LITERATUURSTUDIE
Evaluatie & diversiteit

Joke Ysenbaert, Piet Van Avermaet & Mieke Van Houtte

LITERATUURSTUDIE
Evaluatie & diversiteit

Joke Ysenbaert, Piet Van Avermaet & Mieke Van Houtte

Promotor: Prof. Dr. Piet Van Avermaet

Copromotoren: Prof. Dr. Koen Jaspaert, Prof. Dr. Kris Van den Branden & Prof. Dr. Mieke Van Houtte

Het Steunpunt Onderwijsonderzoek is een samenwerkingsverband van UGent, KU Leuven, VUB, UA en ArteveldeHogeschool.

Voorwoord

Het hoeft geen betoog meer dat elke school in Vlaanderen vandaag in vele opzichten divers is. De gepercipieerde en feitelijk toegenomen diversiteit (M-decreet, vluchtelingcrisis, migratie, toenemende cijfers op vlak van kinderarmoede, …) stelt directies en leerkrachten voor uitdagingen. Ook zo op vlak van hun evaluatiebeleid en –praktijk. Deze literatuurstudie heeft als doel om een referentiekader aan te bieden van waaruit de huidige stand van zaken met betrekking tot evaluatiebeleid en –praktijk en hoe diversiteit hierbinnen een plaats krijgt, verklaard kan worden.

Het voorgestelde referentiekader vertrekt vanuit een holistisch perspectief waarbij 'evaluatie' niet als een losstaand gegeven bekeken wordt, maar in relatie gezien wordt met andere domeinen van de samenleving. Onder 'evaluatie' wordt begrepen het evaluatiebeleid van een school en de evaluatiepraktijken van leerkrachten, maar evenzeer de opvattingen en overtuigingen met betrekking tot evaluatie van relevante stakeholders. Dit alles vormt aldus het voorwerp van onderzoek. Het referentiekader geeft weer hoe 'evaluatie' ingebed is in zowel een onderwijskundige context alsook in een socio-culturele en politiek-economische context. Deze contexten zijn nauw verstrengeld met elkaar en beïnvloeden elkaar voortdurend en oefenen aldus invloed uit op het evaluatiebeleid van een school, de evaluatiepraktijken van leerkrachten en de opvattingen met betrekking tot evaluatie. Daarenboven bepalen deze contexten ook mee in hoe deze evaluatiepraktijken zich verhouden tot de toenemende diversiteit.

Inhoud

Beleidssamenvatting

Het hoeft geen betoog meer dat elke school in Vlaanderen vandaag in vele opzichten divers is. De gepercipieerde en feitelijk toegenomen diversiteit (M-decreet, vluchtelingcrisis, migratie, toenemende cijfers op vlak van kinderarmoede, …) stelt directies en leerkrachten voor uitdagingen. Ook zo op vlak van hun evaluatiebeleid en –praktijk. Deze literatuurstudie heeft als doel om een referentiekader aan te bieden van waaruit de huidige stand van zaken met betrekking tot evaluatiebeleid en –praktijk en de mate waarin diversiteit hierin een plaats krijgen, verklaard kan worden.

Het voorgestelde referentiekader vertrekt vanuit een holistisch perspectief waarbij 'assessment' niet als een losstaand gegeven bekeken wordt, maar in relatie gezien wordt met andere domeinen van de samenleving. Het onderstaand model geeft weer hoe 'assessment' ingebed is in zowel een onderwijskundige context alsook in een socio-culturele en politiek-economische context die elkaar voortdurend (on)rechtstreeks beïnvloeden. Onder 'assessment' wordt begrepen het evaluatiebeleid van een school en de evaluatiepraktijken, maar ook de opvattingen en overtuigingen met betrekking tot evaluatie van relevante stakeholders. Deze 'binnenste cirkel' vormt aldus het voorwerp van onderzoek. De buitenste cirkels kunnen in rekening genomen worden om deze binnenste cirkel te verklaren aangezien deze hier invloed op uitoefenen.

Alvorens dieper in te gaan op de verschillende cirkels van het model, biedt de literatuurstudie een overzicht van de verschillende termen die vaak in verband gebracht worden met evaluatie: 'testen', 'assessment', 'formatieve evaluatie', 'summatieve evaluatie', *'assessment of learning', 'assessment for learning'* en *'assessment as learning'.* De terminologie met betrekking tot evaluatie is de laatste jaren sterk uitgebreid, wat volgens de literatuur geleid heeft tot begripsverwarring en initiatieven om duidelijkheid te scheppen wat resulteerde in een complex probleem van begripsafbakening en definiëring.
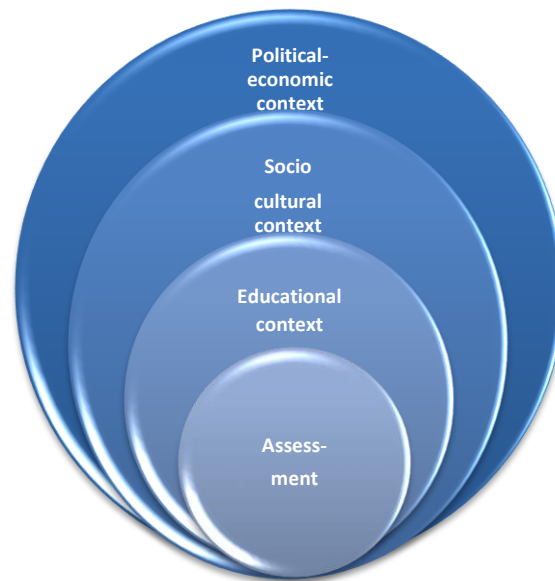
De cirkel die in bovenstaand model het dichtst bij 'assessment' staat, is de onderwijskundige context, die vandaag de dag gekenmerkt wordt door een toenemende feitelijke en gepercipieerde diversiteit. De feitelijke diversiteit heeft betrekking op verschillende facetten: etnische diversiteit, meertaligheid, verschillen in schoolloopbanen,… Overheidsmaatregelen, zoals het GOK-decreet en het M-decreet, die aansturen op een inclusiever onderwijssysteem zorgen ervoor dat de gepercipieerde diversiteit ook toeneemt.

De manier waarop er omgegaan wordt met toenemende diversiteit kan verklaard worden vanuit de socio-culturele context. Theorieën met betrekking tot diversiteit geven richting aan hoe met diversiteit omgegaan wordt, en hoe dit kan doordringen tot op het niveau van klaspraktijken en evaluatiepraktijken. Een groepsgerichte benadering van diversiteit leidt al snel naar een categorisering waarbij mensen *'gelabeld'* worden en waarbij kansengroepen zich moeten aanpassen aan de heersende normen van kansrijke groepen (Van Avermaet & Sierens, 2010). De individuele benadering van diversiteit, legt daarentegen de nadruk op het unieke van éénieder en erkent de diversiteit *binnen* groepen (Van Avermaet & Sierens, 2010). Daarnaast wordt de socio-culturele context ook mee bepaald door visies op leren en onderwijs en bijgevolg ook de visies op evaluatie. Afhankelijk van de mate waarin een leerkracht aansluiting vindt bij het gedachtegoed van het behaviorisme, het cognitivisme en/of het constructivisme zullen ook zijn/haar opvattingen omtrent de rol en de plaats van evaluatie in het onderwijsleerproces hierdoor beïnvloed worden. Wat ook een invloed kan uitoefenen op de uiteindelijke evaluatiepraktijken die in de klas gebruikt worden.

