Chapter 13

Towards an Integrative Formative Approach of Data-Driven Decision Making, Assessment for Learning, and Diagnostic Testing

Jorine A. Vermeulen and Fabienne M. van der Kleij

Abstract This study concerns the comparison of three approaches to assessment: Data-Driven Decision Making, Assessment for Learning, and Diagnostic Testing. Although the three approaches claim to be beneficial with regard to student learning, no clear study into the relationships and distinctions between these approaches exists to date. The goal of this study was to investigate the extent to which the three approaches can be shaped into an integrative formative approach towards assessment. The three approaches were compared on nine characteristics of assessment. The results suggest that although the approaches seem to be contradictory with respect to some characteristics, it is argued that they could complement each other despite these differences. The researchers discuss how the three approaches can be shaped into an integrative formative approach towards assessment.

Keywords: Formative Assessment, Data-Driven Decision Making, Assessment for Learning, Diagnostic Testing

Introduction

Within the various approaches to assessment, the importance of increasing student learning by acting on students’ educational needs is emphasized. However, the meaning ascribed to student learning is one of the factors that separates these approaches to assessment: a distinction is made between learning outcomes and the process of learning. This means that some assessment approaches focus on what has to be learned, while other approaches focus on how students learn what has to be learned (best) and the quality of the learning process. Furthermore, assessment approaches differ in the aggregation levels in the educational system (e.g., student, classroom, school) at which the assessment is aimed. Due to these differences, the approach to assessment that is chosen affects the strategies that are used to assess and promote student learning outcomes.
This chapter addresses the differences and similarities of the three approaches to assessment that are currently most frequently discussed in educational research literature. The first approach is *Data-Driven Decision Making* (DDDM), which originated in the United States of America as a direct consequence of the No Child Left Behind (NCLB) Act in which improving students’ learning outcomes is defined in terms of results and attaining specified targets. Secondly, *Assessment for Learning* (AfL), originally introduced by scholars from the United Kingdom, is an approach to assessment that focuses on the quality of the learning process, rather than merely on students’ (final) learning outcomes (Stobart, 2008). Finally, in *Diagnostic Testing* (DT), also referred to as diagnostic measurement, students’ learning outcomes are described as students’ learning processes and factors that resulted in students’ success or failure to do particular tasks (Leighton & Gierl, 2007a, 2007b; Rupp, Templin, & Henson, 2010).

Although all three approaches claim to provide information that can be used to increase students’ learning outcomes, there appears to be no clear study into the relations and distinctions between these approaches. More specifically, these terms are often used interchangeably. Interestingly, the literature on DDDM tends to cite literature concerning AfL, but not vice versa (e.g., Swan & Mazur, 2011). The aim of this study is to investigate the extent to which DDDM, AfL, and DT can be integrated into an approach that can be used for maximizing student learning. By maximizing student learning we aim at both the process and outcomes of learning as optimized at all levels of education. The following research question will be answered: To what extent can DDDM, AfL, and DT be shaped into an integrative formative approach towards assessment?

**The Formative Function**

Tests are a crucial part of education, namely, it would not be possible to check whether a certain instructional activity led to the realization of the intended learning outcomes without testing (Wiliam, 2011). Paper-and-pencil tests are often used within the classroom to gather information about student learning. Besides paper-and-pencil tests, there are various other methods for measuring pupils’ knowledge and abilities. For example, homework, projects, discussions, and observations can provide valuable information about student learning. Whenever such a broad spectrum of instruments is used for gathering information about student learning one speaks of assessment (Stobart, 2008). Traditionally, in education, a distinction is made between summative and formative assessment.
Summative assessments can be used to judge or compare the learning outcomes of students, based on which a decision is made with regard to, for example, selection, classification, placement, or certification. There are also tests that have the purpose of directing the learning process. These tests are called formative tests (Sanders, 2011). However, a test is not in itself formative or summative by definition (Stobart, 2008). Whether a test is used formatively does not depend on the characteristics of the test itself, but on the way the test results are being used, in other words, the function of the test results (Harlen & James, 1997; Stobart, 2008). Whenever a test result plays a role in making (part of the) pass/fail decision, it fulfills a summative function. The same test, however, can also fulfill a formative function, for example, when feedback is provided to students that can be used in future learning. Another example of a formative use is a teacher using test results to evaluate the effectiveness of the instruction. Subsequently, the teacher might make amendments to the educational program in order to meet the needs of the learners.

Formative assessment is a broad concept that has many definitions (e.g., AfL, and diagnostic assessment; Bennett, 2011; Johnson & Burdett, 2010). Initially, the formative concept was introduced by Scriven (1967) to indicate interim evaluation of intervention programs. In 1968, the formative concept was first used in the context of instruction by Bloom. In the years that followed, various meanings have been ascribed to formative assessment. Recently, researchers have come to the insight that a distinction between formative and summative assessment based on time-related characteristics is not a useful one. After all, it is the way that test results are eventually used that determines the purpose the test serves (Stobart, 2008).

Subsequently, for comparison of assessment approaches it is useful to distinguish between formative program evaluations and formative assessments (Shepard, 2005; Harlen, 2007). Formative program evaluations are meant to make decisions at a higher aggregation level than the level of the learner (e.g., classroom or a school) about the educational needs of pupils. Formative assessment, on the contrary, concerns decisions at the level of the learner. Results from formative assessments are used to accommodate the individual educational needs of pupils. In this study, we refer to formative evaluation as the evaluation of the quality of education. We will use the term formative assessment to indicate assessment that takes place within the classroom and is focused on improving instruction in the classroom and for individual pupils. In the following sections, we elaborate on the characteristics of the three views on assessment, after which the three approaches are compared.
Data-Driven Decision Making

Teachers make most of their decisions based on their intuition and instincts (Slavin, 2002, 2003). However, educational policies such as NCLB have caused an increase in accountability requirements, which has stimulated the use of data for informing school practice in the United States of America. The main idea behind NCLB is that by setting standards and measurable goals the learning outcomes of pupils can be raised. Namely, research has pointed out that by setting specific learning goals an increase in student achievement can be obtained (Locke & Latham, 2002). The idea of using data for informing instruction is not new, namely, in the 1980s there was an approach that attempted to make instruction more measurement-driven (Popham, Cruse, Rankin, Sandifer, & Williams, 1985). Furthermore, recent studies draw attention to the importance of using data, such as assessment results and student surveys, in making decisions (Wayman, Cho, & Johnston, 2007; Wohlstetter, Datnow, & Park, 2008). When data about students are used to inform decisions in the school it is referred to as DDDM (Ledoux, Blok, Boogaard, & Krüger, 2009).

