Sorting pupils into their next educational track : how strongly do teachers rely on data-based or intuitive processes when they make the transition decision?

Sorting pupils into their next educational track: how strongly do teachers rely on data-based or intuitive processes when they make the transition decision?

Problem statement

Teacher judgment is an important issue, given the great impact of high-stakes decisions such as placement and promotion on their pupils' educational trajectories (Bonvin, 2003; Eurydice, 2011; Goos, Van Damme, Onghena, Petry, & de Bilde, 2013). For many years, teachers' intuitive evaluation of pupils' competencies was considered to be a solid basis for teacher judgment; only recently have teachers been expected more and more to use data to inform their decision making (Mandinach & Jimerson, 2016; Schildkamp, Lai, & Earl, 2012). This expectation is based on critiques questioning the accuracy of intuitive teacher judgment. Research has shown that intuitive teacher judgment can be inaccurate when prompted by expectancy effects and different sorts of bias (Kahneman & Klein, 2009; Kaiser, Retelsdorf, Südkamp, & Möller, 2013).

Frameworks have been developed to guide teachers' data use for decision making, involving different steps that constitute a systematic decision cycle. Generally, data-based decision making starts from a question or problem definition, followed by data collection, sense making of data and evaluation of alternatives before a decision is made (Coburn & Turner, 2011; Mandinach, Honey, & Light, 2006; Schildkamp et al., 2012). In the final step of the decision process, all information is assumed to be weighed against pre-defined criteria, and teachers are expected to evaluate alternatives until they arrive at the decision that best meets a clearly defined purpose. These maximizing procedures describe how teachers can reach the optimal choice, based on analyses of all relevant data that were collected (e.g., Simon, 1987).

A frequently heard criticism, from both scholars and practitioners, is that these maximizing procedures do not coincide with decision making in complex contexts, as is the case in

education. In practice, teachers might instead make a satisficing decision, as they evaluate options until they find one that is 'good enough' based on a limited set of data. Many (often unknown) factors are always influencing pupils' performance and development, so it is hard for teachers to know when they have made the best decision. Teachers cannot gather all possible relevant data, as, first, they are not likely to be aware of all data that are relevant for a specific case, and second, they are not likely to be able to access and process all of those data. In (classroom) practice, principles of bounded rationality may also apply. That is, the rationality of teacher judgment might be bounded by limited time or limited cognitive capabilities to process all available data (Kahneman, 2003; Kahneman & Frederick, 2005; Klein, 2008; March, 1978; Simon, 1987). Therefore, we expect teachers to use coping strategies based on expertise, as suggested in theories of naturalistic decision making (e.g., Klein, 2008; Simon, 1987). Our hypothesis is that teachers will rely on intuitive processes to define a satisficing decision, rather than collecting and weighing all available data until the optimal decision is reached (Kahneman, 2003; March, 1994; Simon, 1987).

Theories of naturalistic decision making study human judgment in changing circumstances with uncertainty about the future. These theories often start from the idea that human judgment is guided by intuitive evaluations of the situation, based on expertise within a field (e.g., Klein, 2008; Simon, 1987). For example, the recognition-primed decision model describes how experts, such as teachers, develop patterns and mental models that allow them to recognize relevant indicators automatically without a deliberate and systematic search for data (Klein, 2008).

Although a growing body of scholars have agreed that a combination of both data-driven and intuitive processes is needed for wise and professional decision making in a contextualised fashion (e.g., Earl & Louis, 2013; Hammond, Hamm, Grassia, & Pearson, 1987), educational research that addresses the combination of data-driven and intuitive aspects of judgment is almost non-existent. Moreover, there is little insight into the processes that underlie teacher judgment (Little, 2012). The validity of decision-making is mostly discussed as a function of

its outcome: the decision. Nevertheless, teacher judgment greatly influences their pupils' educational trajectories (Allal, 2013; Earl & Louis, 2013) and has an impact on educational equity (Datnow & Park, 2015). In many educational systems, teachers still have great autonomy with regard to decisions having high stakes for pupils' educational trajectories, such as placement in educational tracks and retention or promotion (Bonvin, Bless, & Schuepbach, 2008; Brookhart, 2013). Since little is known about the way teachers make decisions, we must overcome this shortcoming. In this research, we will describe and explain the processes involved in teacher judgment throughout the different steps of decision making, taking into account both data-driven and intuitive processes and how they mutually influence the final decision. To our knowledge, no research so far has disentangled the process of teacher decision making based on this dual-process perspective.

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 thorough, fair decision process increases (Epstein, 2008). Therefore, this study focuses on a specific case of high-stakes decision making, namely the transition decision. The transition from primary to secondary education can involve a decision with high stakes for the pupils involved, since it can be a 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 (unlike other educational systems) 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 in Flanders can choose to use existing standardized tests to inform teachers' decision making, but these results are not binding for the transition decision.

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 pupil's classroom teacher is still of primary 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 subjects, except gym. At the end of primary education, this classroom teacher must make a transition decision for each student in their classroom. First, teachers need to decide whether or not to give a certificate of primary education. Second, teachers make an official transition recommendation, with the following alternatives: future path in general secondary education (GSE) or no future path in general secondary education. In the latter case, pupils are recommended to choose a school that offers a future track in technical secondary education (TSE, technical curriculum), vocational secondary education (VSE, practical curriculum) or artistic secondary education (ASE, artistic curriculum). Because pupils are thus already sorted at a young age into different tracks as they progress through the educational system, the teacher's transition decision is crucial (LeTendre, Hofer, & Shimizu, 2003).

Given the lack of insight into teachers' decision process from a dual process perspective and the high stakes involved with teachers' transition decision, we sought both to gain a deep understanding of the individual steps of the decision process and to understand how these different steps influence how a decision is made. In previous research, we gained in-depth insight into how teachers define a problem, collect data and make sense of these data during an academic year (author, 2017, 2018, 2019). This study adds to that knowledge by investigating how teachers actually make a decision at the end of the year when they bring together all of the data and evaluate the alternatives. Further, this study brings together the separate steps of the decision process and investigates how these steps influence the final decision. For example, it is important to know whether the collection of data during the year actually leads to a data-based decision in the end.