Tot slot speelt ook de politiek-economische context een rol in hoe evaluatie gezien wordt door de maatschappij, welke rol hieraan gegeven wordt en hoeveel belang eraan gehecht wordt. Internationale toetsen zoals PISA (Program for International Student Assessment) en TIMMS (Trends in International Mathematics and Science Study) zijn voor veel overheden een manier om af te toetsen hoe goed het nationale onderwijssysteem het doet. De resultaten op deze toetsen worden dan ook vaak aangewend om een onderwijshervorming door te voeren of om beleidsmaatregelen ingang te laten vinden (Ercikan, Roth, & Asil, 2015; Standaert, 2008, 2014). Zo werden in Duitsland nationale gestandaardiseerde toetsen ingevoerd naar aanleiding van slechte resultaten bij PISA (Teltemann & Klieme, 2016). Hierdoor wordt duidelijk dat evaluatie-instrumenten gehanteerd kunnen worden om een politieke agenda kracht bij te zetten. Een voorbeeld hiervan is de 'No Child Left Behind Act' dat als doel had om in de Verenigde Staten het aantal risicoleerlingen te laten dalen en zo meer gelijke onderwijskansen te creëren. Deze beleidsmaatregel ging gepaard met de invoering van nationale gestandaardiseerde toetsen en een sterk verantwoordingsgerichte visie op evaluatie. Scholen die slechte cijfers behaalden op de centrale toetsen, liepen namelijk het risico om hun subsidiëring te verliezen. In de literatuur worden de negatieve effecten van deze toetsen uitgebreid omschreven: teaching-to-the-test, curriculumverenging en ook het creëren van ongelijke onderwijskansen omdat bepaalde (groepen) van leerlingen uitgesloten kunnen worden (Menken, 2008; Standaert, 2008, 2014).

Typerend voor het Vlaamse onderwijs is dat er geen traditie is op vlak van gestandaardiseerd toetsen. Uit recent onderzoek blijkt bovendien dat er geen draagvlak is voor de invoering van gestandaardiseerde, nationale examens (Vanhoof, De Maeyer, Van Petegem, Penninckx, & Quintelier, 2016). Daarnaast wordt het Vlaamse onderwijs gekenmerkt door 'vrijheid' van onderwijs' wat grondwettelijk verankerd is. Dit houdt in dat de overheid de bevoegdheid heeft om te bepalen wat de doelen zijn die het onderwijs dient na te streven, wat vertaald wordt in eindtermen en ontwikkelingsdoelen. De invulling ervan, namelijk 'hoe' dit nagestreefd wordt, behoort tot de autonomie van scholen en laren. Het evaluatiebeleid en de –praktijken behoren tot het 'hoe', maar evalueren net het 'wat'. Gezien de onderwijsvrijheid en de lange afwezigheid van centraal georganiseerde, gestandaardiseerde toetsen, zouden we kunnen verwachten dat er in de praktijk een brede variatie is op vlak van evaluatiebeleid en –praktijken. De jaarlijkse rapporten van de onderwijsinspectie laten zien dat 'evaluatie' eerder algemeen als een zwak punt beschouwd wordt in de werking van scholen aangezien slechts 50 procent van de scholen een doelgericht evaluatiebeleid heeft en slechts 26 procent van de scholen over een doeltreffend evaluatiebeleid beschikt (Onderwijsinspectie, 2015).

Introduction

Little is known about the assessment practices of teachers and the assessment policies of schools in Flemish education. Additionally, Flemish education is characterized by increasing factual and perceived diversity. A more diverse public of pupils in schools challenges teachers' educational practices and consequently also their assessment practices. Even more, a twofold trend is occurring: on the one hand, schools and teachers are looking for validated assessment instruments to inform parents and pupils and to give a reasoned advice on moments of transition, for example from primary to secondary education. On the other hand the educational field pays more attention to the quality of assessment: assessment needs to be competence-based, in line with the attainment targets and developmental objectives, taking into account a more diverse public of pupils, responsive to alternative forms of assessment…. To research the current assessment practices and according policies, a framework from which these practices and policies can be explained is needed. The following literature review has the aim to provide a framework that approaches assessment from a holistic point of view which is represented in the following model.
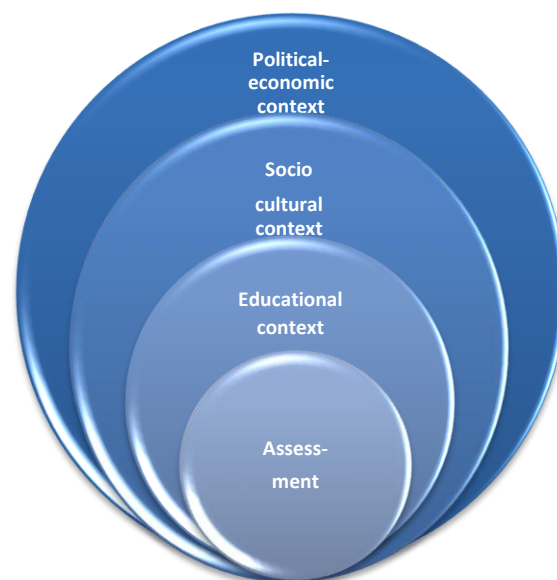
'Assessment' is situated central in the model and refers to beliefs about assessment from different stakeholders, the assessment practices of teachers and the assessment policies of schools'. Corresponding to the model, assessment is embedded in an educational, socio-cultural and political-

economic context. Assessment beliefs, practices and policies can be explained from these interacting contexts which influence each other continuously. This literature study will focus on each of this contexts. Before explaining each circle of the model an overview will be given of terminology used in describing assessment and how this has changed over the years.

The subsequent chapters will explain more profoundly the contexts in which assessment is embedded: an educational context that becomes more and more diverse (chapter 2); a socio-cultural context, shaped by views on diversity and learning (chapter 3); and the political-economic context in which international assessments play a role and have sometimes the power to foster educational reform (chapter 4). The debate about standardized versus non-standardized testing will be mentioned in the chapters about the socio-cultural and the political-economic context as the debate manifests at the level of these contexts.

The fifth chapter will finally focus on the specific context of Flanders. A short description of how Flemish education is organized will be given, as well as an overview of what is already known about assessment practices and policies in Flemish compulsory education. To finish, the model of the interacting circles will be applied to the Flemish context in order to predict what we can expect from the current assessment policies and practices.

1. Evaluation, measurement, assessment, formative, summative, ... See the big picture despite the flood of terms

During the course of the 20th century, most of the research on assessment in education was concentrated on the role of standardized testing. This focus led to the development of a psychometric/measurement paradigm which became very prominent in the educational context in the 20th century (Mc Millan, 2013) and which is still influential for assessment practices in the 21th century (Earl, 2003; Shepard, 2000; Richard J Stiggins, 2002). In the spirit of this paradigm, all attention was focused on the psychometric principles of large-scale testing and the corresponding technical and statistical topics. Stiggins (2002) stated that the measurement community had been focusing too long on the development of ever more sophisticated ways of creating valid and reliable tests. The question of how to assure that assessments serve the learning process and motivation of pupils was ignored, resulting in an assessment crisis (Stiggins, 2002) that in turn led to an evolution in thinking about assessment and its role. This has resulted in a paradigm shift from a testing culture towards an assessment culture.

1.1. Testing culture vs. assessment culture

The changing ideas about assessment are often described as a paradigm shift from 'testing culture' to 'assessment culture' (Birenbaum, 2014, 2016; Dierick & Dochy, 2001; Falchikov, 2005). Often the two views are presented as opposites of one another.

The purpose and the function of assessment are seen differently. Within a testing culture, assessments are all about making the grade and as a result are used for accountability reasons, whereas in an assessment culture, assessment is considered as a means for learning, which drives the teaching and learning process (Birenbaum, 2014, 2016). On the one hand, assessments have to be objective and standardized according to the testing culture. On the other hand, the assessment culture puts emphasis on the possibilities for dialogue and interaction between the assessor and assessed pupil. As a result, both teacher and learner take responsibility for the learning process. The relationship between the two actors differs significantly from that within a testing culture where the assessor is seen as 'the one who knows best'. Both point of views show differences in their expectations about learning. In testing culture, where assessment is seen as accurate quantification, it is considered to be able to measure a pupil's ability. From this point of view, ability is fixed, whereas an assessment culture has trust in efficacy-beliefs: teachers believe in a pupil's innate ability to learn and in the development

of his/her own self-regulated-learning strategies (Birenbaum, 2014, 2016). The two cultures therefore differ in what they believe should be assessed. According to the testing culture, the focus is on the product and on reproduction of knowledge. According to the assessment culture, both product and process get attention. Not only knowledge is emphasized but also the competencies to transfer theoretical knowledge into practice and to different contexts (Dierick & Dochy, 2001; Falchikov, 2005). (Dierick & Dochy, 2001; Falchikov, 2005).

