

This item is the archived peer-reviewed author-version of:

How do teachers make sense of data in the context of high-stakes decision making?

Reference:

Vanlommel Kristin, Schildkamp Kim.- How do teachers make sense of data in the context of high-stakes decision making?
American educational research journal - ISSN 0002-8312 - (2018), p. 1-30
Full text (Publisher's DOI): <https://doi.org/10.3102/0002831218803891>

How do teachers make sense of data in the context of high-stakes decision making?

Kristin Vanlommel & Kim Schildkamp

Abstract

This study examines in depth the way teachers make sense of data in the context of high-stakes decision making, such as decisions related to promotion, retention or placement in educational tracks. Different types of data, data collected rationally and intuitively, may be used in this sensemaking process, and the same data may be interpreted in different ways by different teachers. The results of our study show that teachers base their conclusions on rational processes only to a limited extent. Teachers collect a great amount of data intuitively, they sometimes interpret data collected rationally by personal criteria and triangulate data and elaborate alternative hypotheses to a very limited extent. Since fair educational decisions are informed by a rational collection and a transparent interpretation of data, implications for theory and practice are provided.

Introduction

Within educational research and policy there is growing concern regarding how schools can provide equitable educational opportunities. Teachers need to make high-stakes decisions that have proven to be a major determinant of pupils' progress within educational tracks, as well as access to further educational opportunities (Agirdag, Van Avermaet, & Van Houtte, 2013; Allal, 1988; Brookhart, 2013). Therefore, the quality of teachers' decisions is important. However, not all teacher decisions influence pupils' educational trajectories to the same extent. As the stakes associated with a judgment go up, the need for a solid evidence-base increases (Epstein, 2008). As the stakes go up, there is also pressure to increase standardization in order to promote comparability of conclusions across pupils and occasions,

and thereby, to promote a kind of objectivity (i.e., lack of subjective judgment). As Shepard (2001) noted, standardization involves a basic matter of fairness.

Several types of theories on decision making can be used to study teachers' decision-making process. These theories often differ in the extent to which they rely on rational data-use processes, as in theories centred around data-based decision making, or intuitive processes, as in theories on intuitive expertise (Bertrand & Marsh, 2015; Blackwell, Miniard, & Engel, 2006; Evans, 2008; Hoy & Miskel, 2001; Ikemoto & Marsh, 2007; Kaiser, Retelsdorf, Südkamp, & Möller, 2013; Klein, 2008; Mandinach, Honey, Light, & Brunner, 2008; Schildkamp & Lai, 2013). Common elements that can be found in both types of theories describe how teachers collect and make sense of data to inform their judgements. However, the theories differ in their viewpoint on how these data are collected, analysed and interpreted. Most theories of data-based decision making prescribe fixed and systematic procedures in which data collection, analysis, and interpretation are driven by a deliberate and systematic use of pre-set criteria (Mandinach et al., 2008). Theories of intuitive processes highlight the importance of data that are collected through the spontaneous recognition of cues and highlight the value of the personal knowledge of experts (Klein, 2008).

Key to both processes is that the collected data are used formatively, which is called formative assessment. The field of research on formative assessment emerged based on the shared idea that educators have the responsibility to gather data on the learning process of individual pupils (Black & Wiliam, 1998; Harlen, 2005). Formative assessment involves using the data about pupils' learning processes to monitor and guide these learning processes (Van der Kleij, Vermeulen, Schildkamp & Eggen, 2015). These data can be collected more or less deliberately and systematically. The data can be collected in a non-systematic manner, for example through on-the-fly assessments, by observing pupils, and listening to them talking in a group discussion (Heritage, 2007). These modes of data collection have always been part of

teachers' work, and can be described as teachers' intuitive processes. These data can also be collected in a more deliberate and systematic matter, for example through curriculum embedded assessments (Heritage, 2007) and standardized assessments, which can be described as a more rational process.

Although in empirical research these processes are separated for reasons of conceptual clarity, intuition is not the opposite of rationality. In practice rational and intuitive processes are expected to be intertwined and mutually influence each other (Hammond, Hamm, Grassia, & Pearson, 1987; Kahneman & Frederick, 2005).

Over the past decade, there has been an growing expectation in education that teachers should deliberately and systematically use data to inform their decision making, starting from the hypothesis that rational data use enhances the quality of educational decisions, since it helps prevent and correct the possible biases associated with intuitive judgement (Earl & Katz, 2006; Park & Datnow, 2017; Schildkamp, Poortman, & Handelzalts, 2016; Kowalski, & Lasley, 2009). Moreover, data use has gained a lot of attention since a growing body of literature has shown that data use can lead to school improvement in terms of higher student achievement (e.g., Carlson, Borman, & Robinson, 2011; Lai, Wilson, McNaughton, & Hsiao, 2014; Van Geel, Keuning, Visscher, & Fox, 2016), and can contribute to equity in education (Park, John, Datnow, & Choi, 2017)

Although using data is considered to be an important way to improve education and to detect and correct the pitfalls of intuitive decisions, it is rather simplistic to expect that using data will automatically lead to decisions that enhance student learning. For one thing, teachers' data collection might not be as rational as intended by research and policy. Previous research has shown that teachers' data collection may vary from a rational, deliberate search for data to an intuitive, recognition-primed data collection (Vanlommel, Van Gasse, Vanhoof & Van Petegem, 2017).

But even if data are collected rationally, teachers still need to make sense of the data (Bertrand & Marsh, 2015). The same data might have different meanings to different teachers, or data collected rationally might be interpreted on the basis of teachers' personal beliefs. Decisions can never be completely driven by data; teachers filter data through their own lenses and experiences, and intuition also plays an important role (Datnow, Greene, Gannon-Slater, 2017). Therefore, the sensemaking process will inevitably influence the extent in which teachers' conclusions are supported by the data (Bertrand & Marsh, 2015). To be able to engage in this sensemaking process, teachers need knowledge, skills and dispositions to interpret data effectively and responsibly (Mandinach & Gummer, 2016). According to these authors, data-literate teachers continuously, effectively and ethically collect and interpret multiple sources of data to improve decision-making in a manner appropriate to teachers' professional roles and responsibilities (Mandinach & Gummer, 2016).

Sensemaking is not necessarily a rational process in which an extensive elaboration of alternative explanations based on clear criteria will lead to conclusions. Instead, teachers, as do all people, often use simpler, quick strategies that require less cognitive effort (Kahneman & Frederick, 2005). These judgmental heuristics may lead to false interpretations (fallacies) when teachers try to fit data into a frame that confirms their assumptions without searching for alternative explanations, when their conclusions are based on a limited set of data (lack of data triangulation), or when their interpretation is greatly influenced by beliefs (Hitchcock, 2017; Kahneman & Frederick, 2005; Kaufmann, Reips, & Merki, 2016). In order to prevent these fallacies, data triangulation, testing alternative explanations, and using pre-set criteria for coming to a decision are also identified as important aspects of teachers' data literacy in the interpretation phase (Mandinach & Gummer, 2016).