The following research questions are put forward:

- How data-based or intuitive is teachers' decision regarding sorting a pupil into the next educational track?

- How do data-driven and intuitive processes influence the different steps of teachers' decision process during the year?
- How do teachers make a transition decision at the end of the year? What evidence base is conclusive when teachers evaluate alternatives?

Theoretical framework

Starting from a dual-process approach to teacher judgment, we will elaborate on theories of both data use and intuitive evaluation. Theories of data use (e.g., Datnow & Hubbard, 2016; Mandinach et al., 2006; Schildkamp et al., 2012) and the recognition-primed decision model (Klein, 1997, 2008) will be used as a guiding frameworks. Since we aim at understanding how both aspects of teacher judgment may influence the different steps of the transition decision process, these theories will be discussed and integrated according to the following phases in the decision process: (a) problem definition; (b) data collection; (c) sense making; and (d) evaluation of alternatives.

However, first we need to carefully introduce some nuance into this dichotomous approach to data use and intuitive evaluation. Although we will separate these processes for empirical reasons of conceptual clarity, we need to acknowledge that intuitive evaluation is not the opposite of data use. In practice, both processes are expected to be intertwined and mutually influence each other in constant feedback loops (Hammond et al., 1987; Kahneman & Frederick, 2005).

First step of the decision process: Problem definition

A problem or question is defined when the actual state of affairs is weighed against personal or shared standards with regard to the transition decision that needs to be made at the end of the year (Mintzberg & Westley, 2001; Schildkamp, Poortman, & Handelzalts, 2016). For example, a teacher might define it as a possible problem with regard to the transition when a pupil writes with a lot of mistakes on his or her homework, if the teacher expects pupils to

write without mistakes. This might especially be perceived as a problem when the teacher sees flawless writing as a precondition for future success in secondary education.

Starting from a naturalistic approach to decision making, a decision process may be initiated when a teacher recognizes an indicator without deliberate attention. This intuitive evaluation of a problem is considered to be a valuable aspect of expertise, since it allows teachers to recognize problems at an early stage, even when little data are available (Klein, 2008). However, decision theory stresses the need for further problem diagnosis, using data to test or to elaborate on teachers' problem definition (Cowan, 1986; Mintzberg & Westley, 2001; Schildkamp et al., 2016). Insufficient attention is often paid to the stage of problem definition in the decision process (Hegarty, 1991; Lyles & Mitroff, 1980; Mintzberg & Westley, 2001). Especially when high stakes are involved, decision makers are expected to obtain an accurate understanding of the problem situation, since it influences all of the next steps of the decision process. Intuitive recognition without diagnosis might lead to self-fulfilling prophecies (Blackwell, Miniard, & Engel, 2006; Mintzberg, Raisinghani, & Theoret, 1976). In this research, problem diagnosis refers to the use of at least one output or process indicator collected deliberately and systematically.

Data collection

Before we can study teachers' data collection, first we need to come to a clear understanding of what can be understood as 'data' in the context of teacher judgment.

Theories of data use prescribe a fixed and systematic procedure of data collection following an iterative circle of inquiry (Mandinach, Honey, Light, & Brunner, 2008; Wohlstetter, Datnow, & Park, 2008). Starting from a pre-set goal and guided by a plan, for example, classroom observations may be conducted systematically, using an observation protocol that denotes a form of systematic data collection, and deliberately, as the teacher intends to find out why a certain type of mistake is recurring for a pupil (pre-set goal) (AUTHOR, 2017, 2018).

The recognition-primed decision model, on the other hand, describes how experts are able to recognize indicators without a deliberate focus or without a systematic approach. Throughout their careers, teachers develop a framework of personal knowledge and beliefs based on their learning and experience (Kelchtermans, 2009; Klein, 2008). These personal knowledge frameworks guide teachers' attention (Dane & Pratt, 2007; Klein, 2008). Building on the same example mentioned above, classroom observations are considered to be collected intuitively when they are gathered without a pre-defined, explicit method such as a protocol and without a pre-set, deliberate goal or question.

In our research, the concept 'data' refers to quantitative or qualitative indicators that are collected deliberately and systematically. Teachers' attention might also be drawn by an indicator spontaneously, without deliberate attention or a systematic approach. Evaluations of these indicators are gathered through intuitive recognition of what the teacher perceives to be a relevant cue.

Sense making

The data that are collected do not provide useful input for the decision in the original form in which they are presented (Cousins & Leithwood, 1993). Only after data are analysed and interpreted are they transformed into information that can be used as a basis for decision making. Transforming data into information occurs within a sense-making process in which teachers try to understand what the data mean, in this case in relation to the transition decision they need to make (Datnow, Park, & Kennedy-Lewis, 2012; Spillane, 2012). This sense-making process is highly influenced by the context in which it is taking place (Bertrand & Marsh, 2015).

