed continuous school improvements. The busyness of schools kept them from having more mindful time for using data effectively. This was mostly attributed to a feeling of "change-weariness" within these schools. As a senior leader discussed:

> *Schools are very busy places. I see too often things being implemented in schools that start, but never finish. And part of that reason is because something else comes along that is either more important or more urgent... I think the key to any professional development in a school environment is allowing teachers to own their areas of teaching and learning and for the data to enhance what they are doing. Developing this skill is important for all teaching staff in the school before we start using data in all our decisions…* [CSS2.4].

As this leader went on to discuss, while there were a number of datasets available for school leaders to access and use through highly sophisticated data management systems and the supports that were being provided, senior school leaders did not always make clear connections between what the data was indicating and how they could then lead others in putting evidence-based improvements into practice. In their words:

> *Because within our own school system, we have a data person who will give us data on financials, on NAPLAN, on school surveys that we are asked to do. But who will give us data on other teaching and learning areas? So as a senior-level school leader, I sit there and go, "Great, this is fantastic information, but how do we then follow up to build the school culture and what support do we have in place for that?" I don't want us misinterpreting datasets because of our busyness. Because there's always something that gets in the way. It's either something else or it's more important or it's an emergency that comes in. So, I'm all for building professional learning environments within schools, but I'm not for something new coming in all the time, on a regular basis. With the data that comes in throughout the year we need to know when and how to use it effectively. At the moment, we don't know how to do it* [CSS2.4].

As indicated above, a sense of busyness resulting in a fear of data misuse leading to incorrect interpretations and decisions makes school leaders steer away from using datasets even though these datasets could provide clear insight on their organizational climate as well as their teaching and learning processes. This fear of data misuse directly fed into reducing the organizational trust relationships within senior leaders and amongst senior and middle-level leaders, leading to a trickle-down effect of less organizational trust between middle-level leaders and their teachers. These low levels of trust were indicated through what the researcher identified as *data fortresses* built by school leaders where significant restrictions were placed around data accessibility, so that data was accessible only to a few leaders. Factors contributing to the existence of these data fortresses included an imbalance of leadership influence, a lack of horizontal and vertical organizational communication, and less collaboration amongst staff even though sophisticated data management systems had been installed at this school. A lower level of organizational trust being exercised by senior and middle-level leaders toward widespread data use led directly to a passive disengagement of teaching staff from DIDM practices.

Discussion

Due to the autonomous nature of independent schools (ISCA, 2018), the findings from these two case-study schools indicated that the principals, school boards, and leadership teams needed to work together consistently within their main decision-making processes when considering the complexities and challenges of their schools. However, as suggested by Independent Schools Queensland (ISQ, 2017), along with their autonomy as independent schools, "collaboration" within these schools was essential, since:

> Autonomy alone is no guarantee of good performance, and if the capacity for decision-making is not carefully tailored to the environment, and the needs of students, there is little gain over a highly-centralised system… The impact of school autonomy on performance is enhanced when there is a culture of collaboration between teachers and school leaders in managing a school. This is the "glue" for what makes autonomy work in terms of smart use of resources and intelligent accountability. It involves collaboration at all levels (p. 5).

At both schools, it was found that building a data-informed school culture was necessary before DIDM could be used as a continuous school improvement practice. The findings suggested that data was often used or reviewed simplistically for the purpose it was mainly collected for. This approach lacked the provision of opportunities for deeper analysis and data engagement by these schools. The use of datasets for multiple purposes or for mapping trends around cohorts, grades, or subjects was limited. Similarly, the data management systems had limited access to datasets, providing those with access to positional power over others. These leaders were limited by their own competency and propensity toward the use of data. An efficacy toward the use of data for reflection and continuous school improvement was found limited to a few aspects of school management and administration at both schools.