As described by Birenbaum (2016), the two cultures differ in their attitudes towards diversity. The testing culture does not recognize diversity: assessment is considered to be 'one-size-fits-all' and all the pupils undergo the same strict assessment procedure. Assessment culture, by contrast, acknowledges diversity (Birenbaum, 2016) and provides opportunities for all pupils to prove their progress in the learning process. Assessment culture thus maintains a modest attitude towards assessment, providing unilateral information and often representing a snapshot of pupils' capabilities. For this reason, assessment culture emphasizes the importance of regular and informal assessments in order to provide multiple moments for pupils to show their capabilities, even at moments when they do not realize that the teacher is 'measuring' their achievement.

The following table summarizes the differences between testing and assessment cultures based on the work of Birenbaum (2014, 2016); Dierick and Dochy (2001); and Falchikov (2005).

Testing Culture	Assessment Culture
About making the grade	About learning
Accountability.	Drives teaching and learning
Standardized	Dialogue and interaction
Assessor knows best	Values students participation
'one size fits all'	Acknowledges diversity
Ability is fixed	Efficacy-beliefs
Assessment = accurate quantification.	Modesty in assessment
Product	Product & process

Table 1: Differences between 'testing culture' and 'assessment culture'.

The paradigm shift from 'testing culture' towards 'assessment culture' is characterized by the emergence of many new terms that primarily distinguish between types of assessment, denoting whether the type leans more toward the ideas of testing culture or more toward the ideas of assessment culture. Each emergence of new terms was intended to more soundly define the role and function of assessment, indeed, to create more clarity in the issue. Literature reveals that this was a

complex exercise which sometimes led to oversimplification that then called for nuancing and refining. It seems that this has led to a definitial issue, which we take as our point of departure for the discussion in the following paragraphs.

The rising interest for how to maximize the positive impact of testing on the learning process led to a new focal point in both research and practice: classroom-based assessment (Mc Millan, 2013). A shift took place from a paradigm which was solely focused on the psychometric aspects of measurement towards a paradigm that placed assessment more in the context of teaching and learning, and this shift went hand in hand with an increasing vocabulary for talking about assessment in education. Since the diversity in terminology is still apparent in the current educational context, some frequently recurring terms will be explained more extensively.

In her review study of 2004, Brookhart places 'classroom assessment' at the intersection of three classroom activities: instruction, classroom management and assessment (Brookhart, 2004). As a result, theories of classroom-based assessment are developed in three areas of study: psychology, sociology and measurement, which could explain the proliferation [and variety] of terms. Brookhart (2004) distinguishes 'assessment' from 'evaluation' by seeing the former simply as the collection of information about a pupil's progress. From the moment that the collected information is used to make judgments of the value of something, it is called 'evaluation'. Brookhart notices that the word '*value'* is included in 'evaluation'. When the information, collected through 'assessment', is used to determine whether a pupil passes or fails a subject or a grade, then it is called 'evaluation'. Where the purpose of 'evaluation' is to make a value judgment, the purpose of 'assessment' is to provide feedback to students and to make instructional decisions. The gathering of information to pursue these purposes can be divided into two types of information: information obtained by measurement and/or by assessment. 'Measurement' in this respect refers strictly to applying a set of rules that results in quantitative information. The result of 'measurement' is always expressed through a numerical scale: a number or a score. 'Assessment' refers to the gathering of information about a student's progress of learning, including both quantitative and qualitative information. Strictly speaking, the term 'assessment' takes 'measurement' into account, with the difference that the concept of 'assessment' also includes qualitative information (Brookhart, 2004).

1.2. 'Formative' and 'summative' assessment

During the past few decades the terminology used to refer to where a pupil is situated in his/her learning process has been multiplying rapidly. Terms like measurement, assessment, evaluation,

formative and summative assessment and assessment *for* learning, assessment *of* learning and even assessment *as* learning have shown up in literature. The increase in terms and definitions can be explained in two ways: on the one hand, it is a sign that much is moving in the field of assessment in education. On the other hand, not only for the field of research on assessment in education but also for the practice of classroom assessment, things become more complex as the expansive growth of terms makes it impossible to 'see the forest for the trees.'

In educational literature a traditional distinction is made between formative assessment and summative assessment, which was first described by Bloom in 1969. For Bloom the purpose of formative assessment was *'to provide feedback and correctives at each stage in the teaching-learning process'*. Summative assessment, on the other hand, was used to determine what the learner had achieved at the end of a course or program (Bennett, 2011). The distinction described by Bloom has been elaborated and still holds today. Summative assessment focuses on assessing learning outcomes and occurs at the end of a learning phase (Frey & Schmitt, 2007; Stobart, 2008). Brookhart (2004) refers to summative assessment as useful for final decisions, for example for the assignment of end-of-term grades. Formative assessment is described as the kind of assessment that occurs while learning is happening (Frey & Schmitt, 2007). The purpose of formative assessment is to inform the teacher about the learning processes and to gain insights that can be used to support learning through customized instruction and feedback (Frey & Schmitt, 2007; Stobart, 2008).

'Formative assessment' has received much more attention in educational literature and research than 'summative assessment' because of its power to foster student learning (Black, 2013). The most widely cited authors are Paul Black and Dylan Wiliam. In 1998, they published a pamphlet titled '*Inside the black box: Raising standards through classroom assessment',* which includes a summary of their review study '*Assessment and classroom learning'* (Black & Wiliam, 1998a, 1998b). Both authors were also influential members of the Assessment Reform Group (ARG) in the UK, a voluntary group of researchers brought together as the Policy Task Group on Assessment by the British Educational Research Association (BERA) from 1989 till 2010 (Bennett, 2011; Mc Millan, 2013). The aim of the ARG was to ensure that assessment policy and practice at all levels take account of relevant research evidence (Nuffield Foundation, 2017). Black and Wiliam (1998a) found decisive evidence that 'formative assessment' could improve learning, and published a booklet for practitioners called Inside the Black Box (Nuffield Foundation, 2017).

The focus on 'formative assessment' in both research and practice has led to a variety of interpretations, making it an umbrella concept in which opposing views were developed. In his critical review about 'formative assessment' Bennett (2011) describes the disunity in visions concerning

formative assessment. Among test developers – who are influenced by the ideas of the measurement paradigm – formative assessment is rather seen as an instrument, as a diagnostic test or an item bank from which teachers might create those tests. According to this point of view, tests will be conducted frequently and follow more or less the instructional units of an educational program (Bennett, 2011). This vision is in contrast with the view of formative assessment as a process, a view which is common among educators and researchers (Bennett, 2011). The emphasis on 'formative assessment' as a process puts a focus on understanding the students' learning process. According to this view of formative assessment, the results of assessment are used to adapt the teaching to fulfill students' needs (Bennett, 2011).

According to Bennett's critical review on formative assessment, the definitional issue – whether it has to be seen as an instrument or as a process – is irrelevant:

> "*It is an oversimplification to define formative assessment as an instrument because even the most carefully constructed, scientifically supported instrument is unlikely to be effective instructionally if the process surrounding its use is flawed. Similarly, it is an oversimplification to define formative assessment as a process since even the most carefully constructed process is unlikely to work if the 'instrumentation', or methodology, being used in that process is not well-suited for the intended purpose (Bennett, 2011).*"

Although the plea for a strong conceptualization in which both the process-position as well as the instrument-position are thoughtfully highlighted, the matter of the multiple definitions continues and becomes even more complex when new terms are introduced in the debate. Believers of the process view started putting forward the concept 'assessment for learning' as an alternative term for formative assessment (Bennett, 2011). The use of this concept had been encouraged by the findings in effectiveness research which uncovered the variety of interpretations of 'formative assessment'. These variations had been associated with other fields of research, for example motivation, self-regulated learning, feedback… As a result, studies on one aspect of the use of assessment to improve instruction were used as evidence supporting the efficacy of quite unrelated aspects. This evolution contributed to the creation of a more confusing concept of 'formative assessment', which convinced authors to abandon it, opting for the new concept 'assessment for learning' (Wiliam, 2011).