Recently, DDDM has gained popularity in the Netherlands (the Dutch term that is often used is opbrengstgericht werken), which is seen as a promising method for increasing pupils' learning outcomes (Ledoux et al., 2009). Schildkamp and Kuiper (2010) have defined DDDM as "systematically analyzing existing data sources within the school, applying outcomes of analyses to innovate teaching, curricula, and school performance, and, implementing (e.g., genuine improvement actions) and evaluating these innovations" (p. 482). The data sources that can be used to inform decisions are not only results from tests. Other usable data sources are school self-evaluations, characteristics of the pupils in the school, results from questionnaires taken by parents or pupils, and various assessments sources, such as an external national test or internal school assessments. In this study, for the sake of comparing the three approaches, we will focus on the use of student achievement results for informing the decision-making process. Data about student achievement will be referred to as data-feedback.

Many schools already possess data-feedback, for example from a student monitoring system. These data are often systematically collected via standardized tests and can therefore be valued as objective. Besides these objective data, teachers possess data-feedback from daily practice that has been gathered using various assessment methods. When data-feedback is used in the right way, meaning that its use will lead to education that is more adequately adapted to the needs of the learner, this will eventually lead to better learning results.
At the student and classroom level, data-driven decision making can be a valuable instrument for using assessment results in a formative way. Assessment results are an important source of information about how learning processes could be improved for both students and teachers. Students need feedback to choose the most suitable learning strategies in order to achieve the intended learning outcomes, while teachers need data-feedback in order to act on the pupils' current points of struggle and to reflect on their own teaching practices. However, knowledge about how teachers use assessment results for instructional improvement is limited (Young & Kim, 2010).

According to Wayman (2005), the NCLB policy carries the assumption that whenever data is available it will lead to changes in teaching practice. A general definition of educational measurement comprises four activities that are part of a cyclic process of evaluation: "...designing opportunities to gather evidence, collecting evidence, interpreting it, and acting on interpretations" (Bennett, 2011, p. 16). However, it is not always self-evident for practitioners how (accountability) data should be translated into information that can be readily used to make decisions in the school. Therefore, it is not surprising that recent studies suggest that data-feedback is underused in the majority of Dutch schools (Ledoux et al., 2009; Schildkamp & Kuiper, 2010). These studies have found that the implementation of the evaluative cycle is incomplete in many schools. The results of these studies imply that students are frequently assessed and that the results of the assessments are registered, but that there is no subsequent use of the data-feedback. Moreover, research has also pointed out that many teachers indeed get stuck in the interpretation phase of the evaluative cycle (Meijer, Ledoux, & Elshof, 2011). Thus, educators need to know how to translate raw assessment data into knowledge about student learning outcomes that indicates in what way students’ learning outcomes can be optimized.

The literature makes a distinction between data and information (Davenport & Prusak, 1998; Light, Wexler, & Heinze, 2004; Mandinach, Honey, & Light, 2006). Data are characterized as objective facts, which have no meaning. By interpreting data, they can be transformed into information, for example by summarizing, contextualizing and calculating (Davenport & Prusak, 1998). Subsequently, information can be transformed into actionable knowledge by synthesizing and prioritizing. This actionable knowledge is the basis for a decision about which action to undertake (Light et al., 2004).

The impact of the action is evaluated by gathering new data, in this way a feedback loop is created (Mandinach et al., 2006). However, this is not an easy process because teachers are used to making decisions intuitively (Slavin, 2002, 2003).
"As many educators say, they are data rich, but information poor. By this they mean that there is far too much information with which they must deal, but those data are not easily translatable into information and actionable knowledge" (Mandinach et al., 2006, p. 12).

Within the school, many feedback loops can exist. The frequency in which the feedback loops are completed depends, among other things, on the type of data-feedback that is used. Data-feedback from formal and objective tests, for example from a student monitoring system, are less frequently available than data from informal assessment situations, such as homework assignments. Ledoux et al. (2009) stated that the quality of the evaluative cycle is dependent upon the quality of the data-feedback. This implies that unreliable data can lead to making unjustified decisions. Moreover, wrongly interpreting data can also lead to making unjustified decisions. Being data-literate is thus a necessary prerequisite for successfully implementing DDDM.

Assessment for Learning

AfL focuses specifically on the quality of the process of learning instead of the outcomes of the learning process. Moreover, "it puts the focus on what is being learned and on the quality of classroom interactions and relationships" (Stobart, 2008, p. 145). The theory of AfL has no strict boundaries; it is part of a bigger entity in which curriculum, school culture, and instruction approaches intervene. It is noteworthy that AfL is viewed as a divarication of formative assessment (Johnson & Burdett, 2010; Stobart, 2008).

The Assessment Reform Group (2002) defined AfL as follows: "Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there".

Klenowski (2009) reported on, what she named a 'second-generation definition' of AfL, which was generated at the Third Assessment for Learning Conference (2009): "AfL is 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). This 'second-generation definition' was needed because definitions, or parts of them, are often misinterpreted (Johnson & Burdett, 2010; Klenowski, 2009).
AfL takes place in everyday practice, which means the process is characterized by dialogues between learners and the teacher. This also means that assessments are integrated into the learning process, which is contrary to the traditional approach where assessments are a separate activity that stands apart from instruction. Furthermore, Klenowski’s (2009) definition emphasizes the role of the learners in the learning process and their autonomy. It also emphasizes the nature of AfL in terms of making decisions about which steps to take in the learning process. The information that is used to inform decisions can come from various assessment sources, such as dialogues and observations. These events can be both planned and unplanned. This implies that the evidence that is gathered about the learning process of the learners can be both qualitative and quantitative in nature. The last part of the definition stresses the formative function of assessments. One can only say an assessment serves a formative function when students and teachers use the information for informing and enhancing learning. Therefore, a crucial aspect of AfL is feedback, which is used to direct future learning (Stobart, 2008). The AfL approach encompasses more than the use of assessments and their results. The Assessment Reform Group (1999) described AfL using five core features:

1. Learners are actively engaged in their learning process;
2. Effective feedback is provided to learners;
3. Instructional activities are being adapted based on assessment results;
4. Learners are able to perform self-assessment;
5. The influence of assessment on motivation and confidence of learners is acknowledged.