There was a need for leaders and teachers at each school to feel supported while developing professional learning environments where DIDM was consistently used (Schildkamp et al., 2019; Van Gasse et al., 2016; Vanlommel et al., 2016). It was found that having a strong school culture with organizational trust was essential for promoting DIDM (Fernandes, 2021), as this broke down the need for data fortresses. Holmes et al. (2013) found that the ability to build trust relationships within their schools was one of five main characteristics of effective school leaders. Holmes et al. (2013, p. 276) suggested that "the building of social cohesion and trust is a key factor in ensuring that staff are committed to working toward shared goals and ongoing effort is required to maintain these relationships over time."

In looking for causal explanations around the passive or limited data use evidenced at both schools, Marsh et al. (2006) suggested that

> … equal attention needs to be paid to *analyzing data* and *taking-action based on data*. These are two different steps: taking-action is often more challenging and might require more creativity than analysis. Yet, to date, taking-action generally receives less attention, particularly in the professional development provided to educators. School staff often lack not only the data analysis skills (e.g., knowledge of how to interpret test results), but also guidance in identifying solutions and next steps in addressing diagnosed problems (p. 10).

The findings indicated that while these schools provided substantial allocation of resources for data collection and had sophisticated data management systems, more work was needed on developing the data competence of leaders and teachers so that DIDM could be integrated into regular school improvement processes. The disengagement observed at these schools was due to a lack of emphasis given to building DIDM into the school organizational culture (Schildkamp et al., 2019; Van Gasse et al., 2016; Vanlommel et al., 2016).

Four levels of engagement in data use were found at these schools, with not all leaders ascribing to any single level but rather demonstrating variance in the level of organizational capability. This engagement directly affected how data were used and was directly influenced by senior and middle-level leadership. At Level One – Data Avoidance, the findings indicated that school leaders distrusted data and avoided using it actively or proactively when managing or leading those working with them. The use of data was very limited in such instances and was restricted to just "a chosen few" within the school system. At Level Two – Data Indifference, the findings indicated that school leaders took note of trends in datasets but did not proactively work at improving the system based on evidence from these datasets. The data was sometimes reported to others; when used to inform a limited number of decision-making processes within the school, it was mostly in line with the obvious purpose of data collection. At Level Three – Data-Based, the findings indicated that school leaders used data more proactively but only when it supported their own opinions and decisions and as a means of justification. This instrumentalist approach to data use was found amongst leaders especially when they wanted to inform their teams of decisions that were top-down in their approach. The data use in such cases was still limited as the focus was on confirming for the teaching staff and with the teaching staff the need for mandated change. This did not allow staff to interrogate the data and to establish new ways of thinking and decision making. Finally, at Level Four – Data-Informed, the findings indicated that when these school leaders used the datasets to shape and inform their decisions for school improvement, they confidently used the datasets on a regular basis and for multiple purposes being both data literate and numerate. At this stage, DIDM was firmly embedded within their leadership disposition and practice and its usage then filtered down into proactive data teams at these schools. These four levels of engagement provide insight into how these schools may continue to work at building up the capabilities of their leaders as well as organizational decision-making processes so that this variance across four levels of data use can be decreased over time. Both schools had installed advanced data management systems. However, a lack of strong transformational leadership in data use hindered these schools in having more active use of data within their school improvement approaches. The findings suggest that more work at each of these schools in building up communities of practice where their respective data teams play an active role could be transformational in their continuous school improvement processes.

Conclusion

The findings from this small-scale research study suggest that DIDM can be improved within these schools as they work at enhancing their own system capabilities so that dynamic engagement with DIDM processes is embedded in their school improvement processes (Datnow et al., 2017; Schildkamp et al., 2017; Vanlommel et al., 2018). The findings also suggest that both senior and middle-level school leaders at these schools required the right kind of transformational leadership approach and data expertise to lead a dynamic engagement of data use within their schools. These school leaders needed better understanding of datasets and of the functionality of their data-management systems (Fernandes & Henderson, 2020) to further embed this evidence-based approach to organizational change and development (Fernandes, 2019). As found within this study, data-informed school leaders can in effect build organizational trust through development of collaborative school improvement spaces where DIDM is part of the collective *thinking and working psyche* of the school.