1.3. 'Assessment *of & for* learning'

There is evidence that assessment for learning emphasizes the process-view of 'formative assessment'. Chappuis and Stiggins (2002), for example, describe 'assessment for learning' as taking place "...*during the teaching and learning process rather than after it and has as its primary focus the ongoing improvement of learning for all students".* Assessment for learning is further conceptualized as a powerful tool because of its potential to increase learning by creating a learner-centered environment. 'Assessment for learning' occurs in classrooms when teachers use day-to-day classroom assessment activities to involve students directly and deeply in their own learning process (Chappuis & Stiggins, 2002; Earl, 2003; Shepard, 2000). Teachers use the insights about students' learning process to design the next steps in instruction. These insights are gathered by observations, worksheets and questioning by the teacher in the classroom (Earl, 2003). A crucial aspect in the idea of 'assessment for learning' is the role of feedback, which has to be incorporated in the process to guide future learning (Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). The philosophy behind 'assessment for learning' is that assessment and teaching are integrated. The power of such an assessment does not come from sophisticated technology, nor from using a specific assessment instrument in either a formative or summative way. Its power lies in the recognition of how much learning is taking place in the common tasks of the school day and in how much insight into student learning teachers can extract from these day-to-day practices (McNamee & Chen, 2005).

In this view, 'assessment for learning' is opposed to 'assessment of learning' in which the latter is about grading and reporting, and where the former is intended to make assessment a part of teaching in order to support learning (Earl, 2003). The role of 'assessment of learning' in literature is described to be relatively small as it only has a role to play when there are decisions to be made that require judgments or when teachers and students want to see the cumulative effect of their work. As a consequence, the role of 'assessment for learning' has gained much more attention and praise in literature, a tendency which is defined as problematic in Bennett's critical review (Bennett, 2011). According to his assessment, the use of the concept of 'assessment of learning' as opposite to 'assessment for learning' nullifies the valuable role of summative assessment. This position receives support from authors who show concern regarding the increaseduring the last two decades in studies of formative assessment without a similar development with respect to summative assessment (Black, 2013). The introduction of the concepts 'assessment of/for learning' can be regarded as a step backwards in the 'definition problem', as it tends to remove any responsibility from the summative function in assessment of its in supporting the learning process. Bennett posits that the definition problem and the corresponding changes in terminology disregards the complex relationship between

the summative and formative function of assessment. He refers to the complexity by asserting that summative assessment has a primary function, namely assessment of learning, and a secondary function, assessment for learning. Formative assessment, at the same time, has a primary function as assessment for learning and a secondary function in assessment of learning (Bennett, 2011). The complexity described by Bennett (2011) has received support by other authors(Bennett, 2011; Black, 2013; Brookhart, 2004; Rea-Dickins, 2006; Shepard, 2000) who have emphasized the necessity of the amalgamation of both assessment functions in classroom practices. Both functions are a natural part of the learning process. Therefore both formative and summative assessment should be used together to redirect teaching and learning, instead of using just one or the other. The neat distinction between formative and summative assessment had already been nullified by Brookhart, since she noticed the blurred boundary between the two functions in the practice of classroom assessment (Brookhart, 2004). She maintained that even the use of the concepts 'formative' and 'summative' to refer to types of assessment is problematic insofar as it isolates two practices that are integral to each other: In practice, we see that an assessment can be used in a formative way, but also in a summative way, whereby the lines between the two dissolve. Hence, the character of an assessment instrument is not determined by the instrument itself, but is determined by the way the instrument, and more specifically the results of the instrument, are used by the teacher. In consequence, an assessment instrument cannot simply be labelled as either summative or formative in the practice of classroom assessment, as the character of the instrument is determined by how it is used (Brookhart, 2004). The information from an assessment can be used both in a formative way and also in a summative way.

As outlined above, the attempts to distinguish and define the different types of assessment have led not infrequently to an oversimplification of the notion 'assessment'. Research has often emphasized one function of assessment to the detriment of other functions, which is problematic because it led to conflicting beliefs and competing narratives about assessment (Bonner, 2016). The definition problem is characterized by a 'pendulum movement' in the literature, whereby the balance between these various functions of assessment sometimes got lost. The attention for only one function of assessment and the denial of the other functions conflicts with the natural character of assessment, namely its embeddedness in both learning and instruction (Earl, 2003; Heitink et al., 2016).

1.4. Towards a concept of assessment that recognizes the true nature of 'learning'?

Assessment was for a long time considered as a separate activity and was not seen as an integrated part of the learning process. The natural character of assessment, which is determined by the integration of teaching, learning and assessment as a whole (De Backer, Van Avermaet, & Slembrouck, 2016; Earl, 2003; Heitink et al., 2016) has partly been influenced by the way 'assessment for learning' called attention to the aspect of teaching. Different authors have described the relatedness with teaching by outlining assessment as an instructional tool or as an integration of both teaching and assessment (Heitink et al., 2016). From this point of view, the teacher takes a central role, it being his/her task to use assessment to monitor pupils' performance against objectives, to inform the next teaching steps, to give feedback for improvement, to learn something about pupils' learning (Hargreaves, 2005). The role of the teacher thus overshadowsthe role of the pupil. Some authors did pay attention to the role of the pupil and have recognized the interrelatedness of 'assessment' and 'learning' (Hargreaves, 2005). Chappuis and Stiggins (2002), for example, stated that assessment can also be seen as 'inquiry'. From this perspective, the purpose of assessment is to get a deeper understanding of pupils as learners instead of performers: 'assessment' here is thus all about the pupil taking control of his/her own learning and as a result, turns assessment into a learning event (Hargreaves, 2005). This model puts much more emphasis on the learning aspect and on the engagement of the one who is being assessed, as the assessment is an event to stimulate pupils to reflect, review, discover, and learn about their way of learning (Hargreaves, 2005). Earl (2003) talks about 'assessment *as* learning' to emphasize the neat distinction with 'assessment *for* learning', since assessment has the potential to stimulate the process of developing and supporting pupils metacognition. An important role is assigned to the pupil, not limited to a *contributor* to the assessment and learning process, but elevated to *critical connector* between the assessment and learning process. Assessment *as* learning is considered as a personal affair of the pupil who acts like an active critical thinker and makes sense of information to construct new learning.

Assessment cannot be separated from 'teaching', nor from 'learning'. Assessment is also related to the educational context in which teachers act and where 'assessment practices' take place. The following section will describe what the current educational context looks like.

2. Classroom assessment

The measurement community has been focusing for long time on large-scale assessment, mostly for high-stakes accountability testing. These large-scale tests influence what happens in the classroom, for example by the standards that are emphasized by these tests (Mc Millan, 2013). Furthermore, an increasing trend to devolve responsibility for assessment to classroom teachers, together with a growing awareness of the impact of assessment on learning (Black & Wiliam, 1998a) has put the focus on classroom assessment. Advances in learning and motivation theory have contributed to the increasing focus on classroom assessment (Mc Millan, 2013). As it is the nature of classroom assessments to affect student motivation and learning, these classroom assessment measures (whether summative or formative) impact what and how students study and what they learn. How teachers conceptualize assessments that they use in the classroom and how they are integrated (or not) with instruction also have a direct influence on student engagement and learning. This point of view has become the focus of a field of research that has been developing over the past two decades: the field of classroom-based assessment (Mc Millan, 2013). This field of research has been influenced by changes in several areas: advances in measurement, in technology, in high-stakes testing and in standards-based education. At the same time, the field of learning and motivation theory evolved together with advances in formative assessment (Mc Millan, 2013). Since the past two decades, these advancements and developments have created the context for a field of research that focusses on classroom assessment research.

