

Formative Assessment: Data Use for School Improvement

Mariya Ivaylova Adamska

University of Twente

Abstract

The purpose of this study was to determine the extent to which two formative assessment approaches, data-based decision making and assessment for learning, are used in classrooms by teachers as well as the extent to which the different strategies of AfL. In addition, we studied to what extent user characteristics, namely knowledge and skills and attitude, influence those two approaches. Data was gathered by using an online survey. The context of the research was secondary education. In total, 434 respondents filled out the survey who were teachers from all subjects from 17 schools in the Netherlands. The results of this study show that data-based decision making and assessment for learning were not used often by the teachers in their lessons. A possible explanation for that might be that the teachers do not know how to use assessment data for school improvement. Moreover, attitude did influence significantly one of the approaches of AfL. However, knowledge and skills did have a significant influence. Also, the results of this study indicate that professional development in the use of those approaches is needed, in order for them to be used in their full potential.

Keywords: formative assessment, secondary education, data-based decision making, assessment for learning, user characteristics

Acknowledgement

I would like to express my gratitude to my first and second supervisors for being very supportive, positive and understanding. It was an amazing experience to work on this project together.

1. Introduction

Pressure from society, as well as teachers' devotion to improve their classrooms practices, drive schools into providing good quality education to their students. Usually, a prove of effectiveness of a particular school program is being asked for, as well as higher student achievement. Even though education is commonly characterized as a sector where the decision making process is based on intuition and instinct (Schildkamp & Kuiper, 2010), the desired situation is that decision making is not only based on intuition but also on (assessment) data, as this is a proven way to increase student achievement (Van der Kleij, Vermeulen, Schildkamp & Eggen, 2015).

According to Van der Kleij et al. (2015), assessment is formative when assessment results are used to give direction to the learning process of students. Formative assessment involves providing useful feedback from the teacher to the student on tests and homework, as well as information about specific errors and propositions for improvement (Marsh, 2007) and it provides teachers with information about the students' skills and teachers can adapt their instruction accordingly (Sharkey & Murnane, 2006). Formative assessment consists of several approaches and the combination of these approaches can create a more informed learning environment (Van der Kleij et al., 2015; March, 2007). Two of those approaches are data-based decision making (DBDM or data use, for short) and Assessment for Learning (AfL).

DBDM focuses on what has to be learned. Moreover, it relates to data use to inform decisions in schools and its main emphasis is on systematically collected data (Van der Kleij et al., 2015). According to Schildkamp and Kuiper (2010), DBDM can be defined as "systematically analysing existing data sources within the school, applying outcomes of analyses to innovate teaching, curricula, and school performance, and, implementing" (p.482).

AfL focuses on the quality of the learning process, classrooms interactions and relationships rather than the outcomes (Van der Kleij et al., 2015). It was defined as: "part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning" (Klenowski, 2009, p. 264). It has five key strategies: sharing learning intentions/success criteria, questioning/classroom discussion, feedback and peer/self-assessment (William, 2007).

Currently, very often formative assessment is not being successfully implemented or effectively used in schools. Most teachers do not use (assessment) data properly or do not use it at all, and they are still relying on their intuition to make decisions (Schildkamp & Kuiper, 2010). A reason for that might be that teachers do not have the skills to use assessment data effectively; the necessary data to make informed decision is not available or teachers think that they do not need data to make decisions (Schildkamp, Lai & Earl, 2013). Furthermore, according to Verhaeghe, Vanhoof, Valcke & Van Petegem (2010), lack of time and support are the main reasons for why data is not being effectively used.

In order for DBDM and AfL to be used effectively in classrooms, it is important that teachers know the benefits of using data and that they are motivated to work with it. Besides, they should have the specific knowledge and skills to understand and interpret data in an adequate way, as well adapt their instruction and feedback based on it.

This research is going to check to what extent (assessment) data are being used in schools in the Netherlands. The first aim of this study is to investigate to what extent formative assessment (AfL and DBDM) is being used in classrooms and the second aim is to check how big the influence of user characteristics, such as knowledge and skills is on the use

of formative assessment. The result of this study can be used to support teachers in becoming better (assessment) data users.

The rationale above leads to a set of research questions that will form the basis of the study:

1. To what extent is formative assessment being used in classrooms?
 - a. To what extent is AfL being used in the classrooms?
 - b. To what extent is DBDM being used in the classrooms?
2. To what extent do user characteristics influence the use of formative assessment in classrooms?
 - a. To what extent do user characteristics influence the use of AfL?
 - b. To what extent do user characteristics influence the use of DBDM?

This study aims to explore current use of formative assessment in secondary education. Differently from other studies on formative assessment, the present study provides more detailed information on which approaches of formative assessment are used and to what extent, and it even proves deeper insight at the use of the different strategies of one of the formative assessment approaches, in particular AfL. From the gathered results more knowledge is going to be gained on the use of formative assessment in the classroom. This knowledge could be used to support teachers in their efforts to improve their teaching practices, therefore the student achievement. This research will contribute to both theory and practice. Besides evaluating formative assessment in secondary schools, this research gives insight in the size of the impact of user characteristics, as well as more insight in the types of formative assessment teachers' use in classrooms. Some research exists about used characteristics, but it is mostly qualitative research and it is not known how big the impact is of these user characteristics. Although the general literature indicates user characteristics as a factor influencing data use (Schildkamp et al., forthcoming). Practitioners can benefit from the results of this study by using the knowledge gained from the results, for instance giving the teachers a perception of the extent to which they use formative assessment in their lessons, to support teachers and students. Subsequently, teachers could support their students to become better learners by enhancing their instructions.

2. Theoretical overview

2.1 Formative assessment

Learning, teaching and assessment are recognized as dependent on each other (Van der Kleij et al., 2015) where assessment plays a crucial role in education. According to Hornby (2003), assessment has two main roles: (a) summative, to provide information about attainment at the end of the course, (b) formative, to provide support for future learning. Summative assessment mainly emphasizes assessing learning outcomes and it is given periodically to determine at a particular point in time what students know and do not know. The goal is to measure the level of success. On the other hand formative assessment aims to gain perceptions into learning processes that can be used to support learning through clear and detailed instruction as well as feedback (Heitink, Van der Kleij, Veldkamp, Schildkamp & Kippers, 2016).

In most of the definitions similar key components could be observed. The goal of formative assessment is always improvement of teaching (Schriven, 1967; Marsh, 2007; Sharkey & Murnane, 2006) and improved student learning and outcomes (Benette, 2011; Black & Wiliam, 2003; Peterson & Irving, 2008; Van der Kleij et al., 2015; Marsh, 2007).

Another important component is the feedback (Marsh, 2007; Heitink et al., 2016), where teachers as well as the students are feedback users, meaning that the students also could be responsible for giving comments on their work. Formative assessment is also about teachers' instructional adjustments adapted to the students' needs (Sharkey & Murnane, 2006; Marsh, 2007). Therefore, a definition which fits the most with the present study was designed and it concludes that formative assessment involves providing feedback to students, active involvement of students in their own learning, adjustments to teaching to take account of the results of assessment, improvement of teaching and learning and the understanding how this to be done.

Several formative assessment approaches exist. Two important approaches are DBDM and AfL. In this paper these two approaches are going to be investigated.

2.2 Assessment for Learning

Initially, AfL was introduced by UK scholars (Van der Kleij et al., 2015) and in research literature, there is a wide range of definitions of AfL. AfL is an approach to formative assessment that happens as part of ongoing classroom practices and focuses on the quality of the learning process (Heitink et al., 2016). Even though in most of the definitions similar key components, such as that is occurring as a part of ongoing classrooms practices and that it focuses on the quality of the learning process, could be observed (Klenowski, 2009; Clark, 2012; Drummond, 2006; Wiliam & Leahy, 2015) the definition that we are going to use is the one from Klenowski (2009), where he defined AfL: "part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning" (p. 264). Assessment should be used as a way to provide useful and helpful feedback to the students' particular needs (Heitink et al., 2016). As noted by Heitink et al. (2016), feedback is always incorporated in the AfL process in order to guide future learning, on a group as well as on individual level and students play a vital role in AfL as they are also engaged in providing feedback on their own or on their peers' work. According to Timmers and Veldkamp (2011), AfL is responsible for the facilitation of the learners knowing what the desired levels of performance and understanding are, also to help teachers and students to compare the level of performance with the desired level. It also helps teachers and learners to create learning activities which are beneficial for closing the existing learning gaps (Timmers & Veldkamp, 2011). Likewise in a broad review of the relevant research literature, Black and William (2003) stated that AfL, when implemented well, results in better student learning. According to Heitink et al. (2016) one of the main goals of AfL in classrooms is to help the students to learn how to learn. Through AfL, teachers find out what students know, what they partly know and what they do not know so that the follow-on activities can advance learning.