Stobart (2008) argued that AfL is a social activity that influences both learner identity and the type of learning that will take place.

Whenever the primary goal of testing is to measure a result this can lead to a misleading image of what the level of the learner is, because one cannot measure the full scope of a curriculum. Moreover, Stobart states that on some standardized tests results can improve without students actually learning more. Also, when too much emphasis is placed on achieving specific goals, by frequently using these types of standardized tests, the learning process is narrowed and teaching to the test is promoted (Stobart, 2008), which elicits surface learning (Harlen & James, 1997).
Therefore, multiple methods for gathering evidence about student learning will lead to a more complete picture of students’ knowledge and abilities (Harlen & Gardner, 2010), and are needed in order to achieve deep learning (Harlen & James, 1997).

Hargreaves (2005) compared various definitions of AfL in the literature and interpreted definitions as formulated by teachers and head teachers in a survey. She concluded that there are two approaches within AfL: a measurement and an inquiry approach. In the measurement approach, AfL is viewed as an activity that includes marking, monitoring, and showing a level. In this view, (quantitative) data are used to formulate feedback and to inform decisions. Assessment is seen as a separate activity to show that a predetermined level has been achieved. On the contrary, in the inquiry approach, AfL is a process of discovering, reflecting, understanding and reviewing. It is very much focused on the process and assessments are integrated into the learning process. Qualitative sources of information play an important role. In both approaches, feedback is used to steer future learning. However, in the first approach, feedback might be less immediate and less suited to meet the needs of the learners because of the more formal character of the assessments.

In our opinion, the measurement approach towards AfL can easily turn into a misinterpretation. To illustrate our point of view we will take the example of monitoring (personal communication, M. Johnson, January 3, 2012). Monitoring in itself does not have to be bad in the light of AfL. Teachers can keep accurate track of students’ learning progress and this can help them to make more informed decisions about their students. This way, quantitative data can be used to inform qualitative actions, such as providing elaborated feedback. However, as soon as the monitoring takes place at a level higher than the class level, the measures used to monitor student learning lack a contextual link to the particular performances to which they relate. In other words, the distance between the monitoring action and the learner is too large, which makes it lose its qualitative potential. At this level, monitoring is reduced to ticking boxes, which is at odds with the spirit of AfL (Johnson & Burdett, 2010).

**Diagnostic Testing**

Making diagnoses originates from the field of physical and mental health care in which the aim is to diagnose a disease or disorder and to advise on the treatment (De Bruyn, Ruijssenaars, Pameijer, & Van Aarle, 2003; Kievit, Tak & Bosch, 2002).
In education, DT was initially used for identifying students who were unable to participate in mainstream education because of their special educational needs. Currently, DT is still used for the identification of students with learning deficiencies and/or behavioral problems. However, it is currently also believed that for instruction to be effective educators need to take into account all students’ learning needs (Wiliam, 2011).

Before addressing the process of diagnosing the educational needs of students, the distinction between formative assessment and DT will be explained, as well as the difference between DT and diagnostic assessment. Articles on formative assessment sometimes use the concept of diagnosing when explaining the purpose of formative assessment (e.g., Black & Wiliam, 1998). Similarly, in the literature on DT (or assessment), DT has been defined as equal to formative assessment (e.g., Turner, VanderHeide, & Fynewever, 2011). However, not every diagnostic test fulfills a formative function because, as we explained in the sections above, whether or not a test serves a formative purpose depends on how the results of that test are used. According to Keeley and Tobey (2011), DT is mainly concerned with measuring students’ preconceptions and reasoning styles. This includes the identification of the use of inadequate reasoning styles, and skipped or wrongly executed procedural steps as a result of, among other things, misconceptions. Formative tests, on the other hand, take into account any information about students’ learning outcomes that can be used for adapting instruction and providing feedback. In our view, a diagnostic test fulfills a formative function when the diagnosis is used to optimize students’ learning processes.

Moreover, we make a distinction between DT and diagnostic assessment. DT refers to the use of computerized (adaptive) diagnostic tests, whereas diagnostic assessment refers to the use of various assessment methods, such as diagnostic interviews (Moyer & Milewicz, 2002). Diagnostic assessment is very time-consuming and labor-intensive, and mainly uses qualitative methods. On the other hand, DT is less time-consuming and labor-intensive for teachers, because students can work independently and the test results are delivered automatically in a format that can readily be used to support decision making.

From a diagnostic testing point of view, the development of diagnostic tests requires (statistical) measurement models that fulfill the role of the diagnostician. It also means that data gathered with these tests are likely to be more objective and reliable than alternative assessment methods.
And also, even though data obtained with a diagnostic test have a quantitative nature, the reportage of that test can be the qualitative interpretation of those data. It goes beyond the scope of this chapter to elaborate on these measurement models, therefore we refer the interested reader to Leighton and Gierl (2007b) and Rupp, Templin, and Henson (2010).

The utility of DT stems from its potential to inform instructional decisions by providing information about students’ learning needs. Stobart (2008, p. 55) described diagnosing learning needs as: "...[identifying] how much progress can be made with adult help...". This statement is in accordance with Vygotsky’s criticism on the test culture of the 1930s.

He believed that in order to promote student learning, tests should focus on what students are able to learn, rather than what they have learned so far (Verhofstadt-Denève, Van Geert, & Vyt, 2003). In other words, tests should assess a student’s zone of proximal development (ZPD); which is defined as what a student can do with the minimal help of adults or peers (Verhofstadt-Denève et al.).