In education, it is important that teachers try to make high-quality decisions, as these decisions will influence students' lives, especially when the stakes are high (e.g., passing or

failing, graduating or not graduating). Sensemaking is a critical aspect of teacher judgment to consider in the light of educational decisions, but in-depth insight into how teachers make sense of data is still lacking (Coburn & Turner, 2012; Datnow, Park, & Kennedy-Lewis, 2012; Kane, 2013; Little, 2012; Spillane, 2012). An emerging field of research indicates that data use may not follow a rational model (Bertrand & Marsh, 2015; Coburn & Turner, 2011; Datnow & Hubbard, 2016; Jimerson, Cho, & Wayman, 2016; Schildkamp & Lai, 2013). On a related note, this reconceptualization acknowledges that teachers may use data in non-normative ways.’ Moreover, disparities in education are often deeply rooted in teachers’ daily classroom activities and beliefs about what it means to work toward equitable educational trajectories, which is why sensemaking needs more attention as a central process in teacher judgment (Braaten, Bradford, Kirchgasser, & Barocas, 2017). Therefore, in order to critically examine teachers’ sensemaking, the following research questions are put forward:

How do teachers make sense of data in a high-stakes decision making process?

- *What data sources do teachers use when making inferences?*
- *To what extent do teachers triangulate data when they develop inferences based on data?*
- *To what extent do teachers evaluate alternative explanations when they develop inferences based on data?*
- *What criteria do teachers use when they make sense of data?*

Theoretical Framework

Teachers’ data sensemaking

Sensemaking theory can be used to describe the process by which teachers give meaning to data. In this chain of reasoning, teachers make inferences that lead from data to

conclusions, starting from the question: what do the data tell me? (Coburn, Honig, & Stein, 2003; Weick, 1995). In this study, we will take into account a broad array of data that teachers may use when they judge pupils' competencies, quantitative as well as qualitative data, collected rationally as well as intuitively:

- Rational data collection: quantitative and qualitative data that were collected deliberately and systematically, as described in most theories of data-based decision making. Examples include assessment data, survey data, and structured classroom observations (Schildkamp & Lai, 2013; Earl & Louis, 2013; Mandinach & Jimerson, 2016)
- Intuitive data collection: spontaneous, recognition-primed data collection. Based on their experiences, teachers' attention can be drawn by certain cues they recognize in all the information that surrounds them, without a pre-set question or goal, or without a thought-out and systematic method. Examples include spontaneous observations during daily classroom activities, a talk with a student, a conversation with parents (Vanlommel, Van Gasse, Vanhoof & Van Petegem, 2017).

Independent of the rational or intuitive nature of teachers' data collection, data need to be interpreted before they can inform decisions (Bertrand & Marsh, 2015). Throughout the sensemaking process, mental models (teachers' beliefs about causal relationships) will be used to give meaning to data (Spillane & Miele, 2007). Data interpretation may unfold in different ways because teachers' process of making sense of data may be influenced by the beliefs teachers have about (groups of) pupils, by the (dis)trust they have in the data, or by personal feelings of knowing (Schneider & Ingram, 1993). Because data need to be understood by the individual teacher, individual assumptions, preferences or feelings might

lead to invalid interpretations and thus biased conclusions (Kahneman & Frederick, 2005). In order to critically examine teachers' inferences, a clear elaboration of these inferences and the criteria used is therefore needed, for which purpose interpretive arguments can be constructed by the teachers.

Teachers' interpretive arguments make explicit the teachers' inferences in a chain of reasoning that leads from data to conclusions. An inference is an explicit statement of how the teacher interprets the data with regard to pupil competencies. For example, a teacher might state that 50% on a test means that the pupil did not meet the curricular goals. A conclusion with regard to pupils' competencies, for example that he or she is not fit for a general curriculum, may be based on one or on more inferences. In this paper we will explore how teachers make sense of these data. The argument provides an elaboration of the intended interpretation of the data, and it includes the assumptions and criteria involved in that interpretation. An explicit construction of the sensemaking process and transparent criteria are needed to foster the traceability of teachers' conclusions (Cronbach, 1988; Kane, 2013).

It is suggested that, although rational models prescribe optimal procedures for coming to valid conclusions (Bosker, Branderhorst, & Visscher, 2007; Leonard, Scholl, & Kowalski, 1999), in practice people are more likely to take mental shortcuts (heuristics) to come to quick and easier conclusions (Evans, 2006; Kahneman, 2003; Klein, 2008). Heuristics can be defined as simple procedures for reaching satisfying, but possibly invalid conclusions. Teachers, who often report high work pressure, might be especially likely take mental shortcuts in order to keep moving on with their work (Ballet & Kelchtermans, 2009; Pelletier & Sharp, 2009). False inferences (fallacies) may be drawn when teachers' conclusions are not supported by the data because of a biased interpretation (Evans, 2006; Kahneman & Frederick, 2005).

We will critically study the interpretive arguments made by teachers, by examining whether and how teachers (a) triangulate data (b) consider alternative explanations and (c) use pre-defined criteria when they draw conclusions about pupils' competencies.

Data triangulations

When studying teachers' inferences, data triangulation is an important concept. Data triangulation is not only an attempt to explain the complexity of conclusions related to pupils' competencies in a more detailed and balanced way by studying them from more than one viewpoint (Cohen, Manion, & Morriison, 2008), it is also an important means to cross-check data from different sources (Cresswell & Miller, 2000). This is especially important with regard to the use of assessment data, as all educational tests have some degree of measurement error (Gardner, 2013). For example, measurement error may arise due to variation in human performance, variations in the environment within which measurements are obtained, variations in the evaluation of responses, and variation arising from the selection of the test items used (Feldt & Brennan, 1989). However not only assessment data may be subject to bias as all data are likely to have a certain degree of bias. This is why using multiple data sources is important, particularly when it comes to making high-stakes educational decisions (AERA, APA, & NCME, 2014; Gulek, 2003; Jones, 2007). Triangulation of data will help teachers make better decisions about students or programs (Gulek, 2003).

Furthermore, teachers still rely strongly on data collected intuitively during their daily classroom activities (Schildkamp & Ehren, 2013; Vanlommel, Van Gasse, Vanhoof & Van Petegem, 2017). Since these data are collected spontaneously on the basis of recognition, there is a risk of confirmation bias (Harteis, Koch, & Morgenthaler, 2008; Kahneman &

Frederick, 2005). Confirmation bias refers to the idea that once teachers have a hypothesis about pupils' competencies, they tend to interpret all data so as to confirm rather than challenge this assumption. In this manner, teachers interpret data in a way that confirms their subjective assumptions about pupils' competencies, which may lead to self-fulfilling prophecies and stereotyping (Agirdag et al., 2013; Brophy, 1983). For example, a teacher who believes that a certain pupil lacks motivation to learn might interpret a pupil asking whether it is almost lunch time as a sign of lack of interest while, in reality, the pupil is hungry because he/she left home without breakfast.

Studies have shown, for example, that students from disadvantaged home situations are more likely to be placed in (specific) lower educational tracks, despite making average test scores (Callahan, 2005; Marks, Cresswell, & Ainley, 2006; Park et al., 2017). Bertrand and Marsh (2015), for example, found that several teachers in their study attributed student results to certain unchangeable student characteristics, such as being English Language Learners (disproportionally composed of students of colour), which further hinders equity in education, as it may reinforce low expectations. Recently, data use for equity has gained increased attention as for example visible in the special issue on this topic by Datnow, Greene and Gannon-Slater (2017).

To prevent this confirmation bias, the deliberate and systematic collection of data from multiple sources is important in order to question and complement information derived from data collected intuitively (Earl & Katz, 2006; Kahneman & Klein, 2009). Using multiple data sources (triangulation) can help with addressing these false inferences, as data from one source can help confirm or disconfirm information from another (Gulek, 2003).