The definition of classroom-based-assessment put forward by Hill and McNamara (2011) includes both the formative (or assessment for/as learning) and the summative (assessment of learning) function of assessment:

> *"Any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes."*

This definition is purposefully broad and takes into account the full spectrum of classroom-based assessment practices, including the type of assessment embedded in routine classroom activities: this definition encompasses every action from the teacher, the learner or peers, every interaction or artifact able to provide information about a learners' progress.

McNamara (2001) sets out three dimensions of classroom-based assessment – evidence, interpretation and use - which was subsequently elaborated by Hill and McNamara (2011).

Evidence	Data	*What is assessed?*	• Valued enterprises
	Approach	*How is evidence collected?*	• Planned (e.g. McNamara, 2001); Incidental (e.g. Torrance & Pryor, 1998) • Visible (e.g. McNamara, 2001); Embedded (e.g. Rea-Dickins, 2006)
	Target	*Who is assessed?*	• Individual (e.g. McNamara, 2001); Group/Class (e.g. Torrance & Pryor, 1998)
	Agent	*By whom?*	• Teacher/Learner (e.g. McNamara, 2001)
Interpretation	Reflection	*Level of attention*	• Sustained (e.g. McNamara, 2001); Fleeting (e.g. Rea-Dickins, 2006)
	Criteria	*Values guiding assessment*	• Explicit (e.g. McNamara, 2001); Unconscious (e.g. Wiliam, 2001) • External/indigenous (e.g. Leung & Teasdale, 1997)
Use	Purpose	*How is evidence used?*	• Assign level (*reporting*) • Teaching (*plan/modify*) • Learning • Management (*of behaviour or teaching*) • Socialization (*to culture of assessment*) (e.g. Tunstall & Gipps, 1996)
	Agent	*By whom?*	• Teacher/Learner (e.g. Black & Wiliam, 1998) • School (e.g. Rea-Dickins, 2001)

Fig. 1: Dimensions of classroom based assessment (Hill & McNamara 2011).

The following sections will explain the three dimensions of the framework. Each dimension will be extended based on the findings from review studies.

2.1. Evidence

The dimension of 'evidence' brings together the answers on the questions 'what is assessed?', 'how is evidence collected and by whom?' and 'who is assessed or about whom is data collected?'.

In classroom-based assessment, there are various approaches to the collection of data. This gathering of data is usually headed by the question 'what is to be assessed'? The answer is in any matter induced by valued enterprises (Hill & McNamara, 2011). Messick (1989) discussed the social dimension of validation research and situates all test constructs in the realm of the social arena and its values (McNamara, 2001). In school-based assessment, institutional needs determine what is to be assessed and what the procedures will be.

In contrast to the question 'what is to be assessed', 'approach' refers to the question 'How is evidence collected in classroom-based assessment?' In literature the answer gets situated on a continuum where one end refers to the kind of assessment that is planned in advance, takes place at a certain

moment and with a specific goal in mind (McNamara, 2001). The other end of the continuum refers to unplanned assessment activities or *incidental assessments* (Torrance & Pryor, 1998). This distinction corresponds to the continuum that distinguishes visible assessment activities (McNamara, 2001) from embedded assessment activities (Rea-Dickins, 2006). Based on the review of Black and Wiliam (1998b) which emphasizes the interrelatedness between assessment and learning, classroom assessment can be considered an essential part of teaching and learning. From this point of view, assessment activities are embedded in the teaching and learning process in such a way that it even becomes impossible to distinguish an assessment activity from a teaching activity, as teachers are engaged in an ongoing appraisal of their students. They provide feedback appropriate to learners' needs, feedback that is consequently often embedded within the classroom interaction (Rea-Dickins, 2006).
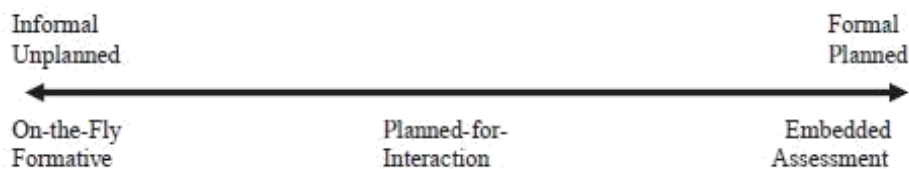
The continuum also represents a distinction between planned versus unplanned assessment. Planned or formal assessment is planned in advance to take place at a certain moment and with a specific goal in mind (McNamara, 2001). Yorke (2003) refers to this with the term 'formal formative assessment' and contrasts it with 'informal formative assessment,' which occurs during the day and does not take the form of tests but consists rather of answering questions or giving feedback. These 'informal formative assessments' are situated on the other end of the continuum: the unplanned assessment activities or *incidental assessments* (Torrance & Pryor, 1998). On this end, we find continuous, informal assessment activities or interactions between teachers and pupils. One example would be when the teacher goes around in the classroom while students are working in groups and he/she stays to give feedback on what he/she heard in the conversation of the students of one group (Shavelson et al., 2008). Ruiz-Primo (2011) describes informal assessments as conversations and dialogues between teachers and students which make the thinking of students explicit. These interactions can also be planned by the teacher when he/she is consciously looking to detect the gap between what students already know/can do and what they do not yet know or cannot yet do. This can be done by planned interactions, by formulating questions; in this way, the teacher gets insight into where students are in their learning process (Shavelson et al., 2008).

In the literature there is evidence for the positive effects of informal or on-the-fly formative assessment methods on the learning process of students. The positive effects of these methods are considered as having the potential to diminish the gap between what a learner already knows and can do, and what a learner does not yet know/cannot yet do. These informal assessment methods foster effective learning by increasing learners' motivation and performance (Cauley & McMillan, 2010). By way of illustration, effective questioning – also referred to as rich questioning – is widely described as an indispensable competence of teachers to gain insight into learners' thinking (Bennett, 2011; Black & Wiliam, 1998b; Clark, 2012; Hendrickson, 2012; Pryor & Crossouard, 2008; Ruiz-Primo, 2011; Taras, 2009; Wiliam, Lee, Harrison, & Black, 2004). By asking questions teachers detect pupils' misconceptions, which in turn enables them to encourage deeper learning. Rich questioning challenges pupils to think and a pitfall here is that teachers do not give enough time to let the pupils think and formulate an answer (Black et al. 2003). In their review study, Sluijsmans, Joosten-ten Brinke, and Van der Vleuten (2013) indicated some important issues in the context of effective questioning. First, it is important to challenge pupils with open questions because those type of questions stimulate pupils' thinking. Second, the questions have to focus on the thinking of the pupil to increase their involvement. Third, the questions must evoke/involve the different kinds of knowledge. A variation of questions (what, why & how) therefore supports deeper learning. Torrance and Pryor (2001) emphasize the importance of a supportive learning environment so that pupils do not feel uncomfortable or vulnerable when they are asked questions. Together with making observations, giving feedback and questioning, creating a supportive environment is crucial for clarifying the task and the criteria (Tang, 2010; Torrance & Pryor, 2001).

Another effective informal assessment method is the instructional dialogue between teacher and pupil. These dialogues are useful for clarifying learning goals to the pupils and to determine, together with the pupils, the criteria for success (Ruiz-Primo, 2011). Initially, 'instructional dialogue' referred to a didactic strategy where a dialogue was conducted after a formal test. Over the years, this concept was given a wider meaning and is now seen as a form of informal assessment (Ruiz-Primo, 2011). In literature instructional dialogues are often referred to with other terms: 'assessment conversations', 'class talk', 'group discussions', 'class discussions' and 'discours' (Black & Wiliam, 1998b; Black & Wiliam, 2009; Boxham & Campbell, 2010; C. Chin & L.-Y. Teou, 2010; Furtak et al., 2008; Hayward & Spencer, 2010; Kingston & Nash, 2011). Based on review studies, affective instructional dialogues show the following characteristics:

- – They are based on learning goals
- – The dialogue is interactive and the response of pupils is used as input for the dialogue

- These dialogues are part of a continuing process
- These dialogues consist of effective questioning
- They involve pupils in their own learning process and social participation is stimulated (Sluijsmans et al., 2013)

Some authors use the term 'reflective lessons' to label those lessons where the earlier mentioned effective informal assessment methods (effective questioning and assessment dialogues) are frequently used (Ayala et al., 2008; Furtak et al., 2008; Shavelson et al., 2008; Yin et al., 2008). During these lessons, teachers aim to incorporate activities that enable them to reflect on pupils' learning process in order to adjust the next teaching events to learners' needs. 'Reflective lessons' are in other words synonymous with lessons where pupils are stimulated to argue, to reason, and to develop their creative thinking. These lessons are characterized by interactive teaching, with plenty of group work and class discussion in which pupils are strongly involved and challenged to make their thinking explicit.