There are multiple ways of assessing the ZPD, for example, a diagnostic test may include tasks in which the student is offered help in terms of feedback when an incorrect answer is given (e.g., a number line in an arithmetic test). This kind of help is also known as scaffolding (Verhofstadt-Denève et al., 2003). Note that a diagnostic test will become more similar to what can be described as learning material when students receive rich feedback about their mistakes.

Another method to assess the ZPD is to diagnose students’ line of reasoning when solving the tasks within the diagnostic test. From a cognitive psychological point of view, student ability is more than being able to solve specific tasks; how students derive the answers to the items is viewed as an important indicator of students’ ability levels. For this method of DT a cognitive diagnostic model is necessary, meaning a theory about how students with different ability levels solve the tasks that are the object of assessment (Leighton & Gierl, 2007a, 2007b). Such a theory includes commonly developed misconceptions and errors that are frequently observed, and their relation to different ability levels. Because students within similar cultural and educational contexts are likely to follow comparable learning processes, it is more efficient to focus on frequently observed errors than on identifying students’ errors that are uncommon or the result of ‘slips’ (Bennett, 2011). Additionally, Bennett explained that the identification of ‘slips’ has less diagnostic value because they cannot inform teachers about how to change their practice.
Whatever method for developing a diagnostic test is used, it is the process of developing "...items (i.e., tasks, problems) that can be used to efficiently elicit student conceptions that these conceptions can be related back to a hypothesized learning progression" (Briggs & Alonzo, 2009, p. 1). A similar cyclic process of formulating and testing hypotheses about the nature of students’ achievements is followed within child health care services. Although variations might exist, the diagnostic cycle consists of the following four phases (De Bruyn et al., 2003; Kievit et al., 2002):

1. Identifying the problem (complaint analysis);
2. clarifying the problem (problem analysis);
3. explaining the problem (diagnosing); and
4. indication of treatment possibilities (advising and deciding).

An example of diagnosing mathematics difficulties following phases similar to the four phases of the diagnostic cycle is described by Rupp et al. (2010, p. 14).

Table 1 illustrates how each phase of the diagnostic cycle could be used in educational contexts. The first three phases result in rich descriptions of the student’s knowledge and skills, whereas the fourth phase will result in a diagnosis that prescribes which decision concerning the learning environment will have the highest probability of successful outcomes for the student. As explained in the second paragraph of this section, a diagnostic test can only have a formative function when it is used to improve students’ learning processes. As shown in Table 1, this is the case when Phase 4 is completed. Furthermore, we consider DT an approach to assessment that primarily focuses on the learning processes of individual students. For that reason, we frequently refer to the student. However, from a practical point of view, it might be more feasible for teachers to address the needs of (small) groups of students. This could be done by classifying students into, for example, instruction groups based on the similarity of their diagnosis. However, caution is needed with this approach because (depending on the measurement model used) DT is not meant for comparing students. Also, because of the degree of detailed information, it might be difficult to group students with a similar diagnosis who would benefit from the same adaptations to the learning environment.
### Table 1 Objectives and Outcomes of the Four Phases of the Diagnostic Cycle

<table>
<thead>
<tr>
<th>Diagnosis Type</th>
<th>Objective</th>
<th>Outcome(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acknowledging (Phase 1)</td>
<td>Assessing whether the student’s learning process is optimized given the student’s characteristics and the characteristics of the learning environment. With this diagnosis it can only be decided whether a student might benefit from adaptations to the learning environment.</td>
<td>The probability of the student reaching learning goals given his/her current test performance on tasks that are associated with knowledge and abilities necessary for achieving those goals.</td>
</tr>
<tr>
<td><strong>Exploratory</strong> (Phase 2)</td>
<td>Describing student’s current achievement on the test in terms of strengths and weaknesses.</td>
<td>A list of strengths (i.e., things he can do or knows that are improbable based on his overall performance on other domains) and a list of weaknesses (i.e., things he cannot do or does not know that are improbable based on his overall performance on other domains).</td>
</tr>
<tr>
<td><strong>Explanatory</strong> (Phase 3)</td>
<td>Investigating which hypotheses about the nature of the student’s achievements on the test are most probable.</td>
<td>A description concerning which errors the student has made and why these errors occurred. The student report might describe the relation between the student’s errors, misconceptions, as well as his wrongly applied or skipped procedural steps.</td>
</tr>
<tr>
<td><strong>Prescriptive</strong> (Phase 4)</td>
<td>a. Indication</td>
<td>a. Determining which intervention or changes to the learning environment are most likely to be effective in optimizing the student’s learning process.</td>
</tr>
<tr>
<td></td>
<td>b. Selection</td>
<td>b. Determining whether the student should be selected.</td>
</tr>
<tr>
<td></td>
<td>c. Classification</td>
<td>c. Determining to which group the student belongs.</td>
</tr>
</tbody>
</table>

### Comparing the Three Approaches to Assessment

Rupp, Templin, and Henson (2010, p. 12) organized various uses of diagnostic assessment into seven categories:

1. The object of the assessment;
2. the object of the decision;
3. the time point in which decisions need to be made;
4. the objective of the assessment;
5. the assessment methods;
6. the types of intervention; and
7. the power differentials between agents.

In order to compare DDDM, AfL, and DT we applied these characteristics of assessment to all three approaches. This method enables us to systematically identify differences and similarities between the approaches. Additionally, we added ‘characteristics of the assessment process’ and ‘learning theory’. Also, we broke down some of the features to smaller aspects. The explanation of the assessment features and their aspects are described in Table 2.