It has been suggested that although data use models prescribe optimal procedures for data analysis and interpretation based on pre-defined criteria (Bosker, Branderhorst, & Visscher, 2007; Leonard, Scholl, & Kowalski, 1999), in practice teachers are more likely to take mental shortcuts (use heuristics) to come to quick and easier conclusions, using personal criteria

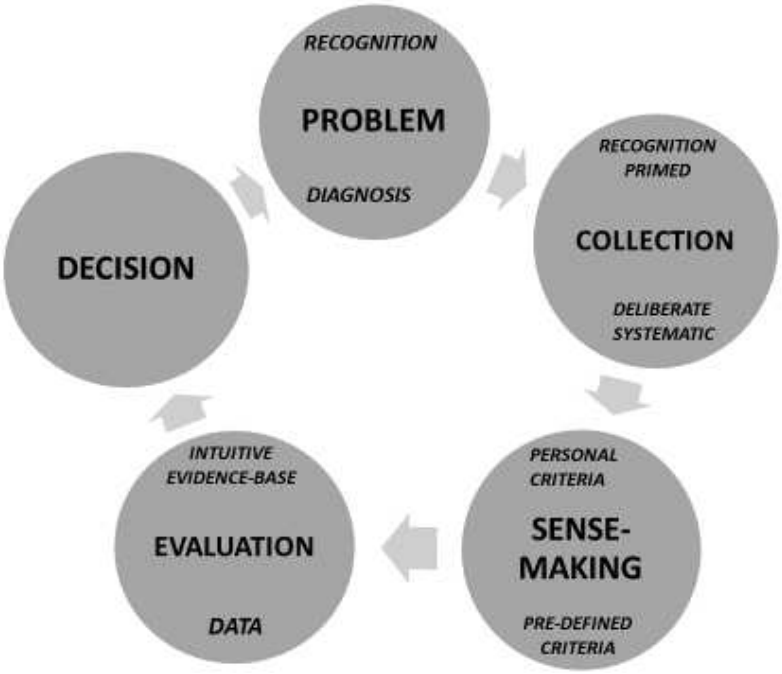
(Evans, 2006; Kahneman, 2003; Klein, 2008). Applying the recognition-primed decision model, patterns and mental models stored in teachers' memory may create expectancies for the future of this pupil (Klein, 2008). Based on experiences with (perceived) similar pupils or situations in the past, mental models will trigger scenarios about future success in secondary education. For example, when a non-native pupil scores below average on a math test, the teacher might recall cases of non-native pupils in the past who failed their final exam, despite all the extra efforts the teacher made during the year. The combination of being a non-native pupil and achieving a low test result may be seen as a pattern, triggering expectancies without systematic analyses of data. In this manner, teachers might jump from data to conclusions without a thorough process of data analysis and interpretation.

Evaluation of alternatives

Finally, teachers are faced with alternative options deriving from the decision process as carried out thus far. Although teachers might have collected a wide array of data during the decision process, this does not necessarily mean that all of the information deriving from these data are taken into account in the final evaluation of alternatives (Blackwell et al., 2006; Kahneman & Klein, 2009; March, 1994). For example, when teachers need to make the transition decision, they think about the different options (Can this pupil receive a certificate of primary education? Do I recommend that this pupil should start in general secondary education?) based on the evidence base they gathered during the year. Teachers need to compare what the information deriving from data tells them with their intuitive evaluation and weigh the importance of both evidence bases. Information deriving from data-driven and intuitive processes may coincide and support the choice of one alternative, or it may provide contrasting information that leads to a different alternative. In this case, an important question concerns what evidence base teachers find conclusive for their final decision.

In summary, teacher judgment is expected to be a complex, iterative process. An overview of the guiding framework is provided in Figure 1.

Figure 1: The process of teacher judgment from a dual-process perspective



Method

Participants

The focus of this study was on 6th grade (pupils aged 11-12) primary teachers in Flanders (Belgium). Twenty-five teachers were randomly selected from a list of teachers in 6th grade with at least 5 years of teaching experience. In research 5 years of experience is often used as a the minimum criteria to identify expert teachers (Palmer, Stough, Burdenski, & Gonzales, 2005). Theories of naturalist decision making suggest that only experts are able to recognise relevant cues spontaneously because they developed mental models based on experience (Klein, 2008). After these 25 teachers had been contacted by researcher 1 in a phone call in which the purpose of the interview-based study was explained, a total of 16 teachers agreed voluntarily to participate. The other teachers who were called, but did not agree to participate in the study, 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 (9) of the teachers had more than 10 years of experience, and the remaining 7 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.

Design

We conducted a longitudinal extended case study to develop an in-depth description of teachers' decision process in a contextualised way (Yin, 1994). 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). This qualitative research design allowed 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. In our research, the case being studied is the teacher decision involved in the transition from primary to secondary education. Using a longitudinal prospective approach, data were collected repeatedly at fixed intervals. In our study, the same teachers were interviewed three times during the academic year (within a month after the start of the school year, six months later, and at the end of the academic year).

The most important distinction between longitudinal and cross-sectional studies, for our purposes, is the timeline. Instead of a researcher collecting data from varying subjects in order to study the same variables, the same subjects are surveyed multiple times with the aim of finding patterns (Yin, 1994). Table 1 provides an overview of the measurement points and aims of the semi-structured interviews that were conducted throughout one school year. All 16 teachers discussed possible transition problems involving 2 specific pupils, which meant that a total of 32 cases were discussed at the start. We asked the teachers to discuss two pupils for whom it was not yet clear that they would get an recommendation for general

secondary education at the end of the year. According to Klein (2008), if you can get decision makers like teachers to tell you about tough cases, then you have a pathway into their perspective. Since two pupils left school during the year, a total of 30 cases (decisions) were available to be examined.

Table 1: Overview of measurement points and aims for the semi-structured interviews with 16 teachers (2 pupils left school)*

Measurement Point	**Month**	**Aim**	**Pupils**	***Total***
MP 1	October	Problem definition	A + B	32
MP 2	March	Search for Data, Sense Making	A + B	32
MP 3	June	Evaluation + Decision	A + B	30*
Total				***94***

Interviews and Procedure

Participants answered open-ended questions that explored their judgments about pupils' competencies and characteristics in relation to their decision process regarding the transition from primary to secondary education. Examples of questions are: "What is your recommendation for this pupil with regard to the transition from primary to secondary education?"; "What are the conclusive arguments for this recommendation?"; "What is the evidence for this argument?" (measurement point 3). The open-ended questions in the interview protocol addressed all of 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 on average one hour and were conducted by a single researcher. The same interview protocol was used in all 16 interviews at a given measurement point to ensure methodological consistency (Cohen, Manion & Morrison, 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 conducted, in which the different methodological choices, data analysis procedures and interpretations were critically examined (Creswell, 2005). With the aim of enhancing the reliability of our research, we clearly described our chain of evidence, so that the external observer can trace the steps in either direction (from conclusions back to research questions or from questions to conclusions).