According to the model of Hill and McNamara (2011), 'target' refers to the question 'who is assessed?'. The focus can be on the individual (McNamara, 2001), but a group of pupils or the whole class can also be the target of an assessment (Torrance & Pryor, 1998). In the context of internal quality assurance, a year group can be the target in order to compare the results over the years to detect trends or to make decisions regarding school policy and didactic choices.

The agent refers to who collects the data, which is often the teacher, because he/she is usually the primary actor wanting to obtain information about the learner. In self-assessment, however, the learner him/herself can function as agent (McNamara, 2001). Several studies have revealed that self-assessment is an effective assessment-method to enhance learning (e. g. Allal & Lopez, 2005; Birenbaum, Kimron, & Shilton, 2011; Black & Wiliam, 1998b; Black & Wiliam, 2009; Brookhart, 2001; Brookhart, 2007; Cauley & Mc Millan, 2009; Clark, 2012; Hattie & Timperley, 2007; Nicol & Mcfarlane-Dick, 2006; Stiggins, 2005; Wiliam et al., 2004). Self-assessment refers to the evaluation carried out by the learners themselves, in particular an assessment of their performance or a learning task with an objective determined by a specific set of criteria. Self-assessment reinforces the development of self-regulation skills. Through evaluating their own work, students will reflect on their learning processes and outcomes, which is effective for deeper learning (Ayala et al., 2008; Brookhart, 2007). It increases their involvement by giving them both responsibility for as well as insight into their own learning process. Brookhart (2007) notes how self-assessment exercises can teach students the crucial skill of evaluating their own work. It is essential for teachers to involve pupils when setting criteria, to show

them how to apply these criteria, to give them feedback on their own self-assessment and to help them customize their next steps in their own learning process.

Besides the teacher and the learner himself/herself, the agent can also be a group of learners. When a learner's performance or learning task is assessed or evaluated by the classmates or peers, this is referred to as peer-assessment. The benefits of peer-assessment are similar to those of self-assessment insofar as it facilitates effective self-assessment and self-regulation (Allal & Lopez, 2005; Birenbaum et al., 2011), but the added value here is the cooperation that it teaches (Allal & Lopez, 2005; Birenbaum et al., 2011; Black & Wiliam, 1998b; Black & Wiliam, 2009; Sadler, 1989; Wiliam et al., 2004). As with self-assessment, students should be trained to do peer assessment (Brookhart, 2007).

2.2. Interpretation

The next step following the collection of data is the interpretation of the collected data. The model of Hill and McNamara (2011) distinguishes two subdimensions: reflection and criteria. Reflection refers to the level of teachers' attention when interpreting collected data about students' performance. According to McNamara (2001) teachers and learners can be involved in systematic reflection on the characteristics of a performance in order to redefine the learning goals. From this point of view, the level of attention is sustained, in contrast with 'fleeting reflection'. 'Fleeting reflection' emerges throughout lessons when potentially formative assessment events occur and unfold (Rea Dickins, 2006). This type of reflection brief and momentary and can occur intermittently throughout the classroom discourse.

The interpretation of assessment information is driven by criteria which can be explicit (McNamara, 2001) or, according to Wiliam (2001), unconscious. These criteria refer to values that guide the assessment.

2.3. Use

The dimension of 'evidence' and 'interpretation' is supplemented with the dimension of 'use', which refers to how the collected and interpreted data is used and by whom.

The subdimension 'purpose' refers to how the collected and interpreted data is used. This dimension distinguishes assessment as for 'teaching', 'learning', 'reporting', 'management' and 'socialization'. In practice it is not always possible to separate one purpose from another and sometimes two or more purposes can be derived from an assessment event (Hill & McNamara, 2011).

The purpose of reporting refers to the use of assessment-related information to inform decisions about students' end-of-the-year report and to assign levels (Hill & McNamara, 2011). McNamara (2001) sees the function of reporting as a managerialist demand in contexts where policy makers and system managers are forcing schools and teachers to report outcomes that demonstrate effort. From this point of view, reporting serves accountability demands instead of informing decisions about a student's schooling career.

Literature shows that assessment information can also be used for the socialization of learners into the conventions of teaching and assessment (Hill & McNamara, 2011; Torrance & Pryor, 1998). An example of this would be when learners are introduced to the concept of 'rubrics' in order to explain the criteria for success to them and in order for them to reflect on their own performances.

Assessment information can be used for managerial functions too, rather than pedagogic functions (Hill & McNamara, 2011), by reinforcing positive behavior, encouraging students and creating a positive atmosphere in the classroom.

Possibly the most obvious purposes of assessment, however, are for learning and for teaching. The purpose of teaching refers to the role of assessment information for teachers. Assessment results enable teachers to evaluate their own didactic functioning in order to modify their teaching activities towards the needs of the learners (Chin & L. Teou, 2010; Hill & McNamara, 2011; Sluijsmans et al., 2013). For example, the pace of teaching can be adapted. From the perspective of the learner, assessment information can be used by students to foster their learning. In her review study Brookhart (2007) describes how the attention for the perspective of the learner has increased throughout the years on the part of scholars, policy makers and teachers alike. Scriven (1967) describes formative assessment as the obtaining of information about the learning process. This definition is broadened by Bloom, Hastings and Madaus (1971), who specify that the information is useful for teachers to reflect about their instructional decisions. The literature of Sadler (1989) refers for the first time to the perspective of the learner by stating that the assessment information can also be used by learners to advance their learning. In more recent literature this learners' perspective has been expanded by adjusting the motivational effect of assessment (Black & Wiliam, 1998b; Brookhart, 2007; Crooks, 1987).

Time	Information about the learning process (Scriven, 1967)			
	Information about the learning process (Bloom et al., 1971)	Which can be used by teachers in order to make instruction-related decisions.		
	Information about the learning process (Sadler, 1989)	Which can be used by teachers in order to make instruction-related decisions..	And can be used by learners to foster their learning	
	Information about the learning process (Black & Wiliam, 1998b; Brookhart, 2007; Crooks, 1987)	Which can be used by teachers in order to make instruction-related decisions..	And can be used by learners to foster their learning	And motivates learners
	Definition			

Table 2: An elaboration of concepts with respect to 'formative assessment' over the years (Brookhart, 2007).

The perspectives of both the learner and the teacher come together in the concept of 'feedback'. Teachers give feedback based on the information gathered through assessment activities and learners can be stimulated and motivated by this feedback. The importance of feedback is widely described in the literature and is mentioned as one of the most effective contributing factors in the learning process (e.g. Allal & Lopez, 2005; Ayala et al., 2008; Bennett, 2011; Birenbaum et al., 2011; Black & Wiliam, 1998b; Black & Wiliam, 2009; Brookhart, 2001; Brookhart, 2007; Cauley & Mc Millan, 2009; Clark, 2010, 2012; Hattie & Timperley, 2007; Hayward & Spencer, 2010; Hill & McNamara, 2011; Kingston & Nash, 2011; Nicol & Mcfarlane-Dick, 2006; Pryor & Crossouard, 2008; Ruiz-Primo, 2011; Shavelson et al., 2008; Sluijsmans et al., 2013; Torrance & Pryor, 2001; Yorke, 2003). Despite all the references concerning the positive effects of feedback, Rea-Dickins (2006) notices the nuances of feedback: not all feedback is positive in terms of promoting learning; it can also have negative effects, such as taking away learners' motivation for learning.