Alternative explanations

As stated above, 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, 2005). The focus is often on confirming hypotheses, not challenging them. Another way to tackle an invalid interpretation of data because of confirmation bias is to search for contrasting explanations that question pre-held beliefs and assumptions (Kahneman & Frederick, 2005). Assessing plausible rival explanations in order to question a priori assumptions is an important precondition for assuring validity (Onwuegbuzie & Leech, 2007). Furthermore, similar heuristics (mental short-cuts) may also lead to false causality (Kahneman & Frederick, 2005). This means that teachers make the false assumption that because there is a correlation between two variables, therefore one caused the other, without taking into account the other factors that might be involved. Again, this is an easier and quicker way to come to a conclusion without deliberate weighing of alternative explanations. For example, although there is a correlation between low results on a math test and the fact that a pupil is a non-native speaker, this does not necessarily mean that the pupils' language status causes bad mathematics results. Other alternative explanations might be possible; for example, the non-native-speaking pupil might not be familiar with the techniques used for calculation. In this case, a deliberate consideration of alternative explanations can enhance the validity of teachers' inferences.

Using pre-defined criteria

Looking at the criteria used when teachers make inferences is also important. Heuristics may lead to quick conclusions that are mainly based on feelings of knowing (i.e., personal criteria) instead of rational data analyses (i.e., pre-defined criteria) (Kahneman & Frederick, 2005). The affect heuristic is a mental shortcut in which emotional responses allow

teachers to come to a decision that feels good and therefore is considered to be the right decision (Kahneman & Frederick, 2005). In this case teachers make inferences based on feelings or personal beliefs instead of criteria based on pre-defined standards. For example, a pupil might score 60% on a standardized test; the teacher makes the inference that this grade is just a lucky shot, because the teacher feels that this pupil is not ready for secondary education. Although the data are collected rationally, the criterion that is used in the argument is based on the teacher's subjective belief, and different criteria might be used depending on the pupil. Since teacher expectations may lead to confirmation bias, as stated above, there is a need for clear, pre-defined criteria within the sensemaking process (Creighton, 2007), which means that the conclusions should be supported by data and not just by subjective beliefs.

In summary, in order to understand how teachers make sense of data, we need to critically examine teachers' interpretive arguments that specify how teachers make sense of data to reach conclusions based on the data they encounter. A first important precondition is that teachers clearly explicate the construction of and criteria used in their interpretive arguments. Subsequently, we can investigate teachers' interpretive arguments by questioning how teachers triangulate data, consider alternative explanations and use pre-defined criteria. An overview of the theoretical framework is provided in Figure 1. Although we depicted it here as a linear process, we acknowledge that in practice it is a cyclic, iterative process (Ikemoto & Marsh, 2007; Mandinach & Gummer, 2016).

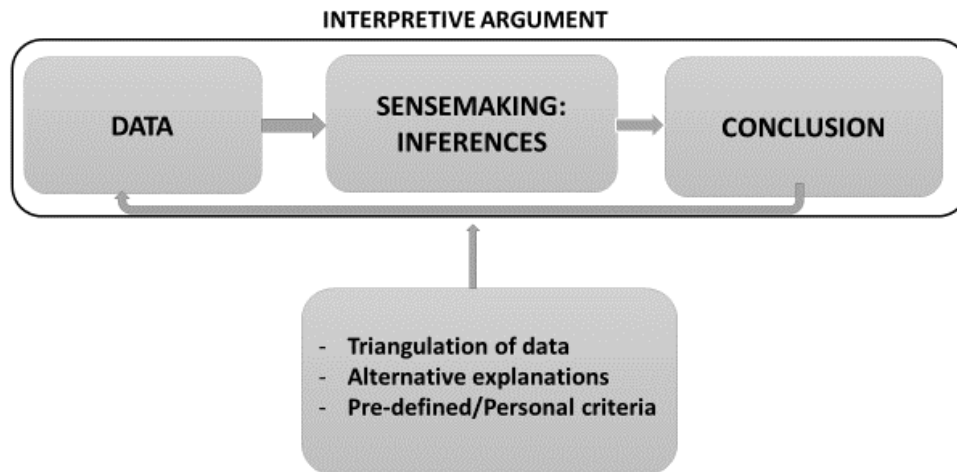


Figure 1: Overview of the theoretical framework

Method

Context of this study

Not all teacher decisions influence pupils' educational trajectories to the same extent. As the stakes associated with a judgment go up, the need for a solid evidence-base increases (Epstein, 2008). As the stakes go up, there is also pressure to increase standardization in order to promote comparability of conclusions across pupils and occasions, and thereby, to promote a kind of objectivity (i.e., lack of subjective judgment). As Shepard (2001) noted, standardization involves a basic matter of fairness.

Therefore, this study focuses on teacher sensemaking in the process where they will need to make a high-stakes decision at the end of the year, namely a transition decision that places pupils in different educational tracks. The transition from primary to secondary education involves a decision with high stakes for the pupils involved, since it is a first major transition towards a future position in society (Terwel, 2006) in which the judgment of the individual

teacher still plays a prevailing role (Eurydice, 2011). This is especially the case in the liberal and autonomous educational system of Flanders (Belgium), which does not use a binding nationwide standardized test at the end of primary school that affects pupils' future educational careers (Eurydice, 2011; Penninckx, Vanhoof, & Van Petegem, 2011). Schools can choose to make use of existing standardized tests to inform decision making, but these results are not binding for the transition decision. Mostly, teachers use or adjust tests that have been developed by the publishers of a particular educational method, or teachers make their own tests. Other examples of data collected rationally teachers may use in their decision making concerning the transition are homework, certificates of learning disorders, or class assignments.

The context of Flanders is also characterized by high decision-making autonomy for the individual teacher. The transition decision is officially a team decision, but in practice it appears that the judgment of the individual teacher is still of decisive importance (Eurydice, 2011). In Flanders, pupils typically make the transition from primary to secondary education by the age of 12. In primary education, pupils have one teacher for all subject, except gymnastics. At the end of primary education, teachers need to make the transition decision. In exceptional cases, teachers may decide not to give a certificate of primary education. In other cases, teachers will make an official transition recommendation with the following options: general secondary education (GSE, broad curriculum preparing for more demanding academic careers in university or college), technical secondary education (TSE, technical curriculum), vocational secondary education (VSE, practical curriculum) and artistic secondary education (ASE, artistic curriculum). Because of this early orientation in which pupils at a young age are already sorted into different tracks as they progress through education, the teacher's transition decision is crucial (LeTendre, Hofer, & Shimizu, 2003).

Since decisions about pupils' placement and promotion are influenced to a great deal by the judgment of the individual teacher, questioning the quality of teachers' inferences about pupils' competencies is an important matter in the light of equity and fairness (Brookhart, 2013). Therefore, we will critically examine how the teachers in our study make sense of data when they make inferences on their pupils' competencies in relation to the transition decision.

Design

We used a case study design in our study, because our focus is on understanding how teachers make sense of data, which requires an in-depth description of the underlying processes in a contextualised way (Yin, 1994). This qualitative research design allows us to gain a rich understanding of the complexity of the phenomenon in a real-life context, trying to understand the viewpoint of the teachers. A case study design is suited for investigating a phenomenon in depth within its real-life context, especially when such understanding is strongly embedded in the specific context (Yin, 1994). The central focus among all types of case studies is that the case study tries to illuminate a decision or set of decisions: why and how they were made (Schramm, 1971, in Yin, 1994). In our research, the cases being studied are the inferences teachers make when they make sense of data. Using semi-structured interviews, we seek to answer our research questions, which aim to gain a deeper understanding of teachers' individual sensemaking processes and reasoning in their specific context.