Coding and analysis

We conducted analyses of the data from all three waves of the longitudinal study. The interviews were transcribed verbatim and analysed with the aim of capturing variation across cases in types of problem definition, data collection and sense-making as well as evaluation of alternatives. A coding scheme was developed, based on the theoretical framework and was discussed in a peer-debriefing session. After both researchers had come to an agreement on the content of the coding scheme the coding scheme was revised and one interview from a specific measurement point was coded and discussed by both researchers. This discussion, for example, stressed the need for a better conceptualisation of what was meant by 'pre-defined' versus 'personal' criteria. Subsequently, the same interview and two other randomly selected interviews were coded by both researchers independently, using the revised coding scheme. The interrater reliability (Cohen's Kappa) was 0.90 for measurement point 1 and 0.72 for measurement points 2 and 3. Disagreements in the codings were resolved by discussing and reflecting on the contents of the different concepts and their boundaries. In the last step of the coding process, researcher 1 went back and re-coded the interviews that had been coded for the interrater reliability check; finally, all interviews were coded by researcher 1, based on the revised coding scheme. Table 2 provides an overview of the final coding scheme. After within-case analyses for each measurement point, a cross-case analysis over the three measurement points was conducted to explore patterns in teachers' decision process (Creswell, 2005).

Table 2: Overview of the codes

Code	Conceptual characteristics	Example
Problem diagnosis	The teacher mentions at least one output or process indicator that was collected deliberately and systematically to define the problem related to the transition decision.	*At the start of the year, a teacher sees a problem with reaching the curricular goals for French because first test results show that a pupil is not able to write French words that are supposed to be known in 6th grade.*
Problem recognition	The teacher mentions no output or process indicator that was collected deliberately and systematically to define the problem. The teacher describes how he/she was able to recognize a certain cue that indicates a problem.	*When a teacher notices a pupil staring outside the window during class time, he/she indicates that lack of motivation might be a problem in relation to the transition.*
Combined problem definition	The teacher mentions both how he/she recognized a cue intuitively and the deliberate and systematic use of an indicator.	*A teacher automatically recognizes a specific kind of mistake on a writing task. Subsequently, the teacher administers a test to check whether the mistake may be related to a learning disability.*
Data collection	Indicators collected deliberately and systematically	*For example: test results for different subject matters, standardized tests, deliberate and systematic observations, planned conversations with parents or colleague, and so forth.*
Intuitive recognition of indicators	Indicators that are not collected deliberately/systematically but spontaneously, recognition-primed	*These examples mostly refer to spontaneous observations during daily practice, or spontaneous conversations with parents or colleagues.*
Pre-defined criteria	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 on to general secondary education.*
Personal criteria	Based on teachers' personal beliefs or feelings	*I don't believe a pupil with 65% on mathematics will make it in general secondary education.*

Evaluation of alternatives	The different options teachers consider based on the evidence	*Based on the test results, the general educational track might be possible, but I feel she is not motivated for the general track.*

Results

Sorting pupils into the next educational track: a result of data use or intuitive evaluation?

In this study, we aim to explain how teachers make the transition decision between primary and secondary education, taking into account both data-driven and intuitive processes in the steps of problem definition, collection, sense-making, and evaluation, and how this interplay leads to their decision.

How do data-driven and intuitive processes influence the different steps of teachers' decision process during the year?

Table 3 provides an overview of the different steps of the decision process each teacher reported for their two pupils with regard to the transition decision, as discussed in the method section. We investigated whether the problem that initiated the process was based on data-driven problem diagnosis (Dia), on intuitive recognition (Rec), or on a combination of both (Com), and how many unique indicators teachers collected deliberately and systematically (Data-driven = Da) or as recognition-primed (Intuitive = In) during the year. In the case of data use, we examined how often data were interpreted by pre-defined criteria (instead of personal criteria). Further, we also took into account to what extent data (Da), an intuitive evidence base (In), or a combination of both (Com) was conclusive when teachers evaluated alternative options. The last columns of table 3 give the outcome of the decision process: the decision to recommend a pupil for general secondary education (GSE) (Yes) or (No). If a pupil instead did not get their certificate of primary education, that is also indicated (*).

Table 3: Overview of the different steps in the decision process for each pupil

Teacher		Problem Definition	Data Collection	Interpreted by Pre-defined criteria	Recognition primed collection	Evaluation of Alternatives	Recommend GSE
Emma	*P1*	Com	7	6	1	Da	*
	P2	Com	4	4	2	Da	*
Amy	*P1*	Com	5	4	2	Da	Yes
	P2	Dia	4	3	2	Da	No
Ann	*P1*	Com	1	1	3	In	No
	P2	Dia	5	3	1	Da	*
Joyce	*P1*	Com	5	4	6	Com	No
	P2	Rec	2	2	4	Com	No
Katy	*P1*	Com	2	2	6	Da	Yes
	P2	Dia	2	2	1	Da	No
Lisa	*P1*	Com	6	4	3	Da	*
	P2	Com	12	8	8	Da	*
Frank	*P1*	Rec	/	/	/	/	/
	P2	Rec	1	1	7	In	No
Bart	*P1*	Rec	2	1	4	In	No
	P2	Rec	1	1	0	In	No
Roy	*P1*	Rec	1	0	6	In	No
	P2	Rec	3	2	7	In	No
Peter	*P1*	Rec	2	0	3	In	No
	P2	Rec	2	1	1	In	No
Mary	*P1*	Rec	2	0	4	In	No
	P2	Rec	3	2	3	In	No
Sophie	*P1*	Com	3	2	6	Com	No
	P2	Rec	2	1	4	Com	No
Bob	*P1*	Rec	3	2	5	Com	*
	P2	Com	2	2	7	Com	No
Julie	*P1*	Rec	/	/	/	/	/
	P2	Com	1	1	9	Com	*
Pam	*P1*	Rec	2	1	6	Com	Yes
	P2	Com	1	1	1	In	No
Liz	*P1*	Rec	1	1	5	In	No
	P2	Com	1	1	5	Com	No

Note: P = pupil, / = pupil left school, * = pupil did not get certificate of primary education. Dia = diagnosis, Rec = recognition, Com = combination, Da = data use, In = intuitive.

