Formative assessment in schools: From school board level to the classroom level

Chair: Kim Schildkamp, University of Twente
Discussant: Ellen Mandinach, WestEd

Introductory abstract
This session focuses on the implementation of different approaches to formative assessment, at the school board level as well as school level. Effective use of formative assessment promotes student learning (Bennett, 2011; Van der Kleij et al., 2015), because teachers and students can use the information derived from assessments as a form of feedback to support learning (Bennett, 2011; Black & Wiliam, 1998). Teachers can adapt their instruction to the needs of the diverse learners based on this feedback, and students can use this feedback to direct their own learning process. The use of feedback can have a strong impact on student learning (e.g. Black & William, 1998; Hattie & Timperley, 2007). A distinction can be made between different approaches to formative assessment, such as data-based decision making and Assessment for Learning. This symposium focuses on these different approaches, both from a perspective of a school board (paper 1), as well as from the perspective of the implementation in schools in Belgium (paper 2), The Netherlands (paper 3), and the USA (paper 4). This proposal fits well with conference strand IV “Measuring and Evaluating School Change”, as it focuses on using measurement and evaluation results to improve education.

Paper 1: Formative assessment from the perspective of a school board
Tom Morskieft & Fridse Mobach, Stichting Carmelcollege

The Carmel Foundation is a Dutch private organization in the field of secondary education. Originally catholic-based, the founding fathers in 1922 were monks of Karmel. Nowadays the Carmel Foundation upholds a more widespread perspective: it governs catholic schools as well as other denomination (about 37.000 pupils, 12-18). Six out of twelve schools are located in the east, the others are spread over the country.

Carmel schools have a great amount of autonomy in their formal relation with the board. Within the boundaries of the Dutch law and Carmel-policy, they can make choices as they feel fit. The main concern of the school board is providing the schools with adequate teaching staff, securing finance and building capacity, and ensuring quality standards, both with regard to the standards as set by the Dutch Inspectorate as well as standards set by the Carmel foundation.

In 2013 Carmel completed a discussion about policy by formulating a strategy ‘Course 2020’. One of the outcomes was the decision to continue our co-operation with the University of Twente. Since 2011 Carmel has been working with the University to initiate and support formative assessment in schools, specifically focused on data-based decision making in schools, through the data team intervention. In four years twenty teams worked with the data team intervention to improve the quality of education. Evidence shows that this has led to improved quality and student learning in some of our schools, but not all.
The follow-up, in 2016 was a program ‘education and research’: several interventions, coordinated by a team of managers/teachers/researchers from Carmel (of which two are pursuing their PhD) in cooperation with the University of Twente. Together with the University, Carmel wanted to study the conditions for sustainable data use. Moreover, the data team intervention focussed on the use of systematically collected data, but teachers also collect enormous amounts of data in informal ways, through for example, unstructured classroom observations, and classroom discussions. Therefore, Carmel also wanted to support schools in the use of another type of formative assessment: *Assessment for Learning*. The University and Carmel are closely working together on research and practice at both school board level, and at school and classroom level. At the level of the school board, Carmel provides the professional development opportunities for schools. Schools can decide whether or not to participate. That schools overall appreciate this way of working is reflected, for example, in the fact that all of the schools participated in the data team intervention, albeit sometimes with a little bit of pressure from the board.

We believe that our supportive efforts will have positive effects on the quality of education, but more evidence still needs to collected, especially since the focus on Assessment for Learning is a rather new initiative. Still our approach is not to mandate as such, but to stimulate, activate and facilitate. But does our governmental efforts meet standards of responsiveness and responsibility? In the end, results will matter (as well)...

Kristin Vanlommel (University of Antwerp) & Kim Schildkamp (University of Twente)

**Issue**
Teacher judgement has proven to be a major determinant of pupils’ progress within educational tracks (Brookhart, 2013). Questioning how the quality of teacher judgment can be enhanced is therefore an important matter. Data use, one approach to formative assessment, can enhance the quality of educational decisions, since it helps prevent and correct the possible bias associated with intuitive judgement (Park & Datnow, 2017). However, using data will not automatically lead to better judgment. Teachers still need to make sense of data before it becomes meaningful information that guides decision-making (Bertrand & Marsh, 2015). Therefore, in this study we will critically investigate how teachers make sense of data when they judge pupils’ competences by studying their interpretative arguments.

**Conceptual approach**
During the sensemaking process, teachers engage in a process of analyses and interpretation in which they build meaning for data (Bertrand & Marsh, 2015; Klein, 2008). When we want to critically examine this process, it is important to understand how teachers’ translate data into conclusions. Kane (2013) offers a framework in which an elaboration of interpretative arguments describes the step from observed data to conclusions. According to Kane (2013) it is important that teachers’ arguments and criteria are explicated because it makes it possible to
investigate the validity of the inferences. An interpretation is considered to be valid if the conclusion is supported by the data (Cronbach & Wainer, 1988; Kane, 2013). However, according to decision theory, teachers’ sensemaking is not necessarily a rational process in which an extensive elaboration of alternative explanations based on objective criteria will lead to conclusions. Instead, teachers often use simpler heuristics that require less cognitive effort (Kahneman & Frederick, 2005). These judgmental heuristics may lead to false interpretations when teachers try to fit data in a frame that confirms their assumptions without searching for alternative explanations, or when their interpretation is greatly influenced by feelings and emotions (Hitchcock, 2017; Kahneman & Frederick, 2005). Questioning the validity of interpretative arguments, we need to investigate to what extent teachers search for alternative explanations and the objectivity of the criteria used.