Participants

The focus of this study was on 6th grade (pupils aged 11 – 12) primary education in Flanders (Belgium). Fifty teachers were randomly selected from a list of all primary schools

in the same province. Half of the teachers were contacted by researcher 1 in a phone call in which the purpose of the interview was explained, and a total of 16 teachers voluntarily agreed to participate voluntarily. The other teachers who were called, but did not agree to participate in the interview, all argued that they did not have time to participate. About one-third (31%) of the 16 teachers were male ($n = 5$) and 69% were female ($n = 11$). The majority (56%) of the teachers had more than 10 years of experience, and the remaining 44% had between 5 and 10 years of teaching experience. All teachers signed an informed consent form stating that they had been informed about the goals of the research, that they understood that their anonymity was guaranteed, and that they could end their cooperation at any time. A descriptive overview of the participants is presented in Table 1.

Table 1 *Descriptive Overview of the participants*

	Teacher	Gender	Years of Teaching Experience
1	Emma	female	11
2	Frank	male	32
3	Bart	male	15
4	Roy	male	32
5	Amy	female	13
6	Ann	female	8
7	Joyce	female	5
8	Peter	male	4
9	Sophie	female	8
10	Bob	male	19
11	Julie	female	29
12	Lisa	female	7
13	Mary	female	25
14	Pamela	female	30
15	Liz	female	7
16	Katy	female	8

Interviews and Procedure

All participants were interviewed at the end of the school year, when they had to make the transition decision. Participants answered open-ended questions that explored their judgements about pupils' competencies in relation to the transition from primary to secondary education. Examples of questions include: "What are your arguments for this transition advice?"; "What is the evidence for this argument?"; "How did you make sense of this evidence?". All teachers discussed a transition problem involving 2 specific pupils, which meant that a total of 32 cases were discussed. At the start of the academic year, each teacher chose (a) one pupil of whom the teachers expected that he/she would not be able to make the transition to general secondary education at the end of the year and (b) a more difficult case, a pupil for whom the teacher found it hard to know in which direction the pupil would evolve during the year to come. In this study, we investigated the interpretative arguments and the data that were used to underpin teachers' advice and how they made sense of the data. The open-ended questions in the interview protocol addressed all the concepts discussed in the theoretical framework, ensuring that all of the relevant conceptual topics were asked about across all interviews.

The in-depth interviews lasted for an average of one hour and were conducted by a single researcher. The same interview protocol was used in all 16 interviews to ensure methodological consistency (Cohen et al., 2008). All of the interviews were digitally audio-recorded and the files securely saved for reasons of reliability (Cohen et al., 2008). Peer-debriefing sessions (investigator triangulation) were then conducted, in which the different methodological choices, data analysis procedures and interpretations were critically examined (Cresswell & Miller, 2000). With the aim of enhancing the reliability of our research, we clearly described our chain of evidence, so that the external observer can trace back the steps in either direction (from conclusions back to research questions or from questions to conclusions).

Coding and analysis

The interviews were transcribed verbatim and analysed with the aim of capturing variation across teachers in types of data used, in teachers' inferences and in their conclusions. In step one, researcher 1 developed a coding scheme based on the theoretical framework. Subsequently, this coding scheme was discussed with researcher 2. After both researchers had come to an agreement on the content of the coding scheme, first one interview was analysed and discussed by both researchers. This discussion stressed the need for a better conceptualisation of what was meant by 'pre-defined' versus 'personal' criteria and it appeared that intuitive data could not be interpreted by pre-defined criteria. Therefore, it was agreed that only data collected rationally would be coded as to the extent to which they were interpreted by pre-defined or personal criteria. Subsequently, the same interview and two other randomly selected interviews were analysed by both researchers using the revised coding scheme, and the inter-rater reliability (Cohen's Kappa) was found to be 0.72 (Miles & Huberman, 1994). Disagreements in the codings by both researchers were resolved by discussing and reflecting on the content of the different concepts and its boundaries. No additional revisions were made to the coding scheme at this point. In the last step of the coding process, researcher 1 went back and re-analysed the interviews that had been analysed before the inter-rater reliability check and finally all interviews were analysed based on the revised coding scheme by researchers 1. An overview of the codes is provided in Table 2.

Table 2 Overview of the coding scheme

HEADCODE	SUBCODE	Conceptual characteristics	Example
1. Data		Quantitative and qualitative cognitive and non-cognitive indicators related to the pupil	Test results, conversations with pupils, observations, pupil files
	<i>1.1 Rational</i>	Collection started from a clear goal or question and was carried out according to a thought-out plan or method	A test taken to measure a specific curricular goal.
	<i>1.2 Intuitive</i>	Collection did not start from a clear goal or question, nor was it carried out according to a thought-out plan or method. Indicators are collected spontaneously based on the recognition of a cue.	During daily classroom activities, a teacher notices that a pupil cannot sit still for more than 10 minutes.
2. Inference		An explicit statement of how the teacher interprets the data with regard to pupil competencies.	‘He scores 49% on his test for mathematics, this means that he scores below average.’
3. Conclusion		Teachers come to a determination of a certain aspect of pupils’ competencies, based on the inference.	‘He does not reach the curricular goals for mathematics.’
4. Alternative explanation		A statement that explicates another possibility for making sense of the same data.	(He does not do his homework because he is not interested in school.) First, I thought it was because his parents were not involved, but then I had a conversation with the parents and it appeared they try to motivate him as much as they can.
5. Criteria	<i>5.1 Pre-defined</i>	Based on clear, measurable and shared (school level) goals.	49% on her standardized test is in the E-zone, a pupil in this zone is not allowed to go to general secondary education.
	<i>5.2 Personal</i>	Based on teachers’ personal feelings or beliefs.	I don’t believe a pupil with 65% on mathematics will make it in general secondary education.

Results

In this research, we wanted to understand how teachers make sense of data in a high-stakes decision process by investigating: (1) what data teachers use when they make inferences, (2) to what extent they triangulate data (3) to what extent they evaluate alternative explanations and (4) what criteria they use when they interpret data. In this manner, we wanted to explore how teachers make inferences on pupil competencies when they make sense of data. Table 3 provides an overview of these elements for all teachers in our interviews. Only unique citations were counted; for example, if a teacher mentioned four times that a pupil is bad at mathematics because he/she failed a test, this data source and this inference were only counted once. If teachers mentioned the triangulation of data (at least two data sources collected rationally, potentially supplemented by data collected intuitively) to underpin one of his/her conclusions, this is shown by (+) in table 3. In the next paragraphs we will discuss our findings in depth.

Table 3 *Overview of the different codes for all teachers*

Teacher	Data collected rationally	Pre-defined criteria**	Personal criteria**	Data collected intuitively	Inferences Total	Alternative explanation	Triangulation
Emma	11	10	1	3	14	2	+
Amy	9	7	2	4	13	1	+
Ann	6	4	2	4	10	1	+
Lisa	18	12	6	11	29	1	+
Frank*	1	1		7	8		
Roy	4	2	2	13	17	1	
Joyce	7	6	1	10	17	1	+
Sophie	5	3	2	10	15		
Bob	5	4	1	12	17		
Julie*	1	1		9	13		+
Mary	5	2	3	7	12		
Pam	3	2	1	7	10		
Liz	2	2		10	12		
Katy	4	4		7	11		+
Bart	3	2	1	4	7		
Peter	4	1	3	4	8		
TOTAL	91	69	22	122	213		

(*discussed only 1 pupil, since 1 pupil left school) (**only applicable to data collected rationally)

First, when we study Table 3 we see that the inferences teachers make range from 7 to 29 and that all of the inferences are based on data collected rationally or intuitively to a certain extent. Lisa, the teacher with most inferences (29) based her inferences on both deliberate and systematic (18) as well as non-deliberate and non-systematic data collection (11), but almost exclusively relies on pre-defined criteria to make sense of the data collected rationally (12/18). In our interviews, the teachers who use the least inferences are Frank (8), Bart (7) and Peter (8). Frank, Bart and Peter differ into the extent in which they collect data deliberately and systematically, but they all predominantly rely on personal criteria when they make sense of data. In the following paragraphs we will discuss teachers' evidence base and sensemaking process with regard to pupils' competencies more in-depth.