In our findings, the decision process was shown to be mostly initiated by intuitive problem recognition, in half of the cases followed by data-driven problem diagnosis. Teachers recognized indicators spontaneously, when they observed their pupils in the classroom. For example, teachers reported an uninterested look on the face or a slow reading pace. This recognition triggered expectancies with regard to the transition decision they would need to make at the end of the year. Diagnosis means that the teachers used at least one output or process indicator that they collected deliberately and systematically. For example, teachers referred to initial test results (output) or deliberate and systematic observations of the method

pupils used for calculating fractions (process). In the interviews, some teachers described how they consulted initial test results to gain better understanding of the problem they recognized. In most cases, data confirmed the problem they identified intuitively, and seldom challenged it. Peter, for example, explained how he recognized a problem that made him question Brahim's transition to the general educational track at the end of the year:

> *Brahim, he lacks common sense, when you ask him a question about the weight of a car, and he answers 'one kilo and a half', without the blink of an eye, then you know… As is the case for many pupils in our school, it is probably related to his language skills (non-native pupil). I feel that Brahim sometimes misses parts of the instruction, because of the language. However, I don't see him making a lot of effort, so if his attitude is not going to change, I can already tell that he is not going to make it to the general educational track at the end of the year.*

In the next step of the decision process, all teachers collected data both in a deliberate and systematic sense and intuitively. Data collection predominantly involved cognitive output indicators (non-standardized test results). In our research, we found little evidence of process indicators or non-cognitive indicators (e.g., the well-being of a child) collected in a deliberate and systematic manner.

The teachers in our interviews collected indicators intuitively during their daily practice without deliberate attention or a systematic search. This almost exclusively refers to non-cognitive (process) indicators. For example, indicators with regard to motivation, effort or well-being spontaneously drew their attention while they were teaching. In the interviews, the teachers explained how they spontaneously collected information while they were teaching, without deliberate or systematic search. For example, when they saw pupils slumping in their seat, staring out the window, when a pupil asked 'what soup are we having for lunch' in the middle of a math class, and the like.

Next, we studied how teachers made sense of data (evidence collected deliberately and systematically). In the interviews, we found that data were interpreted by pre-defined criteria

in almost two-thirds of the cases. For example, teachers used curricular goals to make sense of test results or they mentioned a shared criterion that was agreed upon by the school team. This also means that approximately one-third of the data were interpreted using teachers' personal criteria. It appeared that teachers' personal criteria are largely based on beliefs about what matters most for teaching and learning in general and for success in secondary education more specifically. For example, Mary explained: *'She scores only 70% on mathematics, I don't feel that this is the right profile for general secondary education.'*
Some teachers also interpreted a certain (average or low) test score as a good result for a specific pupil because the teacher felt that the pupil had to work hard for this result. For example, Amy explained: *'Mostly, she scores about 29 out of 50, sometimes 33 or 34 out of a total score of 50. When you know she needs to work extra hard because of her disability, these are great results.'*
In contrast, teachers could also interpret an average test result as too low for the general secondary track when this grade resulted from (perceived) lack of effort or motivation. As Roy, for example, told us:

> *Well, he scored 7/10 on that test when he should have been able to do better than that. There is no motivation or effort. You have to understand that 70% in the end will not be enough for general secondary education when it results from lack of effort and his bad attitude…*

In the next section we will discuss the last two columns that describe the last step of the decision process in greater depth.

What evidence base deriving from data-driven or intuitive processes is conclusive when teachers make the transition decision at the end of the year?

Table 4 provides a refined view of the last two columns of table 3. In this table, we give an overview of the different indicators teachers took into account in the process of evaluating alternative options with regard to the transition decision. Unique counts for each indicator are pictured as (+) when they supported and (-) when they questioned in a negative sense a

successful transition to general secondary education. Table 4 also shows which indicators were conclusive and whether the decision was for a positive recommendation to the general educational track in secondary education (Yes or No). When the pupil did not get their certificate, this is also indicated (*). Teachers may have collected a wealth of indicators during the year (the collection pictured in table 3), but this does not necessarily mean that these were all taken into account when the decision was to be made.

Table 4: Overview of the indicators teachers mentioned in the evaluation of alternatives related to the decision and the evidence base that was conclusive

		Test results	*Standardized tests*	*Conversation parent*	*Conversation pupil*	*Input indicator*	*Observations effort*	*Observations interest*	*Observations well-being*	*Observations skills*	*Parental support*	Conclusive for evaluation	Positive recommendation GSE
Teacher		**Data**					**Indicators collected intuitively**						
Emma	P1	-				-	+				-	Da	*
	P2	-		-	-							Da	*
Ann	P1	-	-			-	+		+		-	In	No
	P2	-	-						+	-		Da	*
Amy	P1	+					+	-		-	-	Da	Yes
	P2	+							-			Da	No
Lisa	P1	-	-		-		+	-				Da	*
	P2	-					-				-	Da	*
Katy	P1	+	+							+		Da	Yes
	P2	-		+			+		-	-		Da	No
Bart	P1	-	+							-		In	No
	P2	+				-						In	No
Peter	P1	+					+	-		-		In	No
	P2	+					-	-		-		In	No
Mary	P1	+				-	+					In	No
	P2	+					-					In	No
Frank	P1	+					-	-			-	In	No
Roy	P1	-						-	-	-		In	No
	P2	-					+	-	-			In	No
Joyce	P1	-		-	-	-		-				Com	No
	P2	-		-	-				-			Com	No
Sophie	P1	-		-	-	-		-	-			Com	No
	P2	-					-	-				Com	No
Bob	P1	-					+	-		-		Com	*
	P2	-	-			-	-	-		-		Com	No
Julie	P2	-				-	-		-			Com	*
Pam	P1	+	+				+					Com	Yes
	P2	+					-			-		In	No
Liz	P1	-			-			-	-			In	No
	P2	-					-			-		Com	No

Note: P = pupil, / = pupil left school during the year* = pupil did not get certificate of primary education.
Da = data use, In = intuitive, Com = combination.