AfL research studies have demonstrated that certain techniques associated with AfL can help students to learn more effectively. According to Wiliam (2011), there are five core strategies for successful formative assessment practice in the classroom, and they are as follows: clarifying, sharing, and understanding learning intentions and criteria for success; engineering effective classroom discussions, activities, and learning tasks that elicit evidence of learning; providing feedback that moves learning forward; activating learners as instructional resources for one another and activating learners as owners of their own learning.

2.2.1 Sharing learning intentions/success criteria

One of the ways teachers get students to learn is by making them participate in particular activities selected by the teacher purposefully to get knowledge (William & Leahy, 2015). The *sharing learning intentions/success criteria* strategy is about getting the students to really understand what their classroom experience will be, how their success will be measured and the specific direction they are going to. The term *learning intentions* illustrates what the teacher want the students to learn, where *success criteria* illustrates the criteria used by the teacher to check whether or not the learning activities in which the students were engaged were successful or not. This strategy is important according to different studies (William & Leahy, 2015; Heitink et al., forthcoming; Black & William, 1998; Crisp, 2012).

It is essential that the students are kept engaged and enthused. The ability to keep the students feel like that is considered as a high priority for successful teachers. It is important that teachers share learning intentions not at the beginning of the lesson but also while teaching it. The aim of any learning intentions or success criterion is to help students learn, not the help them complete their activity. Another important concept related to this criteria is the concept of the generalizability meaning that the students can apply what they have learned in different contexts. In order to do that it is crucial that various learning activities are mixed up. An important point is also that the teacher uses different learning intentions for the different students (William & Leahy, 2015).

2.2.2 Eliciting evidence about student learning

Usually, the main role in a classroom lesson is played by the teacher and the students are given a supporting role. The *eliciting evidence about student learning* strategy is about developing effective classroom discussions, questions, activities, and tasks. The term *eliciting learning evidence* illustrates the various ways evidence about what students can and cannot do could be gathered. The importance of this strategy is acknowledged in several different studies (William & Leahy, 2015; Heitink et al., forthcoming; Black & William, 1998; Crisp, 2012; Gottheiner & Siegel, 2012). William and Leahy (2015) stated it as beneficial if the teacher, for example, does not pick a student to give an answer or comment by the students raising a hand, but the opposite- the teacher should pick a student randomly, but giving also time to think, so that everyone has a chance to respond. Initially, this way of picking students in classroom might frustrate the students but in the long term it is gets accepted by them. In addition, it is advised that the teacher should look at situations where students give incomplete or incorrect answer, as beneficial situations, where a discussion could be started. It is also helpful for the teacher if prior the lesson they design questions which they can ask at any point of the lesson. This should be done so that the teacher can make sure that the students understand the concepts explained in the lesson and then the teacher can move on teaching (William & Leahy, 2015).

2.2.3 Feedback

Information on the students' progress is given in the form of a feedback in schools all over the world. The term *feedback* illustrates the information given by teachers to students, which discusses the students' progress while they are learning. The importance of feedback is acknowledged in several different studies (William & Leahy, 2015; Heitink et al., forthcoming; Black & William, 1998; Crisp, 2012; Gamlem & Smith, 2013; Gottheiner &

Siegel, 2012; Hattie & Timperley, 2007). According to Hargreaves (2013), children who are given feedback on their work, did like being given feedback as long as the feedback they receive is detailed enough and they felt that their learning was supported by providing feedback. The *feedback* strategy is about providing feedback which is beneficial for moving forward with learning. Usually, giving feedback results in four things. Individuals change their behaviour, or they change or abandon the goal. Lastly, it can result in rejecting the feedback. When feedback is given in classrooms, it is crucial that the teacher knows the students who he judges. Also, for example, the teacher should not only state if the answer is wrong or correct, but he/she should give an advice for future improvement. It is also stated that is better for the students' learning if the teacher does not simply give the right answer but gives the students the opportunity to think with only providing some directions. Also, as a more beneficial way to give feedback is written feedback. There are evidences that written feedback support students in finding their own mistakes. However, if feedback is given in a way of comments, it is better if they clearly state what the students need to do and how (Wiliam & Leahy, 2015).

2.2.4 Peer-assessment

Learning from others is not a new trend. Harris and Brown (2013) stated that peer-assessment has a positive influence on students' achievement. The term *peer-assessment* illustrates formative assessment students provide to one another; assessing each other's work not to judge it but to improve it. The *peer-assessment* strategy is about helping the students in classrooms to learn from one another. This strategy is important according to several studies (Wiliam & Leahy, 2015; Heitink et al., forthcoming; Black & Wiliam, 1998; Bryant & Carless, 2009; Crisp, 2012; Harris & Brown, 2013). In order to help students become better learners, they should be given the opportunity to play an active role and talk about their learning and engage in a peer-feedback activity. However, the students need to have a guidance about what sort of comments to write. In literature, it is also advised that the students are given the opportunity to assess their peers work as it is proven that once noticing the mistake in someone else's work, they are less likely to do the same mistakes in their own work. Moreover, it is important that enough time is provided for the students in order to help them own their own learning. Therefore, the students can use various assessment techniques to review their work (Wiliam & Leahy, 2015).

2.2.5 Self-assessment

It is very important that the students are engaged in their own learning and excited about it. The term *self-assessment* illustrates the ability of students to reflect on their learning by assessing their own work. This strategy is important according to several studies (Wiliam & Leahy, 2015; Heitink et al., forthcoming; Crisp, 2012; Fletcher & Shaw, 2012; Harris & Brown, 2013). According to Harris and Brown (2013), self-assessment is beneficial for students' outcomes. Self-assessment is about that the students should take ownership of their own learning. For example, very powerful way of making the students be active in their own assessment is by asking them to keep learning portfolios. Also, self-assessment should be part of the classwork. It is anticipated that is beneficial for the students to record their learning journey. There should be also time for parent-teacher conferences where the students make a plan beforehand what to be discussed in those conferences (Wiliam & Leahy, 2015).

2.3 Data-Based Decision Making

Another formative assessment approach which is going to be discussed in this study is DBDM. In countries all over the world, data use is becoming more and more popular. In most of the developed world, data have become a progressively important and almost inevitable part of people's lives (Hargreaves & Braun, 2013). If we look at different definitions of DBDM, few essential components, e.g., that it is a process, qualitative or quantitative information, are included in all of them (Wohstetter et al., 2008; Wayman, Jimerson, & Cho, 2012; Schildkamp et al., 2013) where (assessment) data is mainly defined as information collected and organized to represent some aspects of the school. This information, which can be qualitative (e.g. in textual form) and quantitative (e.g. numerical form), could consist of performance of students on tests, observations of classroom teaching, as well as surveys (Schildkamp et al., 2012; Van der Kleij et al., 2015).

Research showed that DBDM can result in improved quality of the education (Wohlstetter, Park & Datnow, 2009; Campbell & Levin, 2009; Kerr, Marsh, Ikemoto, Darilek & Barney, 2006; Hargreaves & Braun, 2013), by monitoring curriculum goals, as well as grouping students differently in order to enhance learning and setting suitable learning goals (Van der Kleij et al., 2015), where these are just few examples of how data can be used to improve student learning. However, for DBDM to lead to improved student learning it is crucial that data are used for instructional purposes, which will be discussed in the following section.