What data sources do teachers use when making inferences with regard to pupil competencies?

In order to answer research question 1, we will first discuss the extent to which teachers use data collected deliberately and systematically when making inferences with regard to pupil competencies.

Data collected deliberately and systematically

In our interviews 91 out of the 213 inferences (43%) were based on data collected deliberately and systematically. Rational data collection predominantly referred to cognitive output indicators, mostly test results. Mostly, these were tests based on a teaching method or developed by the teacher, although in some cases teachers referred to the results of standardized tests. To a lesser extent, rational data collection referred to home or class assignments and project work. However, teachers strongly differed in how transparently and precisely they described these data collected deliberately and systematically. One group of teachers referred to specific test results for a specific subject matter, for example '60% on a

test French vocabulary' (Amy), whereas most teachers referred to data collected rationally in a holistic way, for example a pupil has '*bad test results*' (Peter).

Data collected non-deliberately and non-systematically

The results of our study further showed that 122 out of 213 inferences (57%) were based on data collected intuitively. This spontaneous, recognition-primed data collection predominantly referred to observations during daily classroom activities, with regard to non-cognitive indicators such as motivation, attitude and wellbeing. Sometimes, observations informed teachers about practical competencies with regard to the transition decision. For example, when Sophie noticed that a pupil was handy when he helped her repair the computer, Sophie used this in her argument for technical/vocational education. Furthermore, intuitive data also referred to spontaneous conversations with parents, pupils or colleagues. Also, when it came to the description of these intuitive data, teachers differed in the transparency and clarity with which they described these intuitive data sources. Some teachers said that '*when I see how he acts, I just know that he is not motivated for school*' (Peter). Other teachers described concrete cues that led to a conclusion, as for example when Frank explained: '*She started crying during this assignment for mathematics. It was a really difficult exercise where all things came together, it showed me that it became too much for her.*'

Two categories of teachers

Based on the results presented in table 3 we saw two main categories of teachers emerging: (a) teachers who based more than half of their inferences on data collected rationally and (b) teachers who based more than half of their inferences on data collected intuitively.

Only four teachers, Emma, Amy, Ann and Lisa, predominantly used **data collected rationally** in the decision process, referring to a wide array of test data that provided insight into how well pupils had mastered different parts of the curriculum. All of these teachers described the data they used in clear and transparent terms, referring to specific subject matter. For example, Lisa described the results in specific ways as she explained: *'On his standardized test, he scored average for writing, but he failed for listening skills. He failed grammar exercises too. He failed on his final test for French, but he did well on geography.'* These data collected deliberately and systematically were complemented by data collected in a non-deliberate and non-systematic manner, although these data were used to a lesser extent. These data collected intuitively, mainly observations and conversations with parents and pupils, predominantly referred to general cognitive indicators and attitude. These data did not necessarily confirm the data collected rationally. For example, although the test results of a non-native-speaking boy were low, Emma said: *'based on the way he answers my questions during class, I can tell he has competencies.'*

The largest group of teachers (10 teachers: Frank, Roy, Joyce, Sophie, Bob, Julie, Mary, Pam, Liz and Katy) based more than half of their inferences on data they collected spontaneously, in a non-deliberate and non-systematic manner during daily classroom activities. Most of these inferences were based on observations with regard to non-cognitive indicators such as engagement, motivation, home situation, interest and wellbeing. To a lesser extent, teachers mentioned observations with regard to general and practical competencies. Data collected intuitively were complemented by data collected rationally to a certain extent. However, in many interviews, test results were described in more general terms, referring to all tests, or all grades for this particular pupil. For example: *'I will not recommend him for general secondary education because of his results.(...) On what this is based? What I mean? Of course this is based on all his tests during the year...'* (Pam). In a minority of the interviews,

teachers referred to (different) subject matters. Katy, for example, explained: *'He scores below average for mathematics as well as Dutch.'*

Two teachers (Bart, Peter) could not be placed in either one of these categories since they made almost equal use of data collected rationally and intuitively. However, Bart and Peter also differed from the others in that they used fewer inferences than the other teachers to reach a conclusion. For both teachers, the inferences were almost equally based on data collected rationally and intuitively. Both teachers described grades in a very holistic way, as *'the grades are not great, but good enough'* (David). In both interviews, teachers inferences were also based on non-deliberate and non-systematic observations with regard to non-cognitive indicators.

Data that are collected need to be interpreted by the teacher before they can be used in the decision process. In the following paragraphs, we will study (a) to what extent teachers triangulate data, (b) search for alternative explanations and (c) use pre-set criteria when they make sense of the data they collected.

To what extent do teachers triangulate data when they develop inferences based on data?

In order to answer this research question, we will critically investigate to what extent teachers use multiple data collected either intuitively or rationally or to what extent teachers triangulate both data collected rationally and intuitively.

Teachers using multiple data collected intuitively

We critically examined to what extent teachers used more than one data source to underpin a conclusion with regard to pupil competencies. The type of data triangulation we saw most often in the interviews was when teachers used multiple data sources collected intuitively. Teachers often used non-deliberate, non-systematic observations on (mostly non-

cognitive) aspects of the pupil to underpin a conclusion. For example, teachers used observations with regard to motivation and the situation at home to underpin the conclusion that a pupils' personal circumstances inhibited or promoted a transition to general secondary education. We also often saw that teachers used a combination of observations and conversations with parents, pupils or colleagues to underpin their conclusions.

Teachers using multiple data collected rationally

We also found examples of triangulation in which one source of data collected rationally was complemented with other data collected rationally. This was mostly the case when teachers based their conclusions on test results from different subjects, or on different test results from the same subject. Sometimes, cognitive output indicators were complemented with cognitive input indicators that were consulted in pupils' files (e.g., learning disorders) leading to for example a conclusion that a pupil did not reach the curricular goals. Furthermore, all four teachers who predominantly made inferences on the basis of data collected rationally also triangulated by combining two or more rational data sources in their judgement.

Teachers using both data collected rationally and intuitively

A third form of data triangulation we encountered in the interviews was when data collected intuitively were complemented by data collected rationally. Most often, multiple sources of data collected intuitively were complemented by only one data source collected rationally (test results). For example, when the teacher observed a lack of motivation, and when the parents told the teacher that the pupil refused to do his/her homework, combined with low test results, the teacher concluded that the pupil lacked a proper attitude with regard to the transition. Mostly, we found that data collected rationally were explained in a way that made them coincide with the data collected intuitively. For example, the low scores were seen

as a result of the wrong attitude. So, in these examples, data collected rationally were not used to question data collected intuitively, but rather to confirm it. For example:

'I have seen that he is at his wit's end, he is not ready for general secondary education. He will pass his final test for Dutch, true, but it's not really that difficult. He just passed with 50% while other pupils easily score 90%.' (Sophie)

When we searched for data triangulation in which at least two different data sources collected rationally were mentioned, potentially complemented by one or more data sources collected intuitively, these examples were less frequent. The teachers within the rational data user category mostly triangulated on the basis of multiple data sources collected deliberately and systematically. From the intuitive data user category, only Joyce, Julie and Katy mentioned multiple data sources collected rationally that underpinned their judgement. Their conclusions were, for example, based on test results for languages and technical assignments, a certificate of dyslexia and information about the child's interests that was derived from work on a project. These data collected rationally were combined with observations and conversations to underpin the conclusion of Joyce.