**Table 2 Characteristics of Assessment Used to Compare the Three Assessment Approaches**

<table>
<thead>
<tr>
<th>Assessment Characteristic</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The object of the assessment.</td>
<td>Level at which the data is collected (i.e., individual student(s) or groups of students).</td>
</tr>
<tr>
<td>2. The object of the decision.</td>
<td>Level at which the decision is aimed, which should be equal to the level at which data are aggregated.</td>
</tr>
<tr>
<td>3. The time point in which decisions need to be made.</td>
<td>The timing of assessment, decision and actions that follow the decision.</td>
</tr>
<tr>
<td>a. Timing of the decision</td>
<td>- Who should benefit from the decision?</td>
</tr>
<tr>
<td>b. Frequency of the decision</td>
<td>- What behavioral and cognitive aspects are most important?</td>
</tr>
<tr>
<td>c. Timing of the feedback</td>
<td></td>
</tr>
<tr>
<td>4. The objective of the assessment.</td>
<td></td>
</tr>
<tr>
<td>5. The assessment method.</td>
<td>- Degree of standardization.</td>
</tr>
<tr>
<td>6. Characteristics of the assessment process.</td>
<td>- The resulting data (qualitative vs. quantitative).</td>
</tr>
<tr>
<td>7. The power differentials between agents.</td>
<td>- Cyclic vs. non-cyclic.</td>
</tr>
<tr>
<td>a. Who makes the decision.</td>
<td>- Systematic vs. non-systematic.</td>
</tr>
<tr>
<td>b. Who has to take actions (e.g., providing feedback, deciding what has to be learned).</td>
<td>Discrepancies between the agents who make the decision and the agents who are expected to follow-up on the decisions by taking action. The higher the differential, the higher the stakes.</td>
</tr>
<tr>
<td>8. The types of intervention.</td>
<td>The type of intervention that follows assessment. Formative interventions can be:</td>
</tr>
<tr>
<td>a. proactive;</td>
<td></td>
</tr>
<tr>
<td>b. retroactive; or</td>
<td></td>
</tr>
<tr>
<td>c. interactive (Stobart, 2008, pp. 146-147).</td>
<td></td>
</tr>
<tr>
<td>9. (Learning) theory.</td>
<td>On which learning theories the assessment approach is based.</td>
</tr>
</tbody>
</table>
First, we investigated how the object of assessment (e.g., students, teachers, or schools) as well as the object of the decision making varies across assessment approaches. The reportage of the results should equal the level at which the decision is aimed because it is, for example, more difficult to base decisions about students on data reported at classroom level. Next, we compared the three approaches on the timing of the assessment, which is affected by the time in which the decision has to be made. With this third feature the frequency of assessment was investigated as well. The timing of feedback is included to compare the approaches on the timing of actions taken after the decision.

The fourth assessment characteristic is the objective of assessment, by which it was analyzed who should benefit from the decision, as well as which behavioral and cognitive aspects are essential to complete the assessment task. The behavioral and cognitive aspects that are the primary objective of the assessment affect the assessment method that is most likely to be used. We compared the assessment methods used in each approach on the degree of standardization and the type of data that are collected. Sixth, the assessment process consists of collecting data, interpreting data, making a decision, and taking actions. The assessment approaches differ in the extent to which the assessment process is cyclic and systematic. When the assessment process is very systematic and follows a strict cycle the procedure becomes more formal, whereas a non-systematic and non-cyclic assessment process will be perceived as informal and more integral to the learning process. This degree of formality, in combination with the degree of standardization of the assessment method, affects how students will perceive the stakes of the assessment.

Another assessment feature that affects the stakes of assessment is the power differentials between agents. Power differentials might exist between the object of assessment, the assessor, the decision maker, and the person who follows through with the decision by implementing the intervention. The eighth characteristic is the type of intervention that follows the decision. Because the aim of this comparison is to composite an integrative formative approach to assessment, we studied which formative interventions could follow each assessment approach. Assessment data can also be used for the sake of remediating learning difficulties, which is called retroactive formative assessment (Stobart, 2008). On the contrary, proactive formative assessment leads to the implementation of interventions that should prevent the development of learning difficulties by addressing commonly developed misconceptions. This type of formative assessment is also known as pre-emptive formative assessment (Carless, 2007).
With interactive formative assessment the intervention is the result of the interaction of the learner with the learning environment (e.g., learning materials, teachers, and students), meaning that the line between assessment and intervention is blurred (Stobart, 2008).

Finally, a parallel was drawn between the approaches regarding the learning theories that are used to define what is meant by students’ learning outcomes. Among these theories are behaviorism, (social) constructivism (which is based on the cultural historical theory of Vygotsky), cognitive developmental psychology of Piaget, and information processing theories (Verhofstadt-Denève et al., 2003).

**Results**

The results of the comparison between the assessment approaches are shown in Table 3. The numbers of the headings of the following sections correspond with the numbers of the characteristics of assessment in Table 3.

1. **The Object of the Assessment**

DDDM comprises all educational levels, which means that it considers both assessment and evaluation. AfL comprises classroom and individuals and DT concerns individuals. Since the three approaches comprise different levels at which data are gathered, this suggests they could complement each other.

2. **The Object of Decision Making**

The objects of the decisions differ across the three approaches. In DDDM, the decisions can concern all levels of education, whereas AfL only concerns decisions within the classroom. DT is merely aimed at decisions about individual students. The three approaches comprise different levels at which decisions are made, which suggests that these approaches could be present simultaneously. Figure 1 shows the objects of the decisions in the three approaches.

**Figure 1 Objects of the Decisions in the Three Approaches**
3. The Time Point in Which Decisions Need to be Made

a. Timing of the decision
In DDDM, the timing of the decision can vary from immediately to a couple of years after data collection. In AfL, the decisions are almost always made immediately during the learning process. In DT, the time between the measurement and decision depends on the needs of the learner(s), but it is desirable to keep this time limited.

b. Frequency of the decision
Depending on the stakes of the decision, in DDDM the frequency of the decisions varies from daily to once in a couple of years. High-stakes decisions require careful consideration, as well as reliable and objective data. These types of decisions are generally made less frequently compared to low-stakes decisions. In AfL, teachers continuously make decisions based on the assessment information at hand. Thus, decisions are made very frequently. For DT, the frequency of the decisions will depend on the needs of the learner(s).

c. Timing of the feedback
In DDDM, feedback is usually delivered with a delay; this is especially the case when feedback from a standardized measurement is delivered by an external party. In AfL, feedback is provided continuously according to the needs of the learners. When using DT, it is preferable to inform the learners about the diagnosis as soon as possible after the measurement.