2.4 Data use for instruction

In order for DBDM to lead to school improvement in terms of increased student achievement, it is crucial that assessment data is used for instructional purposes. Mostly, data from standardised tests from a student-monitor system can be used, but also data collected from various other assessment methods (Young, 2006), curriculum-embedded assessments as well as observations from daily practice (Van der Kleij et al., 2015). According to Hoogland et al. (forthcoming) assessment data can be used for making instructional improvement based on it. Data can help teachers to identify the conceptions and misconceptions of students (Schildkamp et al., 2015). This can lead to the design of good quality instructions, based on the needs of all the students, and these instructions can lead to improved student learning and better student achievement (Schildkamp et al., forthcoming; Van der Kleij et al., 2015; Schildkamp et al., 2006; Hoogland et al., forthcoming; Park & Datnow, 2009; Daly, 2012; Schildkamp et al., 2013; Kai et al., 2013). For improved instruction, data can be used by teachers in many different ways. Schildkamp et al. (2013) proposed that data can be used for "setting learning goals for students; determine which topics and skills students do and do not grasp; determine students' progress; tailor instruction to individual students' needs; set the pace of lessons; give students feedback on their learning process; form smaller groups of students for targeted instruction; identify instructional content to use in class; study why students make certain mistakes; and adapt instruction based on the needs of the gifted and struggling students" (p. 4).

2.5 Comparison of DBDM and AfL

This section addresses the theoretical differences and similarities of the two formative

assessment approaches that are investigated in this study, namely DBDM and AfL. It is crucial to state that, according to Van der Kleij et al. (2015), the goals of the two approaches are very different. The goal of DBDM is to help educators to use collected data in order to change their practices for student improvement and improve their instructions whereas the main goal of AfL is to learn the students how to learn. Also, another main difference are the assessment methods used by each of the approaches, e.g., DBDM uses systematically collected data whereas AfL uses any kind of information (Vander Kleij et al., 2015). Another difference might also be that the data that is used in DBDM is mainly quantitative and in AfL is mainly qualitative (Van der Kleij et al., 2015). However, each of the approaches is about assisting learning, which “results in different expectations of the roles of teachers, students and other actors in the learning, assessment and feedback processes” (Van der Kleij et al., 2015, p.335). Even though, according to Van der Kleij et al. (2015), these expectations are opposing as in DBDM the responsibility for the assessment process mainly falls on the teacher whereas in AfL the teacher shares the responsibility with the students.

2.6 Problems with using data in classrooms

Using assessment data in schools is a proven way to increase student achievement (Schildkamp et al., 2013; Schildkamp et al., 2014; Hoogland, Schildkamp, Van der Kleij, Heitink, Kippers, Veldkamp & Dijkstra, forthcoming; Schildkamp et al., 2012; Van der Kleij et al., 2015). However, even though there is so much information in literature of the importance of the use of formative assessment in classrooms in order for increased student achievement and generally better learning, and despite the fact that assessment has been on policy agendas internationally for decades, implementation has proven to be challenging and only a minority of teachers actually use it (Heitink et al., 2016). Numerous studies have found that the teachers often lack crucial skills in order to use data in the classroom adequately and effectively (Ikemoto & Marsh, 2007; Hangreaves & Braun, 2013; Marsh et al., 2010; Wayman, 2010; Verhaeghe et al., 2010; Chen, Heritage, & Lee, 2005). According to Verhaeghe et al. (2010), lack of skills, time and support are the main reasons for why data is not being effectively used. Consistently, Ikemoto and Marsh (2007) stated that a big influence on whether or not teacher feel motivated in using assessment data is whether or not they feel prepared enough. Besides, according to Marsh (2007), empirical studies reveal that there is very little evidence that formative assessment is used frequently in classrooms. It is important to state that the role of the teacher is crucial in implementing formative assessment in the classroom.

2.7 The role of the teacher in formative assessment

According to literature, the implementation of formative assessment in the classroom is influenced by *user characteristics*. The user characteristics which can influence DBDM and AfL, are the knowledge and skills of the teachers, as well as their dispositions to use data (Heitink et al., 2016). Wohlstetter et al. (2008), Coburn and Turner (2011) and Heitink et al. (2016) stated that the teachers should know how to collect, analyse, interpret and use data in order to make informed decisions. They should also be able to diagnose students' needs and adjust their instructions in terms of that, thus they could provide more useful feedback to their students (Heitink et al., 2016).

The teachers' dispositions, in a sense of teachers' beliefs and attitudes to use data, also have an influence on DBDM and AfL. Several studies refer to the teachers' belief and

attitudes as factors, that influence AfL and they all state their importance for implementing AfL. Heitink et al. (2016) pointed out that teachers' "beliefs, attitudes, perspectives and philosophy about teaching and learning influence the quality of AfL implementation" (p.56). According to Heitink et al. (2016), teachers should feel responsible not just for the coverage of the curriculum but also for the achievements of the students, as well as giving adequate feedback and for the revision of the teaching plans, in case it is needed. Teachers should also believe in data use as a way to improve student achievement in order to be motivated to use it (Schilkkamp et al., forthcoming) and believe in the importance of using data in their everyday practice. Besides, according to Heitink et al. (2016), teacher's confidence and experience using assessment data is identified as beneficial for the implementation of AfL. Moreover, in order to implement deeply AfL in classrooms, the teachers should have a constructivist view of learning and adequate pedagogical strategies (Heitink et al., 2016).

This theoretical framework leads to the following model presented in this paper (see Figure 1).



Figure 1. *Theoretical framework on factors influencing formative assessment*

3. Methodology

3.1 Research design

This study used a quantitative data collection method to investigate the use of formative assessment in classrooms in the Netherlands. It was also investigated which to what extent knowledge and skills and dispositions to use data have an influence on the use of formative assessment in schools.

3.2 Research context

In a broad review of the relevant research literature, Black and Wiliam (2003) stated that there is an evidence that formative assessment does raise results on the National Curriculum tests in the UK. However, in the Netherlands for instance there are not a lot national standardized assessments in secondary education. But, there is a final examination, as well as an Inspectorate. The role of the Inspectorate is to hold the schools responsible for the education that they provide and the main goal of the Inspectorate is to assess and improve the quality of Dutch schools. According to Schildkamp (2007), the Dutch schools "have always been free to choose the religious, ideological and pedagogical principles on which they base their education, as well as how they choose to organise their teaching activities" (p. 6). Thus, traditionally the schools in the Netherlands have a lot of autonomy (Schildkamp, 2007). In addition, Verhaeghe et al., (2010) stated that governmental bodies expect from autonomous schools to be accountable for monitoring their internal quality policy. Therefore, schools with more autonomy are more likely to use data (Schildkamp et al., forthcoming; Ebbeler, Schildkamp, & Downey, 2012; Earl & Louise, 2012).

In the Netherlands, there are several types of assessment data used in schools. According to Scheerens, Ehren, Slegers, & de Leeuw (2012), examples of assessment data used in schools in the Netherlands are examination results, national assessment programs, international assessment programs, school performance reporting, examinations and student monitoring systems.

3.3 Sampling and sampling techniques

For this study a convenience sample was used and the target population of this study were all teachers who work in secondary education in 52 schools, under one school board, in the Netherlands. Eventually, the survey was filled in by N=434 teachers of 17 schools (response rate 33%). In the total sample, 48.8 % (N=212) were female and 51.2% (N=222) were male. Also, 40.1 % (N=174) of the teachers give their lessons in the lower grades of secondary education and 59.9% (N=260) give their lessons in the higher grades of secondary education. And 0.9% (N=4) give their lessons in employment-oriented training for students who lack the ability to obtain a qualification, 39.9% (N=173) of the teachers give their lesson in pre-vocational education, 28.6% (N=124) give their lessons in the senior general secondary education and 30.6% (N=133) give their lessons in the pre-university education (see Table 1). This particular school board was chosen because the designers of the survey have already worked with it so the school board is familiar with the topics. Moreover, this is one of the largest school boards in the Netherlands.