To what extent do teachers search for alternative explanations when they develop inferences based on data?

We investigated the extent to which teachers consider alternative interpretations when they make sense of one specific data source. In our study, we found little evidence of alternative explanations. Few teachers described how they question information that is in the pupil's file, for example, when the teacher from the previous year has written down that a pupil lacks motivation. Some teachers described how they use their personal knowledge of this colleague to interpret the information. For example, Emma explained:

'The pupil's file said that Joanna did not have the capacities needed for 6th grade and that she would not be able to get a certificate at the end of primary education. I know I have another

approach than the teacher from 5th grade, I can imagine that her approach didn't work, so I wanted to try if I could get more out of Joanna.' So, Emma questions the data she found in the pupils' file.

So, in the limited examples we encountered, the search for alternative explanations was mainly guided by data collected intuitively. Only Roy and Lisa described a deliberate and systematic search for an alternative explanation. For example, for Roy, at first it appeared that his pupil, Rosemary, did not meet the curricular goals for mathematics. Then he administered an extensive test which showed that she could not count. When she was allowed to use a calculator, she passed all other areas of mathematics, such as geometry and applied math. Although the use of alternative explanations was limited, it appeared that all teachers who mostly used data collected rationally, also searched for alternative explanations when they interpreted data.

What criteria do teachers use when they make sense of data?

Teacher who mostly use pre-defined criteria to make sense of data collected rationally

Data collected rationally can be interpreted by pre-defined criteria that refer to clear and specific measures, but also by personal criteria that refer to subjective beliefs. When we looked at the interviews, 91 out of the 213 inferences were based on rational data collection. Of the 213 inferences 69 (32%) were based on data collected deliberately and systematically and also used pre-defined criteria.

In our results, we saw that Emma, Amy, Ann and Lisa based more than half of their inferences on data they collected rationally. When we zoomed in on the criteria they used, we saw that these teachers also predominantly used pre-defined criteria to make sense of the data collected rationally. For example, they described in specific terms how test results were below

average, how curricular goals were not met based on fixed standards associated with standardized tests, how a pupil failed (below average) on a specific subject matter, or how an increase/decrease in grades could be seen based on the change in test results. An example of this was given by Lisa:

'On his standardized test, he is in the D3-zone. It is a fixed standard, pupils in the D3 and E zone did not reach the curricular goals. On our final test, he has an average score of 55%. His reading level has increased with 6 levels during the past year, that is extraordinary, but they need level 6 to go to general secondary education.' In this school, it is agreed among the teachers that all pupils need to reach level 6 to go to general secondary education. So, Lisa uses a pre-defined criteria to make sense of the results of the standardized test.

We also found examples of broader and more holistic expressions of pre-defined criteria. This is especially so for teachers who collect little data rationally, but if they do, they mainly use pre-defined criteria to interpret these data. For example, Joyce, Sophie, Bob, Julie, Pam, Liz and Katy used pre-defined criteria to interpret the rational data in more than half of the cases. For example, Sophie explains: *'His results for Dutch are low.'* (Joyce), *'He did not pass for mathematics.'*

Teachers who mostly use personal criteria to make sense of data collected rationally

In our research, we also found teachers who predominantly used personal criteria to make sense of data collected rationally. For example Roy and Mary, collected little data rationally, but when they did, they mainly used personal criteria to interpret the data collected rationally. For example, as Mary puts it: *'Her percentages are average, but that is too weak to make it in general secondary education.'* These criteria are based on teachers' personal feelings or beliefs about what matters most for future success in secondary education.

Teachers for example described how the interpretation of grades needs to take into account work ethics or engagement. For example, Peter explains: ‘Yes, you might say that 55% on his final test is okay, but I know he didn’t work for it, so for me, that 55% is not good enough.’

In summary, we saw that teachers differed in the criteria they used to make sense of data collected rationally. The category of teachers who predominantly used data collected rationally also mainly used clear and transparent pre-defined criteria for their interpretation. The second category of teachers who mainly used data collected non-deliberately and non-systematically can be divided into two groups. One group used little rational data, but when they did use rational data, they mainly used pre-defined criteria to make sense of these data. A second group of teachers mainly used personal criteria to interpret the little rational data they used for the transition decision.

Taking all this together, we found some teachers who approached high-stakes decision making very rationally because they predominantly used rational data interpreted by pre-defined criteria, they searched for alternative explanations to some extent, and they triangulated data.

A second group of teachers took a mixed approach. They predominantly used data collected intuitively, however, this was complemented by rational data that were interpreted by pre-defined criteria to a certain extent. These teachers did not search for alternative explanations, nor did they triangulate data.

A third group of teachers approached high-stakes decision making very intuitively, since they mainly used data collected intuitively, they mostly interpreted rational data by personal criteria, they did not search for alternative explanations, nor did they triangulate data.

Conclusion and discussion

With the aim of understanding and enhancing the quality of (high stakes) educational decision making by teachers, this study investigated teachers' sensemaking process. In our study, two categories of teachers emerged for the way they made high-stake decisions. One group of teachers followed rational data use processes, as they mainly collected data deliberately and systematically, used pre-defined criteria for their interpretation, triangulated data and searched for alternative explanations. However, the largest group of teachers based their conclusions on intuitive processes of data use, in which data were mainly gathered spontaneously and recognition-primed, without triangulation or consideration of alternative explanations.

Starting from the idea that rational data use is a valuable contribution to high-stakes decision making, our study shows that in practice rational data use is still limited. The results show that teachers' inferences are only based in part on data collected rationally; teachers still collect data intuitively to a great extent when they make decisions. The data that are collected rationally are interpreted by pre-defined criteria to a certain extent; however, a significant part of rational data is also interpreted by personal criteria. Furthermore, we found little proof of data triangulation and consideration of alternative hypotheses when teachers make sense of data. Finally, the conclusions in this study were not necessarily supported by the rational evidence base that was collected. Some teachers interpreted data collected rationally in a way that made these cognitive indicators coincide with non-cognitive indicators collected intuitively. Therefore, stressing the importance of data use alone is not enough if we want to improve decision making in education. It is also important to confront and change certain beliefs about student abilities (Park, Daly, & Guerra, 2012).

The fact that some teachers based their conclusions ultimately on data collected intuitively, often related to non-cognitive pupil competencies, sometimes even despite test results that indicated the contrary, is worrisome, especially from an equity perspective. Confirming the

result from previous studies (e.g., Bertrand & Marsh, 2015; Kaiser et al., 2013; Urahne, 2015), we found that teacher beliefs about student motivation and work ethics also influenced their judgement. Some teachers believed that high engagement would overcome weak results, while teachers' came to more negative conclusions for pupils with these same or sometimes better results but low (perceived) engagement. Teachers inaccurately base part of their judgment of student achievement on students' perceived behavioural engagement in the classroom, as they assume that high engagement and high achievement go hand in hand (Kaiser et al., 2013; Urahne, 2015). However, as Urahne (2015) pointed out, teachers who perceive pupils as less motivated often underestimate students, which means that these pupils incorrectly end up in lower educational tracks. Studies also have shown that teachers overestimated the academic abilities of pupils when they were perceived as independent and interested (Alvidrez & Weinstein, 1999) or easy to manage during lessons (Hinnant, O'brien, & Ghazarian, 2009). Thus, when teachers assume correlation of non-cognitive data collected intuitively and actual achievement when making conclusions with regard to the transfer, they may assign pupils to the wrong educational tracks.