4. The Objective of the Assessment
The objective of DDDM is assessing and/or evaluating whether or not learning goals are met, and thereby investigating whether changes to the learning environment are necessary. This approach is retroactive, because it aims to resolve problems after a period of teaching. On the other hand, the objective of AfL is improving the quality of the learning process and thereby establishing higher learner autonomy, and higher learning outcomes. This process is characterized as interactive and proactive, and sometimes also as retroactive (Stobart, 2008).

The objective of assessment within DT is measuring the student’s processing activities, and identifying their preconceptions, misconceptions, bugs, and problem solving strategies. The process of DT is mostly retroactive, but it can also be used proactively. Furthermore, when the instrument used for DT provides feedback during the test, the process becomes interactive.
5. The Assessment Methods (Instruments)

The assessment methods used in DDDM are primarily standardized tests that result in quantitatively reported data-feedback, which has to be interpreted by the user. Standardized, in this context, indicates that all students take the same test. Also, the design of the instrument is determined by the learning goals that are being assessed, meaning that emphasis is placed on what has to be learned.

Usually, highly reliable tests are used that can be scored automatically, so they are easy to administer on a large scale. Assessment methods used for AfL are usually non-standardized and therefore students' learning outcomes are described qualitatively in most situations. Because AfL focuses on the quality of the learning process, the form of the assessment is as important as the content of the assessment. For the assessment of deep learning, the use of various assessment methods is essential (Harlen & Gardner, 2010). Instruments used in DT can be either standardized or adaptive, with the latter referring to computerized tests in which the selection and sequencing of items depends on the responses given by the student. The testing procedure of the instrument and the method of scoring should be based on measurement theories designed for DT (e.g., Leighton & Gierl, 2007b; Rupp et al., 2010). Additionally, theories about students' cognitive processes are required to formulate hypotheses about the nature of students' learning outcomes. Because of the complexity of diagnostic measurement models, it is preferred that quantitative results of DT are reported in a readily usable format that contains an easy to understand qualitative description of the student's learning needs.

Furthermore, for each approach the requirements with regard to the quality of the data depend on the stakes of the decision: the need for objective and reliable data increases when the stakes become higher (Harlen, 2010). An example of a high-stakes decision is a pass/fail decision that is used to decide which students receive a diploma and which students will not. Even though it is desirable to use objective and reliable data for daily decisions, this is not always feasible because those data are often gathered with standardized (national) tests.

The time between the test moment and receiving reportage on the results of those tests often exceeds the time in which those decisions have to be made. Therefore, if the feedback loops are small, and the stakes are low, using various in-class assessment methods will usually suffice (Harlen & James, 1997). Additionally, standardized tests are likely to be more focused on specific learning goals than other assessment methods, and for that reason are unable to fully cover the diverse meanings of students’ learning outcomes within school curricula.
6. Characteristics of the Assessment Process

The processes of both DDDM and DT are cyclic, systematic, and formal. However, the stakes that are associated with DDDM are usually higher than those associated with DT. The process of AfL is non-cyclic and non-systematic, because the gathering of data is as a result of student-teacher interactions in daily practice rather than of a planned and separate measurement moment. Therefore, students will be likely to perceive the process of AfL as informal.

7. The Power Differentials between Agents

a. Who provides feedback?

For DDDM, the feedback can be provided by an external party that develops standardized tests. When this is the case, the decision most likely has a high-stakes character. Additionally, in DDDM, feedback might also be provided by agents within the school. In both situations, the teacher is responsible for using this feedback within the classroom and for feeding these results back to the students. For AfL, feedback is only provided by agents within the school, such as a teacher, a peer, or a computer. In DT, the feedback is provided by the computer, either directly to the student, or indirectly via the teacher.

b. Who determines what has to be learned?

The learning outcomes assessed in DDDM are the learning objectives that are mandated by external parties like the government. In AfL, national learning objectives serve as a broader framework of learning goals, but do not determine what has to be learned on a day-to-day basis. Instead, the students and their teacher(s) decide upon the daily learning intentions. Although DT is concerned with the learning needs of individual students, what has to be learned is determined by the agents outside the school who developed the diagnostic test. Specifically, the theories about students’ cognitive processes, common errors, and misconceptions determine what has to be learned in DT.

8. Types of Intervention

In the section about Characteristic 6, the process of DDDM is described as retroactive, meaning that the interventions that follow aim to remediate signaled learning difficulties. Feedback in DDDM has a quantitative nature and can be used to inform decisions on all educational levels. AfL is primarily interactive because interventions implemented in AfL, such as qualitative feedback, are part of the daily classroom discourse (Stobart, 2008).
Specifically, the interventions in AfL are the result of the interaction between the learner and the learning environment (e.g., teacher, peers, and materials). When the quality of the interactions between the learner and the learning environment is optimal, AfL is also characterized as proactive because in that case the development of learning difficulties is highly improbable. DT is primarily retroactive; interventions are mostly used for the remediation of learning problems. For example, this is the case when students are classified and/or selected. However, as described above, DT becomes interactive when the diagnostic tests provide feedback to the student during the measurement moment. Additionally, DT can result in proactive interventions when it is used at the beginning of a teaching period to, for example, measure students’ preconceptions so that the development of learning difficulties can be prevented.

9. Learning Theory
Initially, the improvement of learning outcomes was based on neo-behaviorist ideas, which encompass student learning in terms of learning sub-skills spread out over fixed periods of time within each school year (e.g., six weeks). After that period students’ abilities are assessed, usually with a paper-and-pencil test (Stobart, 2008). As a consequence of this cyclic process, the formative function of those assessments is purely retroactive. As is implied by previously described characteristics, DDDM is based on these theoretical principles. However, critics of the neo-behaviorist principles argued that students’ reasoning styles should be recognized as an indicator of their ability, which is known as the constructivist learning theory. In this theory, it is stated that learners actively construct their knowledge. This theory prescribes that assessments should not only focus on behavioral aspects of learning, but also on students’ reasoning (Verhofstadt-Denève, 2003).