First, table 4 shows that all teachers took into account both data collected deliberately and systematically and indicators they gathered intuitively when they considered the different decision options, but not all of this information was used in the final decision.

In 3 out of the 5 cases, the decision not to grant a certificate of primary education was based on a low level of proficiency with the Dutch language (non-native pupils) was of conclusive importance. Both Emma and Lisa explained that they regretted the decision, because their pupils showed a lot of effort and they made considerable progress during the year. Unfortunately, at this point in time the boys did not reach the curricular goals, so they could not give them a certificate of primary education. When the major curricular goals are not met, teachers do not really evaluate alternative options; one source of data (e.g., a low score for a major curriculum area such as the Dutch language) directly leads to withholding of the certificate.

Another group of teachers exclusively based their decisions on indicators collected intuitively, ignoring the data. Although we must be careful with this conclusion given the limited number of cases, in our study we found that none of the decisions based entirely on indicators collected intuitively led to a positive recommendation to the general track. The negative recommendation was mostly based on indicators with regard to lack of effort, lack of interest in a general curriculum and lack of observed cognitive skills. So, when different decision outcomes were possible (positive or negative recommendation) some teachers only relied on indicators collected intuitively to formulate a negative recommendation. In some cases, they did this even when the data told another story. For two-thirds of these pupils, teachers mentioned average test results, but based on the spontaneous recognition of indicators during the year, teachers found that the general track was not the best decision. For example, when teachers felt that the general secondary track would increase the pressure on a pupil too much, or that failure in a demanding track would lower pupils' self-esteem, an intuitive evaluation of the pupils' well-being was mentioned as conclusive. For example, as Roy explained:

> *The danger of sending him to general secondary education is that he is going to hate school. At this point in time, there is no intrinsic motivation. His results are average, but he doesn't like working for school, there is no parental support and when he fails*

in secondary education because he doesn't work for school, because he lacks the will to study, he will become so demotivated that his school career will be over. Maybe it is better to give him at least a chance by sending him to vocational education where he hopefully will find joy in working on practical tasks. Maybe this can get him motivated, there is more to it than his test results.

Peter and Mary had pupils with overall average test scores but they felt that these pupils did not belong in the general track. According to Mary, for her pupil with an autism disorder, it was more important to lower the pressure; therefore being pushed towards the general track was not a good choice. Peter did not consider his pupil to be a '*student'* because the boy did not like reading books, preferred to play football and did not give smart answers during class.

Conclusion and discussion

Main findings

In many educational, systems teachers still have great autonomy with regard to decisions having high stakes for pupils' educational trajectories, such as placement in educational tracks and retention or promotion (Bonvin et al., 2008; Brookhart, 2013). Yet, little is known about the way teachers make such decisions. A dichotomous view of teacher judgment supposes that teachers still rely too much on intuition, while their decisions should become more data-based.

First, our findings show that we cannot understand teacher decision making by merely looking at the outcome. When we want to understand and enhance the validity of teachers' decisions, we need to see what happens in the different steps of the process. For example, we found that some teachers initiate a decision process solely based on the intuitive recognition of one cue. When these teachers do not search for data deliberately and systematically throughout the year, the danger of confirmation bias may lead to self-fulfilling prophecies (e.g., Kahneman & Frederick, 2005). For example, we found that the competencies of pupils with a lower SES or non-native speakers were often estimated lower

than what was suggested by test results.

Our research also showed that collecting data does not necessarily lead to data-based decisions. We found that some teachers ignore all data when they make the final decision. Instead, they rely on an intuitive evaluation of indicators they collected spontaneously during their daily practice, sometimes despite (standardized) test results that provide contrasting information. In the cases we investigated, this always meant that pupils were sorted into a lower educational track because teachers felt this would be beneficial for their motivation or well-being. In this regard, our findings coincide with previous research showing that judgment mainly based on non-achievement factors generally underestimates what pupils are able to do (Allal, 2013; Timperley & Parr, 2010). Previous research also showed that intuitive judgment disadvantaged certain groups such as low achievers, pupils with special educational needs or pupils from lower social classes (Briscoe, 1991; Brookhart, 2013; Stiggins, 2005).

We found examples where one single source of data, a low score for a major curriculum area (e.g., the Dutch language) led to a negative recommendation, despite considerable progress, effort or high scores for other curricular content (e.g., mathematics). In these cases, teachers did not use any other data to weigh the evidence and consider alternative options. In our research, this practice seemed to disadvantage the most marginalized students, as it was non-native speakers who were judged this way.

We followed thirty pupils who teachers were not certain at the start of the year would get a positive recommendation for the general track in secondary education. At the end of the year, only three of these pupils received positive recommendations. It is possible that once teachers identified a pupil as a transition problem, confirmation bias led to strengthening this belief, rather than challenging it. Deficit thinking may lead teachers to see pupils in terms of expecting something to be lacking and focus their attention on data that support this assumption, and may also influence how data are interpreted. In this way, deficit thinking may implicitly influence teachers' decisions (whether based on data or intuition)

disadvantaging pupils, especially the most marginalized. These findings raise major issues of equity, especially in contexts of high decision-related autonomy for individual teachers and schools. Selecting pupils and allocating them to different tracks within the school system has long-term effects on the social position pupils attain in society (Dekkers, Bosker, & Driessen, 2000). The worrying results of our study coincide with previous findings, showing that elements such as gender, social class or scores of other pupils in the same grade impact decisions on promotion or retention (Brandsma & Doolaard, 2010; Dekkers et al., 2000).