Table 1. *Frequencies of the sample*

	<i>Frequency</i>	<i>Percent</i>
Male	222	51,2
Female	212	48,8
Total	434	100,0

Table 2. *Frequencies of the sample*

	<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cumulative Percent</i>
Valid Employment-oriented training for students who lack the ability to obtain a qualification	4	,9	,9	,9
Pre-vocational education	173	39,9	39,9	40,8
Senior general secondary education	124	28,6	28,6	69,4
Pre-university education	133	30,6	30,6	100,0
Total	434			

Table 3. *Frequencies of the sample*

		<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cumulative Percent</i>
Valid	Lower grades of secondary education	174	40,1	40,1	40,1
	Higher grades of secondary education	260	59,9	59,9	100,0
	Total	434			

4. Instruments

4.1 Survey

The instrument used for this project was a digital 30-item survey, as it was a valuable way to get results from a very large sample. The survey measured the extent of AfL and DBDM, and the role of the teachers. The scales to measure AfL were based on one existing, reliable and valid surveys (Lysaght & O’Leary, 2013) and consists of the following scales:

- 1- Sharing learning intentions/success criteria (e.g. Learning intentions are stated using words that emphasise knowledge, skills, concepts and/or attitudes i.e., what the students are learning NOT what they are doing)
- 2- Eliciting evidence about student learning (e.g. Assessment techniques are used to facilitate class discussion)
- 3- Feedback (e.g. Feedback to students is focused on the original learning intention(s) and success criteria)
- 4- Peer-assessment (e.g. Students use each other as resources for learning)
- 5- Self-assessment (Students are encouraged to record their progress using, for example, learning logs)

The scales to measure DBDM and user characteristics were derived from a reliable and valid survey developed by Schildkamp et al. (forthcoming), and consist of:

- Data use for instruction (e.g. “To what extent do you use data to set the pace of my lessons; To what extend do you use data to study why students make certain mistakes”).
- Teacher knowledge and skills (e.g. “I have the skills to change my teaching based on data”.)
- Teacher dispositions to use data (e.g. “I believe that it is important to use data to establish the individual learning needs of the students”.)

All items, with the exception of the user characteristics, were measured on a 5-point scale ranging from “This practice is embedded (happens approximately in 90% of the lessons)” to “This happens less than 10% of the lessons”. User characteristics were measured on a 4-point scale ranging from strongly disagree (1) to strongly agree (4). The test items are in Dutch, because the respondents who are going to fill in the survey are from schools in the Netherlands. All the scales also included an “I don’t know” option.

4.2 Procedures, reliability and validity

A pilot needed to be conducted, because a new survey was developed, based on existing surveys. The pilot took place in one school in the Netherlands, where the respondents were secondary teachers in this school (N=68). The validity of the survey was checked by conducting two focus groups with four teachers and one expert in each of the groups. Those focus groups also established the time for filling the survey in and whether the items were clear. It was found out that on average, it took a participant about 15 to 20 minutes to fill out the survey. Based on the results from the focus group, minor adjustments were made, mostly in terms of formatting the items more specifically. An example of adjustment that was made after the session with the focus group, was a change in the translation from the original survey from English to Dutch. In the English survey in question 10 “To what extent do you use data to adapt instruction based on the needs of the gifted students”, the word “gifted” was changed to “better” because according to the teachers it is more likely that there is a group of good students, than a groups of gifted students; gifted student might be one in class rather than a group of students. The teachers could contact the researcher if they had questions about the survey (both before and afterwards). Personal details of teachers were not asked. The results was reported on school level. After this pilot, the survey was administered to the 54 schools in our sample and 14 schools responded.

Confirmatory factor analyses and reliability analyses revealed a seven factor structure almost consistent with the theoretical framework (the exact results can be found in Appendix A):

1. Data use for instruction. There were no items deleted from this scale, because all the items scored above 0.5.
2. User characteristics: knowledge and skills. No items were deleted
3. User characteristics: dispositions to use data. No items were deleted
4. Sharing learning intentions/success criteria. Four items were deleted from the factor Sharing learning intensions, namely:
 - “Assessment techniques are used to assess students’ prior learning (e.g., concept mapping...)” (it did not load sufficiently high, 0.39).
 - “Students demonstrate that they are using learning intentions and/or success criteria while they are working (e.g., checking their progress against the learning intentions and success criteria for the lesson displayed on the blackboard or the flipchart)” (it did not load sufficiently high, .30).
 - “Students are involved in identifying success criteria.” (it did not load sufficiently high, .39 and it also loaded on two other scales)
 - “Success criteria are differentiated according to students’ needs (e.g., the teacher might say, “Everyone must complete part 1 and 2....; some pupils might complete part 3)” (because it did not load on this scale)
5. Eliciting evidence about student learning (e.g. Assessment techniques are used to facilitate class discussion). No items were deleted.
6. Feedback (e.g. Feedback to students is focused on the original learning intention(s) and success criteria). Five items needed to be deleted from the factor Feedback, namely:
 - “Assessment techniques are used during lessons to help the teacher determine how well students understand what is being taught (e.g., thumbs up-thumbs down and/or two stars and a which)”. It did not load on the expected scale and it did not load sufficiently higher on any other scale, (.31);
 - ”Feedback to students is focused on the original learning intention(s) and success criteria (e.g., “Today we are learning to use punctuation correctly in

our writing and you used capital letters and full stop in your study, well done John”). (it did not load the expected scale and it did not load sufficiently high on any other scale, .38);

- “When providing feedback, the teacher goes beyond giving students the correct answer and used a variety of prompts to help them progress (e.g., scaffolding the pupils by saying “You might need to use some of the new adjectives we learned last week to describe characters in your story”). (it did not load the expected scale and it did not load sufficiently high on any other scale, .38);
 - ”Students are involved formally in proving information about their learning to their parents/ guardians (e.g., portfolios or learning logs are taken home)” (it did not load the expected scale and it did not load sufficiently high on any other scale, .34) and
 - “In preparing to provide students with feedback on their learning, the teacher consults their records of achievement against key learning intentions from previous lessons (e.g., the teacher reviews a checklist, rating scale, or anecdotal record that she/he has compiled)” (it loaded on the expected scale, however it also loaded on two more scales and also, it did not load sufficiently high, 0.33).
7. Peer-assessment and self-assessment. Peer/self-assessment came in the factor analysis as one factor even though in the theoretical framework they were two separate ones. The reason for that is probably because they are both focused on the responsibility of the student. There was one item deleted from the factor Peer/self-assessment, namely:
- “Time is set aside during parent/guardian teacher meetings for students to be involved in reporting on some aspects of their learning (e.g., pupils’ select one example of their best work for discussion at the meeting)”, because it did not load on the expected scale and it did not load sufficiently higher on any other item (.31). The item “Student are given an opportunity to indicate how challenging they anticipate the learning will be at the beginning of a lesson or activity (e.g., by using traffic lights).” scored below .5 (.49), but due to theoretical reasons and because it was very close to .5 the item was not deleted.

The reliability of the survey instrument was determined through reliability analysis. Reliability analyses results as measured by Cronbach’s coefficient alpha were sufficient: learning intentions (0.76), eliciting evidence (0.83), feedback (0.81), peer/self-assessment (0.82), data use for instruction (0.89), knowledge and skills (0.81) and dispositions to use data (0.83) (see Table 4).

Table 4. *Reliability of the scales in the survey*

<i>Scale</i>	<i>Number of items</i>	<i>Cronbach’s alpha</i>
Learning intentions	4	0.76
Eliciting evidence	6	0.83
Feedback	5	0.81
Peer/self-assessment	6	0.82
Data use for instruction	11	0.89
Knowledge and skills	7	0.81
Dispositions to use data	4	0.83

4.3 Data analysis

To answer the first research question descriptive statistics (mean, mode, median, etc.) were used in order to show how formative assessment was used both with regard to AfL and to DBDM.

In order to answer the second research question and its sub questions, multilevel analysis needed to be performed. In order to do that, zero model was performed, where the subjects were the different schools. That was done so we could check whether the teachers who work in the same school, have the same results. The intra class correlation (ICC) was .11, and this number shows that the school level only has a very small influence on DBDM and AfL. Therefore we can use linear regression in order to answer the third and fourth research questions (see Table 5, 6, 7, 8, 9). Regression analyses were performed five times- first with DBDM as the dependent variable and knowledge and skills and dispositions to use data as independent variables. After, each subscale of AfL (Learning intentions/success criteria; Eliciting evidence; Feedback and Peer/self-assessment) was chosen as a dependent variable and user characteristics (knowledge and skills and dispositions to use data) were chosen as the independent variables.