The theory on which AfL is based takes constructivism a step further by stating that knowledge is constructed through interactions with others (Stobart, 2008). This is called social constructivism. As described in the section about DT, DT is concerned with measuring the student’s ZPD. The ZPD is an element within social constructivism; learning through interactions with others means learning with the minimal help of others. Thus, both AfL and DT are based on principles of social constructivism.
<table>
<thead>
<tr>
<th>Assessment Characteristic</th>
<th>Data-Driven Decision Making</th>
<th>Assessment for Learning</th>
<th>Diagnostic Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The object of the assessment.</td>
<td>Individuals, classrooms, schools</td>
<td>Individuals and classrooms</td>
<td>Individuals.</td>
</tr>
<tr>
<td>2. The object of decision making.</td>
<td>Individuals, subgroups within a classroom, whole classrooms, multiple classrooms, within a school, and across schools.</td>
<td>Individuals, subgroups within a classroom, and whole classrooms.</td>
<td>Individuals.</td>
</tr>
<tr>
<td>3. The time point in which decisions need to be made.</td>
<td>a. Depending on the level at which a decision has to be made this can vary from immediately to a couple of years. b. Varies from daily to once every couple of years, depending on the stakes of the decision. c. Data feedback at the school and classroom level, usually delayed.</td>
<td>a. During the learning process. b. When teaching, teachers continuously need to decide upon their next step. For these decisions they use assessment information that is available at that moment. c. Continuously provided according to the needs of the learners.</td>
<td>a. Depending on the needs of the learner(s). b. Depending on the needs of the learner(s). c. Preferably right after the diagnosis and advice on the intervention is given.</td>
</tr>
<tr>
<td>4. The objective of the assessment.</td>
<td>Determining and monitoring whether learning goals have been achieved, and investigating whether changes to the learning environment are necessary. - Standardized tests; - quantitative results; and - form follows content; larger focus on what has to be learned than how it is learned.</td>
<td>Improving the quality of the learning process and thereby establishing higher learner autonomy, and higher learning outcomes. - Non-standardized; - mainly qualitative results; and - the content and form of assessment that are chosen are directly related to how successful learning is operationalized in terms of subject knowledge and understanding, as well as in terms of self-regulation skills (Stobart, 2008).</td>
<td>Assessment of processing activities, and identifying preconceptions, misconceptions, bugs, and problem-solving strategies. - Standardized or adaptive tests; - quantitative results explained in qualitative student reports; and - the content and form are theory-driven, because based on the outcomes inferences have to be made about why a student has achieved those learning outcomes.</td>
</tr>
<tr>
<td>5. The assessment methods (instruments).</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 (continued) Results of the Comparison of the Three Assessment Approaches Described by Characteristics

<table>
<thead>
<tr>
<th>Assessment Characteristic</th>
<th>Data-Driven Decision Making</th>
<th>Assessment for Learning</th>
<th>Diagnostic Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Characteristics of the</td>
<td>Cyclic;</td>
<td>Non-cyclic (continuous);</td>
<td>Cyclic;</td>
</tr>
<tr>
<td>assessment process.</td>
<td>systematic; and</td>
<td>non-systematic; and</td>
<td>systematic; and</td>
</tr>
<tr>
<td></td>
<td>formal.</td>
<td>informal.</td>
<td>formal.</td>
</tr>
<tr>
<td>7. The power differentials</td>
<td>Depending on the stakes of</td>
<td>Peer, teacher, student(s),</td>
<td>Depending on the</td>
</tr>
<tr>
<td>between agents.</td>
<td>the test, feedback on</td>
<td>or computer.</td>
<td>assessment method</td>
</tr>
<tr>
<td>a. Who provides feedback.</td>
<td>student results is provided</td>
<td></td>
<td>used, feedback is</td>
</tr>
<tr>
<td></td>
<td>by external or internal</td>
<td></td>
<td>provided by the</td>
</tr>
<tr>
<td></td>
<td>parties. The teacher is</td>
<td></td>
<td>teacher or the</td>
</tr>
<tr>
<td></td>
<td>responsible for feeding</td>
<td></td>
<td>computer.</td>
</tr>
<tr>
<td></td>
<td>back these results into</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>the classroom.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Who determines what has to</td>
<td>External party (i.e., the</td>
<td></td>
<td>b. The teacher</td>
</tr>
<tr>
<td>be learned.</td>
<td>government).</td>
<td></td>
<td>and/or the student</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>decide upon the</td>
</tr>
<tr>
<td>8. Types of intervention.</td>
<td>Retroactive;</td>
<td>Interactive;</td>
<td>student's personal</td>
</tr>
<tr>
<td></td>
<td>can inform decisions at all</td>
<td>adaptations to teaching;</td>
<td>learning goals.</td>
</tr>
<tr>
<td></td>
<td>educational levels (the</td>
<td>adaptations to the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>level at which the decision</td>
<td>learning environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>is taken affects the type</td>
<td>on single classroom and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of intervention); and</td>
<td>on student level; and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>quantitative feedback.</td>
<td>qualitative feedback.</td>
<td></td>
</tr>
<tr>
<td>9. Learning theory.</td>
<td>Neo-behaviorism (Stobart, 2008)</td>
<td>Social constructivism (Stobart,</td>
<td>Social constructivism;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2008).</td>
<td>cognitive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>developmental</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>psychology; and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>information</td>
</tr>
</tbody>
</table>
|                               |                             |                        | processing theories.
Discussion

The goal of this study was to investigate the extent to which DDDM, AfL, and DT can be shaped into an integrative formative approach towards assessment. The three approaches to assessment claim to be beneficial with regard to student learning. However, the literature has pointed out that different meanings are ascribed to student learning within these approaches; on the one hand student learning is defined as achieving learning goals, and on the other hand it is referred to as the quality of the learning process (see James & Brown, 2005, for more possible meanings of learning outcomes). If the three approaches can be combined into an integrative formative approach towards assessment, this could maximize student learning, in terms of both the process and the outcomes of learning at all levels of education. The three approaches were compared on nine characteristics of assessment.