Further, in our research a third of all data were interpreted by teachers' personal criteria instead of pre-set, shared criteria. These personal criteria are largely based on beliefs about what matters most for teaching and learning in general and for success in secondary education more specifically. For example, some teachers feel that effort is a better predictor for future success than grades. These beliefs often had an impact on the evidence that was conclusive in the final decision.

Given that the educational context lacks sufficient stability and reliability to allow accurate intuitive judgment, these findings raise questions with regard to the accuracy of these decisions. Intuitive evaluations of pupils' competencies may not guide pupils to the right educational track.

However, our research has also shown that we need to integrate theories on data-driven and intuitive processes to understand how teachers make high-stakes decisions in practice. Although overconfidence in an intuitive evaluation of pupils' competencies may lead to decision bias, wise educational decisions require teacher expertise, taking into account the specific context of each individual pupil (Earl & Louis, 2013). Our results confirm the hypothesis that teacher judgment is based on both data-driven and intuitive processes. In this research, we made an important first step in exploring and explaining the role of both processes in teacher decision-making. Intuitive recognition is important to define a problem quickly, even when little data are available, and to focus attention on relevant indicators in the multitude of data. Deliberate and systematic data use processes, on the other hand, are

important to question and refine problem recognition, to triangulate and challenge indicators teacher recognized spontaneously in order to prevent confirmation bias or self-fulfilling prophecies. We developed and tested a theoretical framework that takes into account both data-driven and intuitive processes, depicted in figure 1. Although this figure is a static overview of what is in practice a complex, iterative process, it proved to be a valuable lens to study teacher judgment. Further research is needed to refine our understanding of how these steps mutually influence each other or how teachers may go back or forward in the process.

Decisions with regard to retention or promotion and recommendation for educational tracks greatly influence pupils' educational trajectories. Our finding that a certain number of decisions are still solely based on intuitive processes raises a critical question with regard to the fairness and equity of the decisions. At the same time, our results point out that both data-driven and intuitive processes need to be integrated in decision theories in order to fully understand and improve teacher decision-making. Both researchers and policymakers have a shared responsibility to investigate why teachers differ in their approaches to decision making and how the fairness and equity of high-stakes decisions can be monitored and enhanced for all teachers. We will discuss implications for policy and practice in the last section, but first we need to mention some limitations of our research.

Limitations and suggestions for further research

At the starting point of this study, no established theoretical framework was readily available that provides in-depth insight into the processes of actual (rather than idealized) teacher judgment. Starting from a dual-process approach to human decision making, we developed and tested an integrated theoretical framework that takes into account both data-based and intuitive processes. This model proved to be a valuable lens to study teachers' decision making, as it takes into account both data-based and intuitive theories and pictures the processes that lead up to the final decision. However, this is a simplified, static image of what

is in practice a complex process, in which teachers may go back to the previous step, or skip a step.

Although our findings are important for gaining a deeper understanding of the processes that underlie teacher judgment, we do have to acknowledge some limitations of this study. First, the choice of a qualitative study in one specific, low-accountability context (no central exams, no obligation to use standardized tests) implies that we need to be careful with generalisations of our findings. Further research is needed, such as in high-accountability contexts where the final transition decision is informed by standardized tests or central exams. It is important to understand how teachers approach high-stakes decision making in these contexts.

Further, we need to acknowledge that we can critically discuss the processes of teacher judgment, but we cannot evaluate the quality of the decision being made. An important 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 is needed to answer this question.

Third, our research did not involve novice teachers, because we aimed at studying intuitive processes starting from the recognition-primed decision model based on expertise (Klein, 2008). According to decision theory, intuitive processes can only be used as reliable and skilled expertise for judgment when a professional has had enough practice in a similar environment and with similar cases (Kahneman & Klein, 2009). We might expect that novice teachers collect more data deliberately and systematically, since they have fewer patterns and mental models stored in memory to guide their intuitive evaluations. For future research, this is clearly something that needs to be investigated.

Implications for policy and practice

In our study, we started from the idea that both data-driven and intuitive processes are important for wise, professional decision making by teachers. However, at the outset of our research there was no framework available that integrated both processes in teacher judgment. This highlights the lack of deliberate attention from policy and practice to intuitive processes in educational decision making. During the last decade, from both an accountability and a school development perspective, policymakers have focused a lot of effort on enhancing data-based decision making in education. Intuitive processes are often either ignored in these policies or are described as unintended forms of teacher judgment that need to be replaced by data-driven approaches. However, our study shows that teacher judgment can only be fully understood through the interplay of data-driven and intuitive process in the different steps of the decision process. Based on this insight, policy initiatives that aim to enhance the validity and equity of teachers' decisions should start from an integrated view of teacher judgment. In our framework, we described and explained how experienced teachers can wisely use both data collected both deliberately and systematically and indicators they noticed intuitively to gain a broad and contextualized view of pupils' competencies.

However, not all teachers in our research combine multiple sources of data or search for alternative explanations before they make a decision. This may be due to a lack of training and support that starts from an integrated perspective. Teachers need to be supported in developing the right knowledge, skills and dispositions to combine different indicators, as specific competencies are required to combine and weigh information deriving from multiple sources. A broader and more encompassing view of teachers' competencies for decision making is needed in order to understand how data-driven and intuitive evidence bases can be combined in a wise manner that enhances the appropriateness of educational decisions.

Over the past decade, many efforts have been made to enhance data use and teachers' data literacy in education. Little effort has been made regarding understanding and supporting the

contributions of intuitive processes to educational decisions. In order to prevent the overconfidence trap often associated with intuitive judgment, teachers need to learn about theories of judgmental errors and conditions that can help prevent bias. Therefore, decision theories should be included in teacher education and teacher training programs.