Table 5. *Bar plot DBDM*

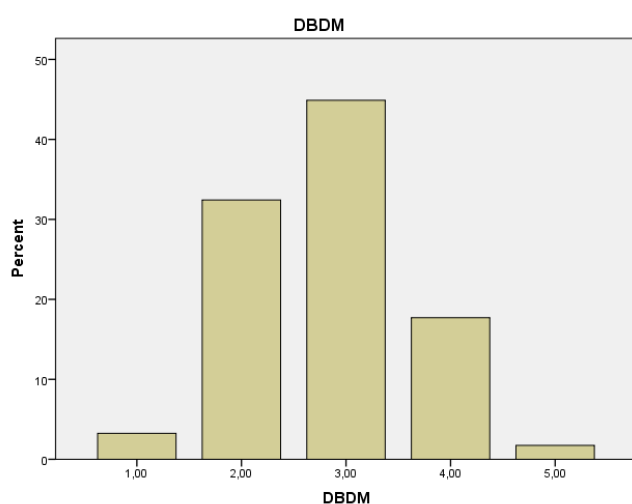


Table 6. *Bar plot learning intentions*

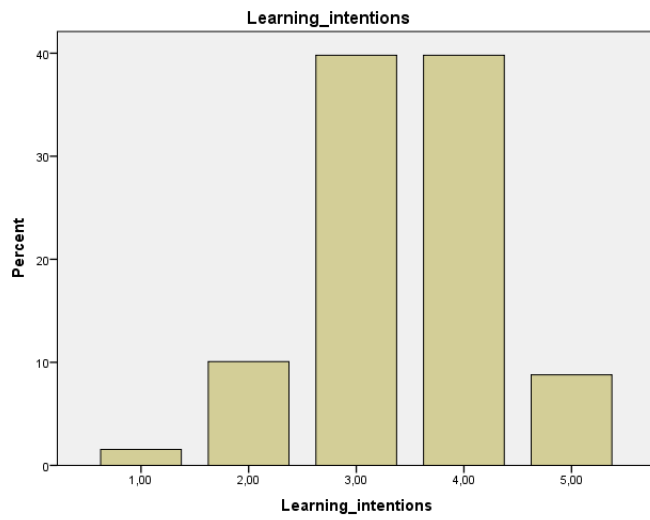


Table 7. Bar plot eliciting evidence

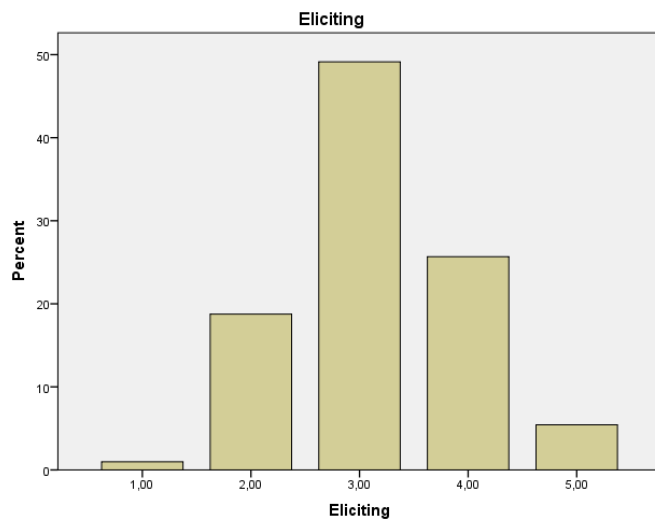


Table 8. Bar plot feedback

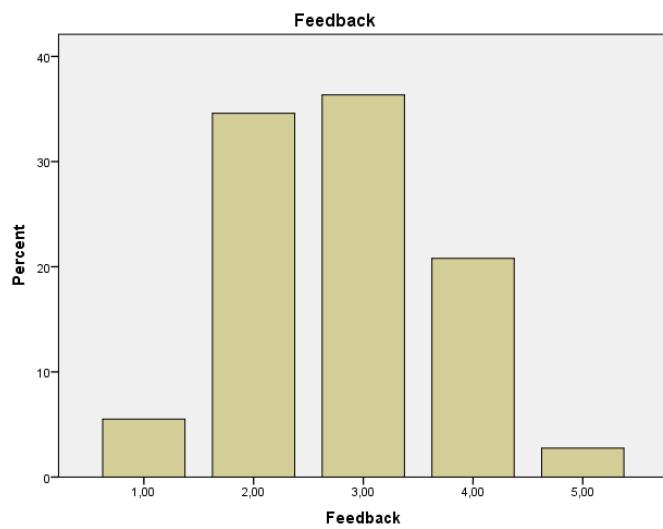
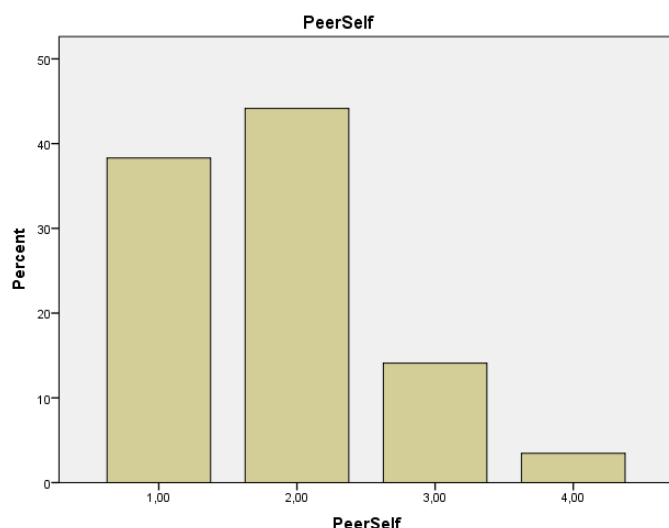


Table 9. Bar plot peer/self-assessment



4.4 Ethical considerations

To guarantee the quality of this research, the research was approved by the Ethic Commission of University of Twente before data collection. This commission checked whether the research is executed following the rules and norms that the university states. Moreover, the schools were asked for approval. Also, the survey contained informed consent given to the participants before distributing the survey. This means that all respondents were informed about the goals and the method of the survey and the autonomy and privacy of the participants is guaranteed. The participation in the survey was on voluntary basis.

5. Results

5.1 DBDM in classrooms

The first part of the first research question concerns the extent to which teachers use DBDM in classrooms. Table 10 shows the minimum, maximum, mean and standard deviation for this factor. It needs to be noted that 1 represents “This happens less than 10% of the lessons” and 5 means “This practice is embedded (happens approximately in 90% of the lessons)”. The standard deviation of .76 meant that there was a big variation between the answers of the teachers. According to the answers of the teachers, on average, this practice is emerging between 25% and 50% of their lessons. This means that teachers generally agree with statements such as: ” I use data to tailor the instruction to the needs of the students”.

Table 10. *Descriptive statistics DBDM*

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
DBDM	434	1.00	4.91	2.82	0.76
Valid N	434				

5.2 AfL in classrooms

For the second part of the first research question regarding the extent to which AfL is used in classrooms, descriptive analyses presenting mean, standard deviation, minimum, maximum and median were conducted again. Table 4 shows the means, standard deviations, minimum, maximum for this factors. The standard deviations were all between .71 and .87 meaning that there was a big variation between the answers of the teachers. Teachers scored the highest on learning intentions, meaning that the practice is emerging between 50% and 75% of their lessons. They also scored relatively high on eliciting evidence, which emerged around 50% of their lessons. This means that, on average, statements such as: "Students are reminded about the links between what they are learning and the big learning picture (e.g., "We are learning to count money so that when we go shopping we can our changes)" occur in their practice. However, teachers scored the lowest on peer/self-assessment meaning that they used this practice sporadically, on average between 10% and 25% or less in their lessons. Therefore, on average, statements such as: "Time is set aside during lessons to allow for self- and peer-assessment" do not occur in their lessons. Feedback was on average used between 25% and 50%.