Methods and main findings
Based on 16 semi-structured interviews with primary teachers in the context of transition decisions in Belgium, first findings show that teachers seldom consider alternative explanations when they make sense of data. Based on recognition-primed cues, teachers have expectations of pupils’ competences and in data they mainly find proof for their assumptions. Secondly, feelings of knowing appeared to be an important aspect of the sensemaking process. Teachers criteria for inferences are mostly based on personal feelings and beliefs and to a lesser extent on pre-set, objective criteria.

Conclusions
Since teachers mostly interpret new data in a way that confirms beliefs, innovative practices are not likely to occur. Research as well as policy needs to take into account the impact of personal feelings of knowing, and insights in confirmation bias, to understand why change is (not) happening.

Christel Wolterinck (University of Twente), Kim Schildkamp (University of Twente), Jeroen van Merriënboer (Maastricht University) & Adrie Visscher (University of Twente)

Issue/question
Assessment for Learning (AfL) is an approach to formative assessment which takes place as part of ongoing classroom practices, and it focuses on the quality of the learning process (Klenowski, 2009). Teachers, students and peers, by collecting formal and informal data, gain insights into where the learner is going, where the learner is now, and how the learner can be moved forward in his/her learning (Black & William, 2009). Although many studies on AfL are available, few studies provide insight into the knowledge and skills teachers need in order to apply AfL in the classroom. In this study a Cognitive Task Analysis (CTA) was performed to explore how teachers in secondary education use AfL in their lessons (Clark et al, 2008).

Conceptual approach
In this study a Cognitive Task Analysis (CTA) has been used to study which constituent skills, the complex skill of Assessment for Learning (AfL), are needed. CTA can provide information about the knowledge, thought process, and goal structures that underlie a certain task performance, in this study AfL. Skills and knowledge that are needed to perform a complex task can be identified, analyzed and structured. The purpose of this study is to provide a coherent description of everything a teacher needs to apply AfL in the lesson together with the students. This includes an overview of all constituent skills and the coherence between these skills, insight into what knowledge a teacher needs, and an overview of the complexity factors that can be used to organize AfL tasks according to their complexity. This information can be used in a later study to develop or improve AfL skills of teachers.

**Method**

For this CTA a variety of techniques have been used such as, document analysis, classroom observations, interviews and expert meetings with ten expert teachers and eight subject-matter experts. Based on these results, an AfL skill hierarchy was developed.

**Main findings**

Findings showed that AfL encompasses a wide range of skills, such as providing feedback to students in different manners, and that AfL during the lesson cannot be separated from the preparation of the lesson, giving the lesson, and the evaluation of the lesson afterwards. Furthermore, aspects related to the complexity of AfL are, for example; heterogeneity of the classroom; whether the activities be planned in advance or that a teacher needs to act on the spot; and more responsibility and ownership for the students.

**Conclusion**

This study shows how a CTA can be applied to study teacher’s behavior. The results provide insights into the knowledge and skills that are needed for the implementation of AfL in the classroom, which is a complex endeavor. The results can be used as input for the development of a professional development trajectory aimed at improving AfL knowledge and skills.

**Paper 4: Some Findings about Observed Formative Assessment Practices in Mathematics and Science Classroom in the United States**

Maria Araceli Ruiz-Primo, Stanford University

**Issue and conceptual approach**

We tend to think that teachers are consistent in their instructional practices – their pedagogical practices are similar across the days, weeks, and months of the school year. We can easily say that a similar belief applies to teachers’ formative assessment practices. Papers on formative assessment rarely mention the variability that teachers exhibit in their formative assessments practices. However, this assumption of consistency rarely holds true. For most teachers, practices vary not only from one day to the next, but from one instructional task to the next. Furthermore, as Stodolosky (1988), we have found that instructional tasks vary according to the subject matter. The study reported in this paper will focus on how formative assessment
practices vary across different instructional tasks (e.g., covering content, problem set, reviewing students work). More specifically, the paper will focus on informal formative assessment practices, also called Assessment for Learning.

Method and main findings
We collected information about the formative assessment practices of 34 mathematics teachers and 24 science teachers. Each teacher was observed three times. Pairs of observers used two different types of observation protocols to capture the teacher’s formative assessment practices in real time. The first observation was carried out with an observation protocol focused on capturing information about how teachers clarify learning goals, expectations, and success criteria with students, and how students participate in this formative assessment activity. The second and third observations were carried out using a different observation protocol which focused on how teachers gather, analyze, and act on information about student thinking in both informal and formal ways. The second and third observations will be the focus of the paper.

The information collected by one of the two observers in the classrooms was analyzed to identify the different strategies teachers used to gather information from students, the type of responses students provided, and the type of actions teachers took to respond to the students’ needs. Using the information collected through the observation protocol, 75 profiles in the mathematics classrooms and 50 in the science classrooms that were identified. Each profile was linked to different levels of quality of practices ranging from 0 (lowest) to 5 (highest). For example, a high percentage of the profiles involved teachers asking fact-based questions which led to obtain responses that could hardly help teachers to know how much students were since the students’ thinking was not explicit.

Conclusion
The paper will report on the variation of the formative assessment practices by instructional task and instructional mode (i.e., whole class discussion and autonomous student work). For example, formative assessment practices during whole class discussions are more likely to occur during covering content and reviewing students’ work tasks.

The paper advances our understanding in at least three areas: (1) the impact that type of instructional task has on the of type of formative assessment practices observed, (2) the impact of instructional tasks on sampling observations, and (3) a methodology that can potentially impact the way in which formative assessment should be studied.

References


Hitchcock, D. (2017). Do the fallacies have a place in the teaching of reasoning skills or critical thinking? *On Reasoning and Argument* (pp. 401-408): Springer.