While at national and state levels there is mandated accountability for school improvement due to the high cost of education to the nation, it would seem that autonomous schools in the independent sector need to give serious consideration to these new forms of evidence-based and data-informed accountabilities and the growing influence they have on the regular practice of school improvement within their respective schools. The exploratory results of this small-scale study provide insight into how a case could be made for similar independent schools to work on embedding the regular practice of DIDM within their school improvement processes. Through collaborative decision-making processes using datasets, these case-study schools could work at diagnosing, repairing, and improving themselves. Through further use of DIDM, these case-study schools could work to develop effective internal accountability measures that assist them in developing consistency in their school improvement processes, especially as they address some of the external pressures that independent Australian schools are facing today.

This current study recommends further research investigating how organizational trust processes can be developed by school leaders in Australian independent schools through DIDM practices, especially within active communities of practice. Another area for further research would be looking into how effective systems around data-access may contribute toward better communication and collaboration at senior and middle-level school leadership.

Funding Acknowledgement

This project was funded under the Dean's Early Career Research Project Award from Monash University, Australia, in the amount of AUD $30,336 over 2017-2018.

References

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method and practice*. Addison-Wesley.

Bernhardt, V. L. (2003). No schools left behind. *Educational Leadership*, *60*(5), 26–30.

Blaxter, L., Hughes, C., & Tight, M. (2001). *How to research?* (2nd Ed.). Open University Press.

Brown, C. (2015). Conclusion. In C. Brown (Ed.), *Leading the use of research and evidence in schools* (pp. 153–160). Institute of Education Press.

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. *American Journal of Education*, *118*(2), 99–111. https://doi.org/10.1086/663272

Cook, S. D. N., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*(4), 381–400. https://doi.org/10.1287/orsc.10.4.381

Datnow, A. (2017, August). *Opening or closing doors for students? Equity and data-driven decision-making* [conference paper]. ACER Research Conference 2017, Australia. http://research.acer.edu.au/cgi/viewcontent.cgi?article=1317&context=research_conference

Datnow, A., Greene, J. C., & Gannon-Slater, N. (2017). Data use for equity: Implications for teaching, leadership, and policy. *Journal of Educational Administration*, *55*(4), 354–360. https://doi.org/10.1108/JEA-04-2017-0040

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*(4), 1–26. https://doi.org/10.1177/016146811511700408

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision-making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2

Datnow, A., & Park, V. (2018). Opening or closing doors for students? Equity and data use in schools. *Journal of Educational Change*, *19*(2), 131–152. https://doi.org/10.1007/s10833-018-9323-6

Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: Harnessing data for school improvement*. Corwin.

Earl, L. M., & Timperley H. (2009). "Understanding how evidence and learning conversations work." In L. M. Earl & H. Timperley (Eds.), *Professional learning conversations: Challenges in using evidence for improvement* (pp. 1–12). Springer. https://doi.org/10.1007/978-1-4020-6917-8_1

Fernandes, V. (2016). Reframing continuous school improvement in Australian schools. In A. Normore, L. Long, & M. Javidi (Eds.), *Handbook of research on effective communication, leadership, and conflict resolution* (pp. 98–124). IGI Global. http://doi.org/10.4018/978-1-4666-9970-0.ch006

Fernandes, V. (2019). Investigating the role of data-driven decision-making within school improvement processes. In R. Hamlin, A. Ellinger, & J. Jones (Eds.), *Evidence-based initiatives for organizational change and development* (pp. 201–219). IGI Global. http://doi.org/10.4018/978-1-5225-6155-2.ch010

Fernandes, V. (2021). Making sense of data-informed decision making in educational contexts. In M. Khosrow-Pour (Ed.), *Handbook of research on modern educational technologies, applications, and management* (pp. 545–563). IGI Global. http://doi.org/10.4018/978-1-7998-3476-2.ch033

Fernandes, V., & Henderson, M. (2020). Activating turnaround leadership solicitation of data-informed decision-making within Australian public schools. In C. V. Meyers & M. J. Darwin (Eds.), *School turnaround in secondary schools: Possibilities, complexities, and sustainability* (1st Ed.), (pp. 59–81). Information Age Publishing.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd Ed.) (pp. 191–215). Sage.