Table 11. *Descriptive statistics AfL*

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
	434	1.09	5.00	3.18	0.76
Learning intentions	387	1.00	5.00	3.30	0.80
Eliciting evidence	405	1.00	5.00	3.07	0.79
Feedback	399	1.00	5.00	2.82	0.87
PeerSelf	376	1.00	4.17	1.77	0.71

5.3 User characteristics influencing DBDM

For the first part of the second research question regarding the influence user characteristics have on DBDM, linear regression analysis was conducted. Table 5 shows the results of the linear regression analyses regarding the variables influencing DBDM. For the analysis with DBDM as the dependent variable, the results (see Table 5) show that attitude ($p < .201$) does not influence DBDM and knowledge and skills ($p < .000$) significantly influence DBDM. These two variables (knowledge and skills) show an effect that remains significant at the .05 level (two-tailed) when applying the Bonferroni correction. The variables together explained 19.3 % of the variance in DBDM.

5.4 User characteristics influencing AfL

For the second part of the second research question regarding the influence user characteristics have on AfL, linear regression analysis was conducted where each of the strategies of AfL was chosen as dependent. The results (see Table 12) show that only knowledge and skills ($p < .001$) show a significant result. The effect of attitude ($p < .05$) reaches statistical significance only on feedback, but did not reach statistical significance on the other strategies. Attitude ($p < .443$) did not significantly influence peer/self-assessment, neither eliciting evidence ($p < .442$) and learning intentions ($p < .420$). Knowledge and skills significantly influenced peer/self-assessment ($p < .000$) and feedback ($p < .000$). Knowledge and skills significantly influenced also learning intentions ($p < .000$) and eliciting evidence ($p < .000$).

Eliciting evidence explained 20.5% of the variance in a combination of attitude and knowledge and skills. Learning intentions explained 28.3% of the variance in a combination of attitude and knowledge and skills. Feedback explained 12.2% of the variance in combination of attitude and knowledge and skills. Peer/self-assessment explained 12.2% of the variance in combination of attitude and knowledge and skills.

Table 12. *Results linear regression*

	<i>Predictor</i>	β	<i>p</i>	R^2
1. Peer/self assessment	attitude	-0.008	0.890	0.243
	knowledge/skills	0.496	0.000	
2. Feedback	attitude	0.117	0.044	0.29
	knowledge/skills	0.494	0.000	
3. Data use	attitude	0.077	0.201	0.193
	knowledge/skills	0.411	0.000	
4. Learning intentions	attitude	0.026	0.670	0.175
	knowledge/skills	0.409	0.000	
5. Eliciting information	attitude	0.048	0.442	0.122
	knowledge/skills	0.333	0.000	

6. Discussion

Formative assessment has been anticipated as very beneficial regarding better student achievement and school improvement (Van der Kleij et al., 2015). DBDM and AfL are approaches of formative assessment. The aim of the present study was to measure the extent to which DBDM and AfL were used by teachers in their daily practice. This study give a very deep insight not only on the use of formative assessment’s approaches but also on the different strategies which one of the approaches, AfL consist of. In addition, greater insight was gained into how large is the influence of knowledge and skills and dispositions to use data on AfL and DBDM.

6.1 Major findings interpretation and explanation

Formative assessment, if used correctly, could be considered as the answer to some of the education’s biggest issues. It could be the answer to many questions, for example, how to improve the graduation rate, how to decrease the dropout rate, and also how to prepare students from school to their future higher education (Mandinch, 2012). However, DBDM, as an approach to formative assessment, is not yet integrated into daily practice of teachers. DBDM was used by teachers is their lessons between 25% and 50% of the time. This finding is in line with the research of Heitink et al. (2016) and Marsh (2007) who stated that only a minority of teachers actually use formative assessment in their lessons. This might be because still prefer to rely on their intuition to make decisions (Schildkamp & Kuiper, 2010).

The results of this study showed that teachers seem to make greater use of some of the AfL strategies (learning intentions, eliciting evidence) than others (peer/self-assessment).

Learning intentions emerged between 50% and 75% of the lessons and eliciting evidence emerged around 50% of the time in the lessons, feedback emerged between 25% and 50% of the time, and peer/self-assessment is used sporadically, 25% or less in their lessons. To conclude, teachers use most strategies on average only in 25% to 50% in their lessons, therefore it could be stated that formative assessment is not integrated in the daily classroom activities (yet).

It could be argued that the reasoning behind peer/self-assessment being the least used strategy by the teachers is due to the fact that students are mostly seen by teachers as background players and teachers do not feel confident letting students actively participate in the assessment practices. However, according to Schildkamp et al. (2013), teachers can improve their teaching practices exactly by supporting the students in developing the ability to monitor their own learning. Also, peer-assessment strategy is considered as important from researchers (William & Leahy, 2015; Heitink et al., forthcoming; Black & William, 1998; Bryant & Carless, 2009; Crisp, 2012; Harris & Brown, 2013). The reasoning behind that is explained by Black and William (2003) where they state that in order to help students become better learners, they should be given the opportunity to play an active role and talk about their learning and engage in a peer-feedback activity. Moreover, Heitink et al. (forthcoming) argue that peer-assessment can foster the integration of AfL in classrooms. Also, Clark (2012) argue that students can improve their understanding of their learning when they discuss the learning process with their peers. Furthermore, according to Heitink et al. (forthcoming) students valued the assessment techniques which were based on peer and self-assessment. Regarding self-assessment, Harris and Brown (2013) discuss the benefits of self-assessment in order for students to reach better outcomes and this strategy also could help them to take responsibility for their own learning (Clark, 2012).

Concerning the user characteristics that influence DBDM and AfL, our results showed that the attitude of the teachers does not have a significant influence on DBDM and on most of the strategies of AfL. According to the results, attitude influenced significantly only feedback as a formative assessment strategy in the classroom. Contrary to our expectations attitude did not influence significantly the other four strategies. A possible explanation for attitude being a poor predictor for the use of formative assessment in classrooms might be due to the average positive attitude of the teachers in our sample or because attitude, as it is a multidimensional variable, was measured partly. The reasoning behind that is that even when teachers have a positive attitude towards assessment data use in their lessons, it doesn't mean that they have the skills to actually use it effectively. However, the dispositions of the teachers to use data are considered as important in various researches (Heitink et al., forthcoming, Schildkamp et al., forthcoming). It is stated that teachers' beliefs and attitudes have an impact on the successful implementation of formative assessment in classrooms (Heitink et al., 2016).

Contrary to the results we had for attitude, our results showed that knowledge and skills influence significantly all five strategies of AfL. They have a big impact and they explained 0.18% of the variance. To sum up, it could be learned from this study that there is an urgent need for professional development for the schools' staff in order to become data literate. The main reason for that is the teachers need to learn how to use assessment data effectively together with the students in order to achieve school improvement. This is in line with other research findings (Schildkamp et al., 2014; Schildkamp et al., 2015). Only then, when teachers are confident with using data, and when they master the knowledge and skills

needed to use data effectively, it could be expected that formative assessment would be embedded in almost every lesson.

6.3 Limitations of the research strategy and implications for future research studies

The present research is not, of course, without limitations. It is crucial to acknowledge that this study was conducted in a specific context, the Netherlands. As already explained in the theoretical framework, the schools in the Netherlands have a lot of autonomy, therefore, they are more likely to use data (Schildkamp et al., forthcoming; Ebbeler, Schildkamp & Downey, 2012; Earl & Louise, 2012). Even though the goal of this study was not to make firm generalizations, but to gain more insights into the use of assessment data in Dutch schools, we do think that the results of this study could apply to other contexts. This could be done because the original questionnaire was designed to be used in Ireland and it is also used in Australia. The convenience sample chosen for this study also affects the generalizability of the results, as convenience sampling occurs when the participants are the easiest and most convenient to reach.

Another limitation of this research might be that it was performed in schools from only one school board, which might influence the outcome of the results. It might be that the teachers from this particular school board are more oriented towards using formative assessment. Also, some of the teachers from this schools were participating in the data team intervention which means that they are already familiar with using assessment data and are more knowledgeable in that field than teachers who never participated in data teams. Data teams are teams of teachers and school leaders who collaboratively learn how to use data, following a structured approach and guided by a facilitator from the university (Schildkamp & Poortman, 2015). Moreover, given the fact that our study was based on secondary education so maybe the results will differ if the participants were primary teachers, as secondary teachers teach only one subject and primary school teachers teach numerous subjects.