Further, supporting a systematic collaborative cycle of inquiry might overcome an individual lack of judgment literacy, since it forces teachers to share, reflect and discuss their beliefs, the inferences they make and the criteria they use when they evaluate alternatives. It is crucial that teachers explicitly discuss their personal beliefs with colleagues and come to a shared understanding of what is important with regard to the transition decision. Collaboration and feedback are not just important to enhance teachers' data use. As Kahneman and Klein (2009) pointed out, enhancing the likely quality of intuitive processes also requires the opportunity to learn through cooperation and feedback.

Further, as we found many examples of deficit thinking, it is alarming that in some educational systems (such as Flanders), teachers still have great autonomy with regard to deciding pupils' educational trajectories. For policy, it is important to investigate what systems, wisely combining (standardized) data and intuitive collection of indicators, strengthen the fairness and equity of educational decisions.

Altogether, as far as theory, more research studying teacher judgment from a dual-process approach is required. More research is needed to understand under what preconditions intuitive processes can contribute to reliable and valid decisions. For practice, an important responsibility lies in enhancing teachers' judgment literacy through teacher education and collaborative in-service training and support. Decision theories and practices deserve more deliberate attention, since teachers' decisions continue to greatly influence pupils' educational trajectories.

References

AUTHOR. (2017).

AUTHOR. (2018).

AUTHOR. (2019).

Allal, L. (2013). Teachers' professional judgment in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20-34.

Blackwell, R., Miniard, P., & Engel, J. (2006). *Consumer behaviour*. Boston: Mason.

Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal*, *52*(5), 861-893.

Bonvin, P. (2003). The role of teacher attitudes and judgment in decision-making: the case of grade retention. *European Educational Research Journal*, *2*(2), 277-294.

Bonvin, P., Bless, G., & Schuepbach, M. (2008). Grade retention: decision-making and effects on learning as well as social and emotional development. *School Effectiveness & School Improvement*, *19*(1), 1-19.

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.

Brandsma, H., & Doolaard, S. (1999). Differences in effectiveness between primary schools and their impact on secondary school recommendations. *School Effectiveness and School Improvement*, *10*(4), 430-450.

Briscoe, C. (1991). The dynamic interactions among beliefs, role metaphors, and teaching practices: A case study of teacher change. *Science Education*, *75*(2), 185-199.

Brookhart, S. M. (2013). The use of teacher judgment for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 69-90.

Cohen, L., Manion, L., Morrison, K., & Wyse, D. (2010). *A guide to teaching practice*. Routledge.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173-206.

Cousins, J., & Leithwood, K. (1993). Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, Utilization*, *14*(3), 305-333.

Cowan, D. A. (1986). Developing a process model of problem recognition. *Academy of Management Review*, *11*(4), 763-776.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of Management Review*, *32*(1), 33-54.

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.

Datnow, A., & Park, V. (2015). Data use--for equity. *Educational Leadership*, *72*(5), 48-54.

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.

Dekkers, H. P., Bosker, R. J., & Driessen, G. W. (2000). Complex inequalities of educational opportunities. A large-scale longitudinal study on the relation between gender, social class, ethnicity and school success. *Educational Research and Evaluation*, *6*(1), 59-82.

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* (pp. 193-207). Dordrecht: Springer.

Epstein, S. (2008). *Intuition from the perspective of cognitive experiential self-theory*. New York, NY: Erlbaum.

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

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378-395.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology*, *51*(3), 323-347.

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.

Hegarty, M. (1991). Knowledge and processes in mechanical problem solving. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 253-285). Hillsdale, NJ: Erlbaum.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In J. H. Keith & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 267-293). Cambridge UK: Cambridge University Press.

Kahneman, D., & Klein, B. M. (2009). Conditions for intuitive expertise. A failure to disagree. *American Psychologist*, *64*, 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.

Kelchtermans, G. (2009). Who I am in how I teach is the message: self-understanding, vulnerability and reflection. *Teachers and Teaching: Theory and Practice, 15*(2), 257-272.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), Expertise: Research and applications. *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Erlbaum.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456-460.

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. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*(2), 143-166.

Lyles, M. A., & Mitroff, I. I. (1980). Organizational problem formulation: An empirical study. *Administrative Science Quarterly*, *25*, 102-119.

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

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

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.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics*, *9*(2), 587-608.

March, J. G. (1994). *A primer on decision-making*. New York, NY: The Free Press.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of "unstructured" decision processes. *Administrative Science Quarterly*, *21*, 246-275.

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

Palmer, D. J., Stough, L. M., Burdenski, Jr, T. K., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. Educational Psychologist, 40(1), 13-25.

Penninckx, M., Vanhoof, J., & Van Petegem, P. (2011). *Evaluatie in het Vlaamse onderwijs. Beleid en praktijk van leerling tot overheid* [Evaluation in the Flemish school system: Policy and practice from pupil to government]. Antwerpen: Garant.

Schildkamp, K., Lai, M. K., & Earl, L. (2012). *Data-driven decision making around the world: Challenges and opportunities.* Dordrecht: Springer.

Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School Effectiveness and School Improvement*, *27*(2), 228-254.

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *The Academy of Management Executive (1987-1989)*, *1*(1), 57-64.

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

Stiggins, R. J. (2005). *Student-involved assessment for learning*. Upper Saddle River, NJ: Prentice Hall.

Terwel, J. (2006). *Is de school een sorteermachine? Schoolkeuze en schoollloopbaan van leerlingen van 10-16 jaar* [Is school a sorting machine. School choice and educational tracks of pupils from 10 - 16]. Amsterdam: Vrije Universiteit Amsterdam.

Timperley, H., & Parr, J. (2010). *Evidence, inquiry and standards*. Wellington: NZCER.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, *19*(3), 239-259.

Yin, R. (1994). *Case study research: Design and methods*. Beverly Hills, CA: Sage Publishing.