The results of our study also show that teachers use different criteria and thresholds for different students in their classroom, influenced by their beliefs about these non-cognitive indicators. Whereas one student who passes a test threshold of 50% is perceived to be ready for a higher track, another student with the same test threshold is considered to be not. Oláh, Lawrence, and Riggan (2010) also found that these types of achievement thresholds differed considerably, and varied by student, by class, and even by time. Therefore, especially when making high-stakes decisions, it is important to develop pre-defined thresholds and criteria beforehand. To validate teachers' conclusions with regard to pupil competencies, it is important to make these thresholds and criteria, as well as the whole sensemaking process,

more public, transparent, traceable and reproducible (Cohen et al., 2008; Kane, 2013; Senge, 2001).

In conclusion, teachers use the majority of data collected intuitively when they make high-stakes decisions, such as the transition decision under study. However, given the high stakes for the pupils involved, decisions with regard to placement and promotion require deliberate and systematic processes of data collection, analyses and interpretation to counterbalance the pitfalls of intuitive judgment (Blackwell et al., 2006; Kahneman & Frederick, 2005).

Furthermore, data use is often considered to be a straightforward process, without sufficient attention to the complexity of the sensemaking processes and teacher beliefs that influence decisions. Having good data does not lead to good decisions when the sensemaking process is biased. An important theoretical contribution of this study is the finding that data that were considered to be part of the cycle of data based decision-making, are less rational than intended. For example, qualitative data, such as observations that is often considered to be part of data use is predominantly collected intuitively. Further, even data that is collected rationally, may not lead to data-based decision making. The results of our study show that often even data that is collected in a deliberate and systematic manner is interpreted using personal criteria to come to a conclusion, making the whole data use process less objective than is desirable, especially from an equity perspective. Fair educational decisions require deliberate, systematic and transparent decision-making processes in which teachers reflect upon data, triangulate data collected rationally and intuitively and elaborate alternative hypotheses.

Limitations and suggestions for future research

Although our findings are important for gaining a deeper understanding of teachers' decision making, we do have to acknowledge some limitations of this study. First, the choice of a small-scale qualitative study in one specific, low-accountability (no central exams, no

obligation to use standardized tests) context implies that we need to be careful with generalisations of our findings. Replication of our findings in other contexts, especially high-accountability contexts, is needed. Although replication studies are often viewed as unoriginal (Lindsay & Ehrenberg, 1993, in Makel & Plucker, 2014), and are not seen as contributing much to the field (Sterling et al., 1995, in Makel & Plucker, 2014), they are needed to develop a robust knowledge base on what works in education, and the conditions under which it works (Granger & Maynard, 2015; Makel & Plucker, 2014).

Unfortunately, we can only critically discuss the processes of teacher judgment, but we cannot evaluate the quality of the conclusions being made. Several decisions were based on data collected intuitively, and influenced by teachers' deficit beliefs about non-cognitive student competencies. However, the next question that needs to be answered is, how do these decisions work out for the students in question? Longitudinal research in which pupils are followed in the different educational tracks to which they were assigned would be needed to answer this question.

In this research, we focused on a specific type of decision that has a discrete set of alternative options related to the transition. Possibly, the decision process differs when teachers are making more open-ended types of decisions based on data, such as how to adapt one's instruction to pupils' individual needs. Also, the decision under study involved high stakes for the pupils, but not for the teachers themselves. For future research, it is interesting to examine if and how teachers' approaches to decision making differ based on the decisions.

We only discussed the criteria teachers used for the interpretation of rational data. Traditional viewpoints that start from a dichotomy between objective and subjective are not suited for understanding the quality and interpretation of data collected intuitively. Objective interpretation of data collected intuitively would be an oxymoron. Since teachers still use a the majority of data collected intuitively to inform their judgment, the question is what criteria

can be used to adequately assess the quality and interpretation of these data collected intuitively. Also, for future research it is interesting to investigate how the criteria teachers use and the level of specificity of their inferences might be related to teachers' data literacy. Finally, our research did not involve novices because we aimed at studying the intuitive processes of data collection from the field of intuitive expertise (Klein, 2008). This means, however, that we have no insight into processes that underlie the judgement of novices. For further research, it would be interesting to study if and how novices and expert teachers differ in the rational and intuitive processes they use in judgement. According to decision theory, intuitive processes can only be used as reliable and skilled expertise in judgement when a professional had enough practice in a similar environment and with similar cases (Kahneman & Klein, 2009). From an educational perspective, this would imply that novices need to rely on predominant rational processes to prevent judgemental bias, because they lack expertise. For future research, this is something that clearly needs to be investigated.

Implications for practice

School leaders can influence data use processes by teachers (e.g., Bertrand & Marsh, 2015; Datnow & Park, 2014; Halverson, Grigg, Prichett, & Thomas, 2007; Knapp, Copland, & Swinnerton, 2007; Datnow et al., 2012). Bertrand and Marsh (2015) suggested that school leaders should encourage teachers to reflect on their sensemaking process and attributions. Moreover, school leaders need to confront cultures of low student expectations for specific sub-groups of students. The focus should be on ensuring more equal student placement (Datnow et al., 2012). Furthermore, school leaders can stress the importance of collaboration around the use of data, and data sharing (Datnow et al., 2012). By collaborating around data

use and sharing data, the decision-making process of teachers can be more public, transparent, traceable and reproducible. The latter is crucial, as teachers' often long-held implicit assumptions about student ability levels and capacity for learning need to be made explicit in order to create more equitable outcomes (Datnow et al., 2012).

Following a systematic collaborative cycle of inquiry might overcome the pitfalls of individual intuitive judgement, since it forces teachers to share, reflect and discuss their beliefs, the inferences they make and the criteria that are used for the decision. It is crucial that teachers explicitly discuss their personal beliefs with colleagues and come to a shared understanding of decision criteria that will be used to evaluate alternative options.

Furthermore, data need to be triangulated and alternative options should be discussed to challenge personal assumptions. Where individual teachers often struggle to analyse and interpret data, collaboration is considered to solve these problems (Hubbard, Datnow, & Pruyne, 2014; Van Gasse, Vanlommel, Vanhoof & Van Petegem, 2016). Collaboration incorporates support and mutual reflection among teachers when making sense of data, alignment in and transparency of decision criteria, and a shared responsibility with regard to the high-stakes decisions (Datnow et al., 2012; Jimerson et al., 2016; Mandinach & Jimerson, 2016).

Finally, although the use of rational data is crucial, especially in the context of high-stakes decision making, this does not mean that data collected intuitively do not serve a purpose in education or that non-cognitive outcomes do not matter. These process data spontaneously collected during daily classroom activities are important as well (Yan & Cheng, 2015), and can be seen as an important part of what is called Assessment for Learning. Assessment for Learning involves continuous data collection during daily classroom activities, for example, through dialogues and observations (Klenowski, 2009). Based on these (often) intuitive data sources, feedback is used to direct further learning (Stobart, 2008). As the feedback loops are

short, and the stakes are low, this type of intuitive data use process can serve the purpose of constantly monitoring and improving the quality of instruction and learning in the classroom.