Moreover, the instrument used in the study was a survey and a survey can help us get to know how the teachers who participated in the study perceive the use of formative assessment in their lessons. As formative assessment is happening during the learning process then this learning process needs to be observed. This study is meant as a starting point for a (qualitative) follow-up studies with the use of observations, which holds great promise for providing the researchers with a deeper inside on studying not only the use of formative assessment but also the effects it has on students regarding their achievement based on that use. Future studies can also study more variables (e.g., knowledge and skills of students) using the findings from this paper as a starting point to design such studies. For example, the attention could be mainly on the strategies of AfL as they are all strongly related to the students too. From the results of this study it was found out that Also, other survey could be designed for measuring the perceptions of students on formative assessment rather than only the perceptions of teachers.

Furthermore, we found that a specific strategy needs more attention- peer/self-assessment. Future research could be developed to study how this strategy could be made more attractive for teachers to use more in their lessons. One way to make this strategy more attractive is by supporting the teachers into proving assessment criteria, which is considered as important for the success of peer/self-assessment (Heitink et al., 2016). Heitink et al. (2016) argue of the importance of appropriate training of the students in order to teach them how to use assessment criteria. Once the students master using the criteria, then they will be able to assess their own and their peers' work which, according to Wiliam and Leahy (2015), is going to lead to better student achievement. Then, the students will not be considered anymore as background players by teachers but would be perceived more like partners.

Various studies put attention on the benefits of using formative assessment in classrooms and the fact that formative assessment use is limited. Our study also found that formative assessment is not completely integrated in classrooms. This means that future researches could also study the support which needs to be offered to school staff as well as students in order to increase the effective use of formative assessment. This support might be provided by school leader (Schildkamp & Kuiper, 2010; Schildkamp et al., 2012). According to Schildkamp et al. (2012) school leader is responsible for the support for data use as well as making sure that collaboration between teachers and teachers, and between teachers and school leaders is present. The school leader is also the one who needs to make sure that the teachers have time to use data in their practices and to stress on the importance of it (Schildkamp & Kuiper, 2010). Also, it is essential that data use is encouraged by the school leader in order to be used in classrooms (Schildkamp et al., 2012; Schildkamp & Kuiper, 2010). Furthermore, according to Schildkamp & Kuiper (2010) the school leader needs to be enthusiastic about using data as this also affects the enthusiasm of the teachers to use it in their practice and reflects on their motivation. Motivational characteristics also “determine a person’s intention to engage in behaviour, and therefore, in this case, to use data” (Prenger & Schildkamp, forthcoming, p.5). According to Prenger and Schildkamp (forthcoming), it is important that psychological factors are also taken into account to increase teachers’ implementation of data use for instruction and from that it follows better school improvement. Their study indicated that the teachers “perceived control of data use, their attitude regarding its benefits and consequences and their intention to use data positively influence their instructional data use” (Prenger & Schildkamp, forthcoming, p. 7).

6.4 Implications for practice

In order for teachers to become data literate, a lot of effort is needed and it is not realistically to consider this possible as long as there is no specific professional development in this direction. Also it is more constructive if the teachers are not only pressured in using assessment data but also they are offered support on doing that. The results of this study show that professional development in the use of formative assessment is urgently needed. In our view an important point of action might be a training specifically designed to support teachers in assessment data use in their everyday practice. An investment in professional development and training of the school staff is crucial so that they know how to collect, analyse, interpret and use data in order to make informed decisions and to be more literate. Schildkamp et al. (2013) also argue that training and facilitation are very essential in order to create experts in using assessment data in schools. Yet, more research is needed into the design, development, implementation and evaluation of professional development for (assessment) data use.

The results of this study are going to be reported to the schools’ staff of the schools which participated. They might give the teachers a perception of the extent to which they use formative assessment in their lessons. As AfL was not used effectively enough, which they will find out from the reported results, might provoke them to become more data literate as well as more critical to their AfL understanding. This could happen by attending professional training, for example.

Also, the survey used in this study, could be used as an assessment instrument by school staff in schools and not only by researchers. By using the survey as an assessment instrument, the school staff will become aware of the extent to which formative assessment is being used by the teachers from the particular school. Then, based on the results, the teachers from the schools where formative assessment is not used effectively, might be given the

opportunity to join a training in order to learn how to use formative assessment but also to understand the benefits of using it.

6.5 Conclusions

According to literature, formative assessment is considered as very beneficial for increased student achievement. Yet, it is not actively and effectively used in classrooms by most teachers. This study provides an overview of the use of two of its approaches, DBDM and AfL and of the user characteristics that influence those two approaches. The results that are presented in this paper call for attention in designing a training for teachers so that they know how to use (assessment) data together with students in classrooms. This paper also offers a starting point for future research and it can be used as a guide by researchers. It is important to remember that formative assessment leads to improved student achievement only if it is successfully implemented.

References:

- Bennett, E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice* 18, no. 1: 5-25.
- Black, P., & Wiliam, D. (2003). In praise of educational research: formative assessment. *British Educational Research Journal*, 29(5), 623-637.
<http://dx.doi.org/10.1080/0141192032000133721>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
<http://dx.doi.org/10.1080/0969595980050102>
- Bryant, D., & Carless, D. (2009). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? *Educational Research Policy Practice*, 9(1), 3-15.
<http://dx.doi.org/10.1007/s10671-009-9077-2>
- Campbell, C., & Levin, B. (2008). Using data to support educational improvement. *Educational Assessment Evaluation Access*, 21(1), 47-65.
<http://dx.doi.org/10.1007/s11092-008-9063-x>
- Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students' learning needs with technology. *Journal of Education for Students Placed at Risk*, 10(3), 309-332.
doi:10.1207/s15327671espr1003_6
- Clark, I. (2012). Formative assessment: assessment is for self-regulated learning. *Educational Psychology Review*, 24, 205-249.
- Coburn, C., & Turner, E. (2011). Research on data use: a framework and analysis. *Measurement: Interdisciplinary Research & Perspective*, 9(4), 173-206.
<http://dx.doi.org/10.1080/15366367.2011.626729>
- Crisp, G. (2012). Integrative assessment: reframing assessment practice for current and future learning. *Assessment & Evaluation In Higher Education*, 37(1), 33-43.
<http://dx.doi.org/10.1080/02602938.2010.494234>
- Gottheiner, D., & Siegel, M. (2012). Experienced middle school science teachers' assessment literacy: investigating knowledge of students' conceptions in genetics and

- ways to shape instruction. *J Sci Teacher Educ*, 23(5), 531-557.
<http://dx.doi.org/10.1007/s10972-012-9278-z>
- Daly, A. (2012). Data, dyads, and dynamics: exploring data use and social networks in educational improvement. *Teachers College Record*, 114,110305, 38-38.
- Fletcher, A., & Shaw, G. (2012). How does student-directed assessment affect learning? Using as a learning process. *International Journal of Multiple Research Approaches*, 6, 245-263. doi:10.5172/mra.2012.6.3.24510.1037/0022-3514.45.2.357
- Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: reconceptualising assessment for learning. *Oxford Review of Education*, 39, 229-246.
<http://dx.doi.org/10.1080/03054985.2013.787922>.
- Hargreaves, A., & Braun, H. (2013). *Data-driven improvement and accountability*. Boston, MA: National Education Policy Center. Retrieved October 24, 2013.
- Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111.
<http://dx.doi.org/10.1016/j.tate.2013.07.008>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81- 112. doi:10.3102/003465430298487
- Heitink, M., Van der Kleij, F., Veldkamp, B., Schildkamp, K., & Kippers, W. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62.
<http://dx.doi.org/10.1016/j.edurev.2015.12.002>
- Hoogland, I., Schilkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B. & Dijkstra, A. (submitted). Prerequisites for data-based decision making in the classroom: a practical literature review.
- Hornby, W. (2003). Assessing using grade-related criteria: a single currency for universities? *Assessment & Evaluation in Higher Education*, 28(4), 435-454.
<http://dx.doi.org/10.1080/0260293032000066254>

- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in: Principles, Policy and Practice*, 20, 150-169.
doi:10.1080/0969594X.2012.749212
- 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* (Vol. 106, pp. 105–131). Malden, MA: Wiley-Blackwell.
- Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *American Journal of Education*, 112, 496–520.
- Kerr, K., Marsh, J., Ikemoto, G., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement. Actions, outcomes, and lessons from three urban districts. *American Journal of Education* 112(4), 496-520.
- Mandinach, E. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educational Psychologist*, 47(2), 71-85.
<http://dx.doi.org/10.1080/00461520.2012.667064>
- Marsh, C. (2007). A critical analysis of the use of formative assessment in schools. *Educ Res Policy Prac*, 6(1), 25-29. <http://dx.doi.org/10.1007/s10671-007-9024-z>
- Marshall, B., & Drummond, M. (2006). How teachers engage with assessment for learning: lessons from the classroom. *Research Papers in Education*, 21, 133-149.
<http://dx.doi.org/10.1080/02671520600615638>.
- Lysaght, Z. & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, 32:2, 217-232, DOI:
10.1080/03323315.2013.784636
- Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: district and school connections in data-driven decision making. *School Leadership and Management*, 29(5), 477–494. doi: 10.1080/13632430903162541
- Peterson, E., & Irving, S. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238-250.