The results suggest that the approaches could complement each other with respect to the objects of assessments, the objects of the decisions, the time point in which decisions need to be made, the assessment methods, and characteristics of the assessment process. With respect to the objects of assessments and the objects of the decisions, DDDM comprises all educational levels, whereas AfL comprises the classroom level and individuals. DT only concerns individuals. If these approaches are to be integrated, this would mean that DDDM would serve as a more overarching approach concerning evaluation and monitoring. Furthermore, in our view, DDDM should be concerned with high-stakes decisions, whereas AfL should concern the daily practice where decisions are made on a continuous basis. DT should be used when needed to gather in-depth information about student learning. The assessment methods differ widely between the three approaches; however, we believe that they should be present simultaneously. Namely, when all three approaches are combined this will lead to a complete picture of both students’ learning processes and learning outcomes. Also, the characteristics of the assessment process differ, but can be complementary. Since DDDM is highly systematic and cyclic, it can be used to maintain and improve the quality of education. Useful tools for this purpose for example are student monitoring systems, in which students’ learning outcomes are assessed using standardized tests once or twice a year. When the results of these monitoring actions suggest that learning goals are not being met, it has to be decided how the learning environment could be changed to improve students’ learning outcomes.
On a day-to-day basis in the classroom, however, AfL can be a very powerful approach. Because of its flexible and responsive character, the learning needs of students can best be attended to. DT can be used in a flexible way, when there is a need for more in-depth information about student learning or particular learning needs.

The objectives of the assessments and the interventions in the three approaches were found to be somewhat contradictory. More specifically, DDDM has a retroactive character, whereas AfL and DT are more proactive, interactive, as well as retroactive. We believe that when the three approaches are implemented simultaneously, the need for retroactive measurers will decline, because learning difficulties will be resolved immediately. Therefore, we argue that when sufficient autonomy is granted to the students and teachers, the three approaches could actually support each other. The power differentials between agents appeared to differ between the three approaches as well. The results suggest that high power differentials between agents, as is the case in DDDM, will give students and teachers less opportunities to take responsibility for the quality of the learning process, a necessary condition for AfL. DDDM is more distant from teachers and students, because it mainly takes place outside the classroom, and therefore the power differentials are high. Thus, power differentials between agents in AfL, and to a lesser extent in DT, are much smaller than in DDDM because teachers decide on the learning activities that take place in their classroom. We believe that when the power differentials at the classroom level are low, this will offer the most optimal learning climate in which feedback can be used to its full potential. However, low power differentials in AfL require that teachers receive autonomy from school principles to design their practice. Moreover, the learning theory on which DDDM is based, known as neo-behaviorism, has in most educational reforms been replaced with the (social) constructivist view of learning (Stobart, 2008). Because our comparison shows that both AfL and DT are based on social constructivism and this appears to be the theory that currently dominates education, we believe it would be best if this theory was also used for DDDM. Social-constructivism can be used in DDDM because it does not exclude neo-behaviorist principles. Moreover, it supplements those principles by addressing the cognitive and social components of learning. Pedagogical-didactical principles in neo-behaviorism focus on conditioning, meaning students internalize how to solve tasks by repeatedly performing the same tasks (Verhofstadt-Denève, et al., 2003).
A DDDM approach that is more social-constructivist oriented includes learning activities that are active (not merely reproducing and repeating the same tasks) and include frequent interactions with peers and the teacher. Furthermore, concerning the ecological validity of assessment methods used within the three approaches it is important to use methods that are similar to the wide variety of learning tasks used in social-constructivist learning environments. However, assessment methods often used in DDDM are standardized tests with a fixed question format. Thus, a DDDM approach based on a social-constructivist view of learning acknowledges the complexity of learning by including various assessment methods.

Furthermore, DT complements DDDM and AfL because it offers teachers a non-labor-intensive way of systematically collecting detailed data on a student's learning needs. Moreover, DT has a lower risk of misinterpretation of data than DDDM, because the quantitative data are described in wording that is easy to understand and can be directly used in practice. This is in contrast to DDDM, where the data has to be interpreted. For AfL, there is a low risk of making unjustified decisions as a result of misinterpretations, because misinterpretations will be directly revealed through teacher-student interactions, and can be contingently restored.

Nevertheless, the three approaches have some characteristics that might limit their formative potentials. In DDDM, for example, monitoring students could easily transform into frequent administration of mini summative tests (Harlen & James. 1997; Stobart, 2008), which would unnecessarily raise the stakes for both students and teachers. Even more important, data-feedback from monitoring activities is most likely aggregated quantitative data that cannot easily be used to enhance the learning of individual students, because those data do provide sufficient detail about individual students. A possible pitfall of AfL is that teachers require extensive knowledge of the assessment domain to be able to, for example, ask questions that will promote students’ learning during classroom discourse (Bennett, 2011; Moyer & Milewicz, 2002). In the section about DT, it was explained that for the development of diagnostic tests, theories are required to make inferences about students’ reasoning during the tests. In AfL, teachers need to develop similar theories to make distinctions between students’ ‘slips’ and errors (Bennett, 2011). Besides formal teacher training programs, the use of DT might enhance teachers’ knowledge about students’ thinking with regard to a specific domain. Subsequently, the necessity of theories about how students construct knowledge and about their thinking within both AfL and DT becomes clear from our comparison of the theoretical underpinnings of both approaches.
Finally, a pitfall of DT is that it only advises on the intervention, selection, and classification of students and therefore a formative use is not guaranteed.

A limitation of the current study is that the three approaches were only compared on their theoretical principles as described in educational literature. However, the definitions of the approaches are not uniformly described in the literature. For the sake of comparing the three approaches, interpretation of the meaning of these approaches was unavoidable. Further research should investigate how this integrative formative approach could be implemented in practice. Currently, in schools in British Columbia, Canada DDDM and AfL are the focus of their accountability framework (Ministry of Education, British Columbia, Canada, 2002). They define accountability in terms of the achievements of each student, and take into account the diversity of students within different school districts by adapting expectations based on DDDM and AfL. In addition to our comparison of DDDM, AfL, and DT, it would be interesting to further study initiatives such as these.

References