Hansen, J. M., Gentry, R., & Dalley, T. (2003). Mindful change in a data-driven school. *Principal Leadership*, *3*(6), 37–41.

Harris, A., & Hargreaves, A. (2015). Exceptional effectiveness: Taking a comparative perspective on educational performance. *International Congress for School Effectiveness and School Improvement*, *8*, 7–40.

Harris, A., & Jones, M. (2017). Leading in context: Putting international comparisons into perspective. *School Leadership & Management*, *37*(5), 431–433. https://doi.org/10.1080/13632434.2017.1368864

Holmes, K., Clement, J., & Albright, J. (2013). The complex task of leading educational change in schools. *School Leadership & Management*, *33*(3), 270–283. https://doi.org/10.1080/13632434.2013.800477

Honig, M. I., & Coburn, C. E. (2008). Evidence-based decision making in school district central offices. *Educational Policy*, *22*(4), 578–608. https://doi.org/10.1177/0895904807307067

Independent Schools Council of Australia (ISCA). (2018). *Independent Schooling in Australia – Snapshot 2018*. ISCA. http://isca.edu.au/wp-content/uploads/2018/05/18189-ISCA-2018-Snapshot-A4-2pp.pdf

Independent Schools Queensland (ISQ). (2018). Independent Schools are more accountable than ever. *Briefings: Thought leadership for the independent schooling sector*, *22*(2), 1–8.

Independent Schools Queensland (ISQ). (2017, May). *Autonomy and school leadership: An independent schooling perspective* [Discussion paper]. Independent Schools Queensland.

Lai M., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. Lai, & L. Earl (Eds.), *Data-based decision making in education. Studies in educational leadership* (pp. 9–21). Springer. https://doi.org/10.1007/978-94-007-4816-3_2

Lai, M. K., Wilson, A., McNaughton, S., & Hsiao, S. (2014). Improving achievement in secondary schools: Impact of a literacy project on reading comprehension and secondary schools' qualifications. *Reading Research Quarterly*, *49*(3), 305–334. https://doi.org/10.1002/rrq.73

Lincoln, S. Y., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30–37. https://doi.org/10.3102/0013189X12459803

Mandinach, E. B., Parton, B. M., Gummer, E. S., & Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, *96*(5), 25–28. https://doi.org/10.1177/0031721715569465

Mann, S. (2016). *The research interview: Reflective practice and reflexivity in research processes*. Palgrave-Macmillan. https://doi.org/10.1057/9781137353368

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from Recent RAND Research* (OP-170). RAND Corporation. https://doi.org/10.7249/OP170

Maxwell, J. A. (2011). Epistemological heuristics for qualitative research. In H. Soini, E. L. Kronqvist, & G. L. Huber (Eds.), *Epistemological heuristics for qualitative research* (pp. 10–27). Center for Qualitative Psychology.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools—Teacher access, supports and use*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development. SRI International. https://files.eric.ed.gov/fulltext/ED504191.pdf

Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: district and school connections in data-driven decision-making*, School Leadership & Management: Formerly School Organisation*, *29*(5), 477–494. https://doi.org/10.1080/13632430903162541

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. SAGE.

Phillips, J. (2003). Powerful learning: Creating learning communities in urban school reform. *Journal of Curriculum and Supervision*, *18*(3), 240–258.

Ponelis, S. R. (2015). Using interpretive qualitative case studies for exploratory research in doctoral studies: A case of Information Systems research in small and medium enterprises. *International Journal of Doctoral Studies*, *10*, 535–550.

Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, *60*, 425–433. https://doi.org/10.1016/j.tate.2016.06.010

Robinson, V. M. J., & Walker J. C. (1999). Theoretical privilege and researchers' contribution to educational change. In J. S. Gaffney & B. J. Askew (Eds.), *Stirring the waters: The influence of Marie Clay* (pp. 239–259). Heinemann.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482–496. https://doi.org/10.1016/j.tate.2009.06.007

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai., & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 177–191). Springer. https://doi.org/10.1007/978-94-007-4816-3_10

Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, *28*(2), 242–258. https://doi.org/10.1080/09243453.2016.1256901

Schildkamp, K., Poortman, C. L., Ebbeler, J., & Pieters, J. M. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision-making. *Journal of Educational Change*, *20*(3), 1–43. https://doi.org/10.1007/s10833-019-09345-3

Smeed, J. L., Ehrich, L. C., Kimber, M., & Perry, L. (2011). Data-driven accountability in Australia: An unfolding story with lessons for leaders. *The Australian Educational Leader*, *33*(2), 17–20.

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*(2), 113–141. https://doi.org/10.1086/663283

Thoonen, E. E. J., Sleegers, P. J. C., Oort, F. J., Peetsma, T. T. D., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors and leadership practices. *Educational Administration Quarterly*, *47*(3), 496–536. https://doi.org/10.1177/0013161X11400185

Timperley, H. S. (2005a). Distributed leadership: Developing theory from practice. *Journal of Curriculum Studies*, *37*(4), 395–420. https://doi.org/10.1080/00220270500038545

Timperley, H. S. (2005b). Instructional leadership challenges: The case of using student achievement information for instructional improvement. *Leadership and Policy in Schools*, *4*(1), 3–22. https://doi.org/10.1080/15700760590924591

Timperley, H. S. (2006). Learning challenges involved in developing leading for learning. *Educational Management Administration and Leadership*, *34*(4), 549–566. https://doi.org/10.1177/1741143206068217

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2016). Teacher collaboration on the use of pupil learning outcome data: A rich environment for professional learning? *Teaching and Teacher Education*, *60*, 387–397. https://doi.org/10.1016/j.tate.2016.07.004

Vanhoof, J., Verhaeghe, G., Van Petegem, P., & Valcke, M. (2012). Flemish primary teachers' use of school performance feedback and the relationship with school characteristics. *Educational Research*, *54*(4), 431–449. https://doi.org/10.1080/00131881.2012.734726

Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2016). Data use by teachers: The impact of motivation, decision-making style, supportive relationships and reflective capacity. *Educational Studies*, *42*(1), 36–53. https://doi.org/10.1080/03055698.2016.1148582

Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Teachers' high-stakes decision making. How teaching approaches affect rational and intuitive data collection. *Teaching and Teacher Education*, *71*, 108–119. https://doi.org/10.1016/j.tate.2017.12.011

Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, *42*, 25–34. https://doi.org/10.1016/j.stueduc.2013.11.001

Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the Data-Informed District. *School Effectiveness and School Improvement*, *23*(2), 159–178. https://doi.org/10.1080/09243453.2011.652124

Corresponding author
Venesser Fernandes
Faculty of Education, Monash University, Melbourne, Australia
E-mail: venesser.fernandes@monash.edu

STUDIA PAEDAGOGICA

Studia paedagogica is a peer-reviewed journal published by the Faculty of Arts at Masaryk University. The editorial board is located at the Department of Educational Sciences. The publication frequency is four issues per year.

Studia paedagogica publishes articles on education, upbringing, and learning from all spheres of social life. The articles are theoretical, but mainly empirical. The journal is interdisciplinary – it covers current topics in educational research while at the same time providing scope for studies grounded in other social sciences.

The journal is indexed in international databases (DOAJ, ERA, ERIH PLUS, EBSCO, CEJSH, PKP, ProQuest, SCOPUS).

Cover design
Roman Švaříček

Language editor
Anne Johnson

Publisher
Masaryk University, Czech Republic
This issue is published in December 2021.

Print
Reprocentrum, a. s., Blansko

Circulation
280 copies

Journal website
www.studiapaedagogica.cz

Contact
Department of Educational Sciences
Faculty of Arts, Masaryk University
Arna Nováka 1, 602 00 Brno
Czech Republic
e-mail: studiapaedagogica@phil.muni.cz

The journal is a continuation of SPFFBU, U series.
ISSN 1803-7437 (print)
ISSN 2336-4521 (online)