- Prenger, R. & Schildkamp, K. (submitted) Data-based decision making for teacher and student learning: a psychological perspective on the role of the teacher.
- Schildkamp, K. (2007). The utilisation of a self-evaluation instrument for primary education. Enschede: University of Twente (published thesis).
- Schildkamp, K., Lai, M. K., & Earl, L. (2013). *Data-based decision making in education. Challenges and opportunities*. Dordrecht: Springer Science + Business Media.
- Schildkamp, K., & Kuiper, W. (2010). Data- informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and teacher education*, 26(3), 482-496.
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (submitted) Factors promoting and hindering data-based decision making in schools.
- Schildkamp, K., Ehren, M., & Lai, M. (2012). Editorial article for the special issue on data-based decision making around the world: from policy to practice to results. *School Effectiveness and School Improvement*, 23(2), 123-131.
<http://dx.doi.org/10.1080/09243453.2011.652122>
- Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation*, 42, 15-24.
- Scriven, M. (1971). General strategies in evaluation. *Curriculum Theory Network*, (8/9), 182.
<http://dx.doi.org/10.2307/1179202>
- Sharkey, N. S., & Murnane, R. J. (2006). Tough choices in designing a formative assessment system. *American Journal of Education*, 112, 572–588.
- Scheerens, J., Ehren, M., Slegers, P., & de Leeuw, R. (2012). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*. The Netherlands.
- Supovitz, J. A., & Klein, A. (2003). *Mapping a course for improved student learning: How innovative schools systematically use performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania Graduate School of Education.

- Timmers, C. F., & Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers in Education, 56*(3), 923e930. <http://dx.doi.org/10.1016/j.compedu.2010.11.007>.
- Van der Kleij, F., Vermeulen, J., Schildkamp, K., & Eggen, T. (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. <http://dx.doi.org/10.1080/0969594x.2014.999024>
- Verhaeghe, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2010). Using school performance feedback: Perceptions of primary school principals. *School Effectiveness and School Improvement, 21*(2), 167–188.
- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the Data-Informed District. *School Effectiveness and School Improvement, 23*(2), 159–178.
- Wayman, J. C., Cho, V., & Johnston, M. T. (2007). *The Data-Informed District: A district-wide evaluation of data use in the Natrona County school district*. Austin: The University of Texas.
- Wiliam, D. 2011. *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment*. West Palm Beach, FL: Learning Sciences.
- 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.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*, 521–548.

Appendix A: Results of the confirmative factor analyses and reliability analyses

Factor	Cronbach's alpha
<i>Learning intentions (0.76)</i>	
Learning intentions are stated using words that emphasise knowledge, skills, concepts and/or attitudes, i.e. what the pupils are learning NOT what they are doing	(0.75)
Pupils are reminded about the links between what they are learning and the big learning picture (e.g. 'We are learning to count money so that when we go shopping we can check our change')	(0.73)
Success criteria related to learning intentions are differentiated and shared with pupils	(0.68)
Child-friendly language is used to share learning intentions with pupils (e.g. 'We are learning to make a good guess (prediction) about what is likely to happen next in the story')	(0.66)

Factor	Cronbach's alpha
<i>Eliciting evidence (0.83)</i>	
Assessment techniques are used to facilitate class discussion (e.g. brainstorming).	(0.80)
Questions are used to elicit students' prior knowledge on a topic.	(0.80)
Students are encouraged to share the questioning role with teacher during lessons (e.g., the teacher routinely invites pupils to question their peers' contributions to discussions).	(0.81)
Questioning goes beyond the one right answer style (where the focus is often on trying to guess the answer in the teacher's mind) to the use of more open-ended questions that encourage critical thinking.	(0.79)
Students' incorrect responses are used to guide teaching and learning (e.g., a pupil is asked to explain why he/she gave a particular answer).	(0.81)
Students are encouraged to share the questioning role with teacher during lessons (e.g., the teacher routinely invites pupils to question their peers' contributions to discussions).	(0.81)

Factor	Cronbach's alpha
<i>Feedback (0.81)</i>	

Written feedback on pupils' work goes beyond the use of grades and comments such as "well done" to specify what students have achieved and what they need to do next.	(0.81)
Diagnostic information from standardised tests is used to identify strengths and needs in teaching (e.g. common errors in the comprehension section of the Mary Immaculate College Reading Attainment Tests [MICRA-T] are identified and used in teaching).	(0.74)
Diagnostic information from standardised tests is used to identify strengths and needs in learning	(0.73)
Diagnostic information from standardised tests is used to identify strengths and needs in teaching and learning (e.g., common errors in NAPLAN tests are identified and used in teaching).	(0.77)
The teachers use information about the progress in the learning process of the students into giving feedback to the students	(0.80)

Factor	Cronbach's alpha
<i>Peer/self-assessment (0.82)</i>	
Students are given an opportunity to indicate how challenging they anticipate the learning will be at the beginning of a lesson or activity (e.g., by using traffic lights).	(0.80)
Students are encouraged to record their progress using, for example, learning logs.	(0.78)
Students assess and comment on each other's work (e.g., they are taught how to use the success criteria of a lesson to judge another pupil's piece of work).	(0.78)
Pupils are encouraged to use a range of assessment techniques to review their own work (e.g. a rubric, traffic lights, thumbs up/down, two stars and a wish).	(0.78)
A visual record of students' progress is maintained to celebrate students' learning and show areas of/for development (e.g., a bulletin board displaying progression in story writing over a term).	(0.79)
Time is set aside during lessons to allow for self- and peer-assessment.	(0.78)

Factor	Cronbach's Alpha
<i>Data use for instruction (0.89)</i>	
To what extent do you use data to:	
set learning goals/targets for individual students	(0.88)
determine which topics and skills students do and do not possess	(0.88)
determine progress of students	(0.89)
tailor instruction to individual students' needs	(0.88)
set the pace for my lessons*	(0.89)
give student feedback on their learning process	(0.88)
form small groups of students for targeted instruction	(0.89)
identify instructional content to use in class	(0.89)
study why students make certain mistakes*	(0.89)
adapt instruction based on the needs of gifted students	(0.88)
adapt instruction based on the needs of struggling students	(0.88)

Factor	Cronbah's alpha
<i>Knowledge and skills (0.81)</i>	
I am able to adjust my instruction based on data	(0.76)
I am able to use (assessment) data to diagnose student learning needs	(0.76)
I understand the quality criteria and concepts for (assessment) data use (for example correlation, validity, reliability)	(0.79)
I know how to interpret data and reports I receive (for example exam results, student achievement results from previous years)	(0.80)
I am comfortable in interpreting (assessment) data that are presented in graphs	(0.80)
I have the skills to use (assessment) data to tailor and differentiate my instruction to the needs of the students	(0.78)
In my classes I use a structured method to analyse and interpret (assessment) data for taking action	(0.80)

Factor	
<i>Dispositions to use data (0.83)</i>	
I believe it is important to use data in determining individual students needs	(0.79)

I believe that the use of different type of (assessment) data is important for understanding the learning process of my students	(0.78)
I believe that students benefit when instruction is based on data	(0.79)
I believe that data are important in changing my teaching	(0.78)