Hattie and Timperley (2007) assign a central role to feedback in the context of formative assessment because of its potential positive effects on learning. Based on a meta-analysis, they developed a model for effective feedback which maximizes the positive effect. The model consists of three types and four

levels of feedback. The three types of feedback refer to the position of the learner in the learning process, which can be determined through three questions:

- Where is the learner in his/her learning? → Feed back
- Where is the learner going to? (The goals) → Feed up
- What needs to be done to get him/her there? → Feed forward

Each of these questions works at four levels. The first is task-level feedback, which focuses on misunderstandings in the interpretation of the task or on mistakes in the outcomes. The second level is about the main process needed to understand/perform a task. Feedback on the process-level should be related to the learners' own error-detection strategies, and give cues that lead to better strategies. The third focuses on the self-regulation level: the self-monitoring, directing, and regulating of actions. The fourth level concerns issues of personal evaluations and affect, including such feedback features as praise and judgment (Hattie & Timperley, 2007). According to their meta-analysis, the effect levels of feedback were high in cases where learners receive feedback on the task or the process. Lower-level effects were found for feedback on the self when only praise or punishment was given (Hattie & Timperley, 2007).

Within the dimension of 'use', a distinction is made between the 'purpose' and the 'agent', which refers to who uses the assessment information. As stated earlier, this can refer to the teacher or the learner (Black & Wiliam, 1998b), but also the school can be the agent (Rea-Dickins, 2006). The school can use the results of assessments in the context of internal quality assurance.

3. The role of the educational context in assessment practices

This chapter will zoom in on the second circle of the model: the educational context from which assessment (beliefs and practices of teachers concerning assessment and assessment policies of schools) can be determined and explained. This chapter will discuss the changing educational context which is characterized by increasing diversity. The increasing diversity in classrooms will challenge teachers and school leaders in their educational practices, and more specifically in how they create and use assessments and which role they ascribe to assessment.
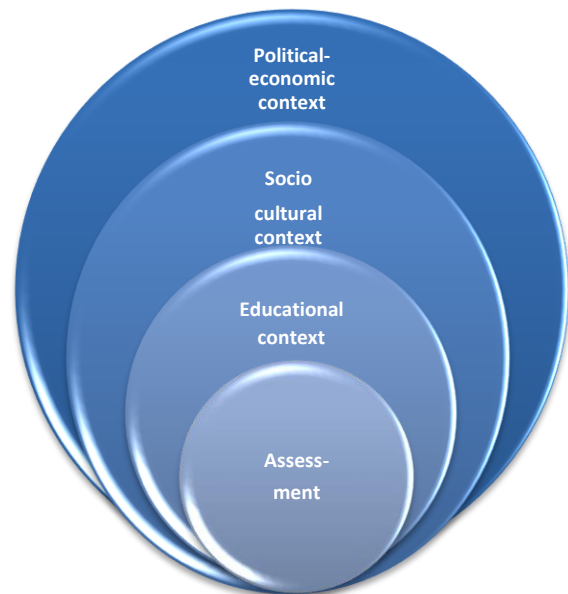
Schools and teachers around the world are challenged by a growing diversity and increasing complexity (Cochren-Smith, 2014). Research in Antwerp notices the problems schools face with respect to the increasing diversity. In 2010, three in ten elementary schools indicated that the load and complexity of diversity was too high (Schraepen, Lebeer, Vanpeperstraete, & Hancké, 2010). Although students always have been different from each other in a variety of ways -- for example, regarding socio-economic status, gender, learning preferences, interests, talents,… -- the factual and perceived diversity has increased over the last few years. In the context of Flanders, different sources and data from diverse domains in society affirm the diversity existing in society and as a result, also in the classroom, and indicate that this diversity will only increase or become more complex.

In Belgium, ethnic diversity increased by 'old' and 'newer' waves of migration (Verhaeghe, Van der Bracht, & Van de Putte, 2012). In the early '60s, people from Turkey and North Africa moved to Belgium because of economic migration. More recently, new waves of migration arrived as a result of the EU expansion, with new memberships from Poland, Romania, Slovakia and Bulgaria. Current migrants come from the Middle East, more specifically from Syria, Iraq and Afghanistan, where people are leaving their country because of war and violence (CGVS, 2015). The waves of migration are represented in statistics which affirm the actual diversity: in 2011 12,7% of the adolescents in Flanders had non-Western roots (Noppe & Lodewijckx, 2013). That number will further increase, as statistics show that 30% of children under the age of 12 have a foreign origin (Kind & Gezin, 2013).

As a result of the increase of ethnic diversity, linguistic diversity also rises. The number of children with a language different from the official language in Flanders and the school language is increasing in both elementary and secondary education. In September 2014 there were 1114 such pupils in elementary education, and in September 2016 that number rose to 2119. A similar trend is visible in secondary education, where the corresponding number increased from 1712 in 2014 to 3328 in 2016 (Agodi, 2017). The further growth of linguistic diversity was also published in a report of Kind & Gezin (2013), revealing that for 25,5% of the children born in the Flemish region in 2013, the language between mother and child is different from the official language.

Classrooms have always been diverse, e.g. when it comes to the presence of pupils with different capabilities. In Flanders, the educational system is organized in such a way that the diversity in school tracks intensifies the further one progresses in one's educational career. Vocational education, in particular, presents a wide variety of differences between pupils resulting from the fact that pupils start off in a higher-valued track (general education) but then sometimes drop down to a lower-valued track (vocational education) (Van Praag, Boone, Stevens, & Van Houtte, 2015). Another feature of the educational system in Flanders that leads to more diversity in classrooms is grade retention. In comparison with other countries, grade retention in Flanders is rather high (Ikeda, 2011). Results from PISA show that 34,90% of the Belgian pupils at the age of 15 repeated at least one year in their educational career.

Schools did not always have the mission to support differences between pupils in order to help every pupil succeed at school. In the past, pupils' assignments to classroom and schools have fostered segregation rather than encouraged inclusion (Banks et al., 2007). Since the beginning of 1990, policy measures have been implemented in Flanders to reduce segregation in education, with the result of intensifying the implications of the above-mentioned demographic and social-cultural realities. One of the recent policy measures was the 'Decree for equal educational opportunities' in 2002, which provides an integrated framework of support to provide equal educational opportunities in both primary and secondary education. The decree resulted in positive effects, but many opportunities remain unexploited to foster equal educational opportunities (Nicaise et al., 2014). A second recent policy measure is stimulated by the ratification of the UN convention (UN, 2007) on the Rights of Persons with a disability. Article 24 of this convention states that every child has the right to attend a regular school. In order to meet this goal, Belgium agreed to develop a more inclusive educational system. This movement is encouraged by the Flemish M-Decree 'Measures for students with special educational needs', approved in 2014. The decree is to be gradually introduced starting from September 2015 and has as its aim to realize more inclusive education and ensure that in the long term

fewer children with special educational needs are referred to special education. A third and the most recent policy measure is a response to the high socio-economic achievement gap in Flanders, which is among the highest in Western countries. The educational reform in secondary education had the aim to move towards a more comprehensive educational system in which tracking in secondary education is delayed. The aim is to limit the tendency of social tracking in secondary education and dissolve the hierarchical system that has developed among the types and subjects in secondary education (Nicaise et al., 2014).

The move towards more inclusion in education and the efforts to provide equal opportunities result in an increase of diversity in the classrooms. These measures of the government, taken together with the earlier described demographic and social-cultural tendencies, will lead to a an increase of factual and probably stronger perceived diversity. A wide variety of diversity within the student body of any given school is challenging for teachers and more specifically for their assessment policy and practices. For instance, the evaluation of pupils with a different language than Dutch raises questions of validity when tests are developed in the official language. When teachers want to assess summatively what pupils have learned, which tools and practices should they choose in order to ensure that every pupil be able prove himself/herself? How do educational professionals guarantee that every pupil get the opportunities he/she needs to make progress in his/her learning, or in other words, to what extent is 'assessment for learning' modified to the needs of *all* pupils?