References

- Agirdag, O., Van Avermaet, P., & Van Houtte, M. (2013). School segregation and math achievement: A mixed-method study on the role of self-fulfilling prophecies. *Teachers College Record, 115*(3), 1-50.
- Allal, L. K. (1988). Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education, 4*(1), 41-51.
- Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731-746.
- American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ballet, K., & Kelchtermans, G. (2009). Struggling with workload: Primary teachers' experience of intensification. *Teaching and Teacher Education, 25*(8), 1150-1157.
- Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal, 52*(5), 861-893.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.
- Blackwell, R., Miniard, P., & Engel, J. (2006). *Consumer behavior*. Mason, OH: South-Western Thomson Learning.
- 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.

- Braaten, M., Bradford, C., Kirchgasser, K. L., & Barocas, S. F. (2017). How data use for accountability undermines equitable science education. *Journal of Educational Administration, 55*(4), 427-446.
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice, 20*(1), 69-90.
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631-661.
- Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305-328.
- 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. *Educational Evaluation and Policy Analysis, 33*(3), 378–398.
- Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What is the evidence on districts' use of evidence? In J. D. Bransford, D. J. Stipek, N. H. Vye, L. M. Gomez, & D. Lam (Eds.), *The role of research in educational improvement* (pp. 67-86). Cambridge: Harvard Educational Press.
- Coburn, C. E., & Turner, E. J. (2012). The practice of data use. *American Journal of Education, 118*(2), 99-111.
- Cohen, L., Manion, L., & Morriison, K. (2008). *Research methods in education*. New York: Routledge.
- Creighton, T. B. (2007). *School and data: The educator's guide for using data to improve decision making*. London: Sage.
- Cresswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130.

- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. Hove: Psychology Press.
- Datnow, A., & Park, V. (2014). *Data-driven leadership*. San Francisco: Jossey Bass.
- 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.
- Earl, L., & Katz, S. (2006). *Leading in a data-rich world*. Thousand Oaks, CA: Corwin Press.
- Epstein, S. (2008). *Intuition from the perspective of cognitive experiential self-theory*. New York: Erlbaum.
- Eurydice. (2011). *Grade retention during compulsory education in Europe: Regulations and statistics*. Retrieved from:
http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/126EN.pdf
- Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378-395.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: American Council on Education and Macmillan.
- Gardner, J. (2013). The public understanding of error in educational assessment. *Oxford Review of Education*, 39, 72-92.

- Granger, R. C., & Maynard, R. (2015). Unlocking the potential of the “what works” approach to policymaking and practice improving impact evaluations. *American Journal of Evaluation, 36*(4), 558-569.
- Gulek, C. (2003). Preparing for high-stakes testing. *Theory into Practice, 42*(1), 42-50.
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership, 17*(2), 159-194.
- Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *Transactions on Systems, Man, and Cybernetics, 17*(5), 753-770.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *Curriculum Journal, 16*(2), 207-223.
- Harteis, C., Koch, T., & Morgenthaler, B. (2008). How intuition contributes to high performance: an educational perspective. *Education Review, 5*(1), 68-80.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140-145.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*(3), 662-670.
- Hitchcock, D. (2017). Do the fallacies have a place in the teaching of reasoning skills or critical thinking? In *On reasoning and argument* (pp. 401-408). Springer.
- Hoy, W., & Miskel, C. (2001). *Educational administration: Theory, research and practice*. Boston: McGraw-Hill.

- Hubbard, L., Datnow, A., & Pruyne, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation, 42*, 54-62.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the “data driven” mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.), *Evidence and decision making*. Chicago: National Society for the Study of Education.
- Jimerson, J. B., Cho, V., & Wayman, J. C. (2016). Student-involved data use: Teacher practices and considerations for professional learning. *Teaching and Teacher Education, 60*, 413-424.
- Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology, 23*(2), 65-86.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*(9), 697-720.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgement. In J. H. Keith & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 267-293). Cambridge: Cambridge University Press.
- Kahneman, D., & Klein, B. M. (2009). Conditions for intuitive expertise. A failure to disagree. *American Psychologist, 64*(6), 515-526.
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction, 28*, 73-84.
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review, 42*(4), 448-457.
- Kaufmann, E., Reips, U.-D., & Merki, K. M. (2016). Avoiding methodological biases in meta-analysis. *Zeitschrift für Psychologie, 224*, 157-167.

- Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50(3), 456-460.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16, 263-268.
- Kowalski, T., & Lasley, T. (Eds.). (2009). *Handbook of data-based decision making in education*. New York: Routledge.
- Knapp, M. S., Copland, M. A., & Swinnerton, J. A. (2007). Understanding the promise and dynamics of data-informed leadership. *Yearbook of the National Society for the Study of Education*, 106(1), 74-104.
- 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.
- Leonard, N. H., Scholl, R. W., & Kowalski, K. B. (1999). Information processing style and decision making. *Journal of Organizational Behavior*, 20, 407-420.
- LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal*, 40(1), 43-89.
- Little, J. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, 118(2), 143-166.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty replication in the education sciences. *Educational Researcher*, 43(6), 304-316.
- 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.
- Mandinach, E.B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for

- datadriven decision making. *Data-driven school improvement: Linking data and learning*, 13-31.
- 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.
- Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, 12(2), 105-128.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Oláh, L. N., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education*, 85(2), 226-245.
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity*, 41(2), 233-249.
- Park, V., & Datnow, A. (2017). Ability grouping and differentiated instruction in an era of data-driven decision making. *American Journal of Education*, 123(2), 000-000.
- Park, V., St. John, E., Datnow, A., & Choi, B. (2017). The balancing act: Student classroom placement routines and the uses of data in elementary schools. *Journal of Educational Administration*, 55(4), 390-406.
- Pelletier, L. G., & Sharp, E. C. (2009). Administrative pressures and teachers' interpersonal behavior in the classroom. *Theory and Research in Education*, 7(2), 174-183.
- Penninckx, M., Vanhoof, J., & Van Petegem, P. (2011). *Evaluatie in het Vlaamse onderwijs. Beleid en praktijk van leerling tot overheid* [Evaluation in Flemish education: Policy and practice from pupil to government]. Antwerpen: Garant.

- 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.
- Senge, Peter. (2001). *Lerende scholen: een vijfde discipline-handboek voor onderwijzers, ouders en iedereen die betrokken is bij scholing* [Learning schools: a fifth-discipline textbook voor educators, parents and other stakeholders]. Schoonhoven: Academic Press.
- Schildkamp, K., & Lai, M. K. (2013). Introduction. In K. Schildkamp, K. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 1-8). Dordrecht: Springer.
- Schneider, A., & Ingram, H. (1993). Social construction of target populations: Implications for politics and policy. *The American Political Science Review*, 87(2), 334-347.
- Shepard, L. A. (2005, October). *Formative assessment: Caveat emptor*. Paper presented at the ETS Invitational Conference, The Future of Assessment: Shaping Teaching and Learning, New York, NY
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113-141.
- Spillane, J. P., & Miele, D. B. (2007). Evidence in practice: A framing of the terrain. *Yearbook of the National Society for the Study of Education*, 106(1), 46-73.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon: Routledge
- Strayhorn, T., Kowalski, T., & Lasley, T. (2009). Handbook of data-based decision making in education.
- Terwel, J. (2006). Is de school een sorteermachine? Schoolkeuze en schoolloopbaan van leerlingen van 10-16 jaar. [Is school a sorting machine? Educational trajectories of pupils aged 10-16] University of Amsterdam, oration

- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education, 45*, 73-82.
- authors. (2016).
- Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. (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.
- 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.
- 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.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45*, 128-136.
- Yin, R. (1994). *Case study research: Design and methods*. Beverly Hills, CA: Sage Publishing.