4. The role of socio-cultural context for assessment practices

The educational context is embedded in a socio-cultural context and as a result, the socio-cultural context will have an impact on how different stakeholders act in the educational context. Since the educational context is characterized by diversity , the socio-cultural context will play a role in how educational professionals will deal with this changing educational context, depending on their view on diversity. Additionally, teachers always create their learning environments based on their beliefs and convictions about how learning happens. This chapter provides an overview of the three main learning theories, all of which are influential in teachers' practices. Teachers' learning environment and instructional decisions are shaped by these learning theories. As teachers make instructional decisions on the basis of 'fitness for purpose' (James, 2006), an amalgam of different approaches in instruction and assessment appear in classrooms. Thus, attention will be paid to how these theories of learning play into teachers' conceptualization of both 'teaching and instruction' and 'assessment'. Furthermore, the importance of alignment between 'learning', 'instruction' and 'assessment' will be explained by referring to research that has proven the need for alignment in order for learning to happen.
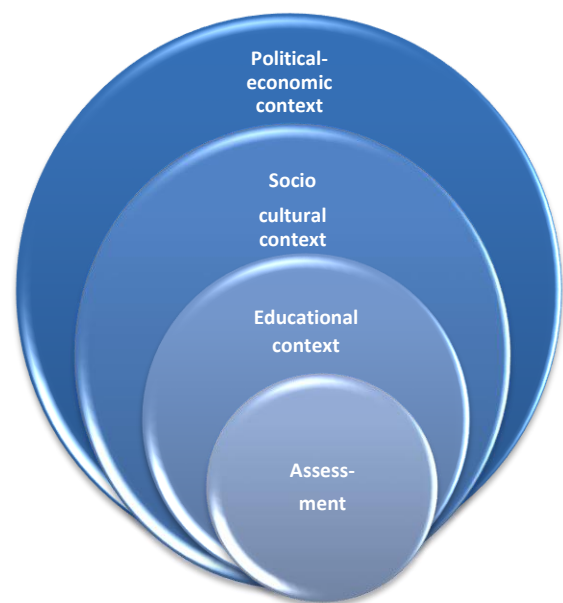
4.1. Diversity theory

The fact that diversity is increasing in society and consequently also in schools and classrooms calls for a description of diversity theory. Diversity theory will facilitate the interpretation of assessment policy and practices in light of the increasing diversity, as people's actions are a predominantly framework or vision.

As described by Van Avermaet and Sierens (2010), two main approaches to diversity can be distinguished: a *group vision* and an *individual vision*.

The *group vision* stems from an assimilationist view, whereby certain groups are (seen as) disadvantaged and needing support in their process of assimilating into society. As a result of equal

opportunities ostensibly being provided to everyone, some groups are labelled as *disadvantaged* or *risk groups* because they are at risk of missing out on the opportunities. This group vision leads to categorizing society in groups, often based on physical or innate characteristics. Consequently, by labelling groups as *disadvantaged,* e.g. women, an opposite group is automatically formed and functions as the advantaged group and thus, the ideal to strive for: e.g. 'women' vs. 'men', 'LGBT's' vs. 'heterosexuals', 'elderly employees' vs. 'young employees', 'natives' vs. 'migrants'. A strong *group vision* encourages practices like labelling, stereotyping and stigmatizing. From this point of view, the group vision creates homogeneity within diversity by using labels for groups, which is in contrast with the *individual vision* on diversity that emphasizes the uniqueness of each individual. This perspective seems to avoids the negative effects of the *group vision.* However, Van Avermaet and Sierens (2010) warn against the negative effect of the individual vision: blindness for the social and cultural influences. We must keep in mind that individuals are always related to communities or groups. According to Van Avermaet and Sierens (2010) the challenge is thus to combine the two approaches, the individual and the group visions, in such a way that the negative effects of both can be avoided in order to end up with an approach that respects the nature of diversity, namely that diversity not only occurs between groups, but also within groups. In that sense, diversity is as much about 'sameness' as about 'otherness'. These different approaches to diversity can help us to understand the practices of schools and teachers who are confronted with increasing diversity in their schools and classrooms. More specifically, the assessment policies and practices of schools and teachers can be interpreted via these theories on diversity as well as the learning theories.

4.2. Learning theories and their view on assessment

The following sections summarize the three most prominent theories of learning: behaviorism, cognitivism and (social-) constructivism (Valcke, 2008). These theories will be explained in the same order as they emerged through history, for each theory can be seen as a reaction to the preceding theory. In reality, the ruling paradigm during a period of transition can be described as a theory in itself, but here we will give an overview of the three main paradigms of learning as they are described in literature. It is important to notice that with the evolution of new theories, the previous ones are not *replaced* by the newer ones, but present a different point of view on how learning happens. As a result, each theory has its strengths and weaknesses which is the reason that all these theories are still meaningful for education today.

In the following paragraphs, attention will be paid to how these theories impact instruction and assessment.

4.2.1. Behaviorist theory of learning

The behaviorist theory of learning emerged in the 20th century and is rooted in the field of psychology, where learning was seen as the association of stimulus and response (Valcke, 2008). The mechanical process of associating stimulus and response, which produces new behavior, constitutes the core idea of the behaviorist theory of learning. Behavior is strengthened by reinforcement, so it is the teachers' task to reinforce the desired behavior. Behaviorists view the learner as a passive person who responds to the stimuli. According to this view, the learner starts as a 'blank page' and the behavior is shaped by reinforcement (Valcke, 2008).

The dominating paradigm during this time was that of the social efficiency movement. A leading figure in this movement was F.W. Taylor, an engineer who wanted to render American industry more efficient. His key principle was the task idea: each worker should be given a narrowly defined assignment that he was to perform at a specific rate using certain predefined procedures. Notice that the focus is on performance, on carrying out the desired *behavior.* This task idea was believed to maximize the efficiency of factories. Consequently, *Taylorism* influenced the area of education as well: educational programs were designed to eliminate waste, and it was wasteful to teach pupils things they would never use (Shepard, 2000). The idea of efficiency also affected the area of curriculum development, which began to pursue a deconstruction of complex performances. Learning could be accomplished by training first the basic skills before then introducing the complex skills in curriculum planning (James, 2006).

The strength of behaviorism is that the success of outcomes is easily measurable, which brings us to the role of measurement practices in the context of a behaviorist point of view. Tests are used to determine whether a pupil masters the desired knowledge or skill. Assessment is seen as a *check* or *control* of the success of instruction. From a behaviorist point of view, assessment is considered as 'assessment of what has been learned.'
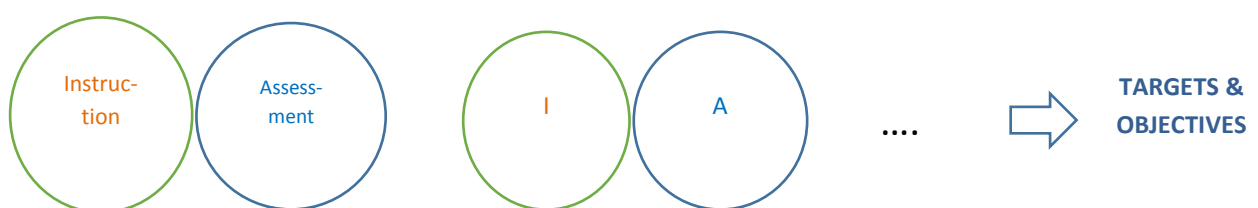

Fig. 2. Assessment in behavioristic theory of learning.

In order to meet the hierarchy of skills, tests should be used frequently to measure progress. Mastery of a skill has to be ensured by a good performance on a test before proceeding to the next level in the hierarchy of skills. Consequently, precise standards of measurement are needed to guarantee that a skill be mastered at the desired level. Poor results on tests lead to remedial activities before progressing to more complex skills. However, the question can be asked as to what extent remedial activities fit into the idea of efficiency. It seems that this question was not perceived as relevant due to a strong, prevalent belief in the predictive value of IQ tests.

Scientific measurement was highly valued in that period, as it corresponded to the driving philosophy behind the social efficiency movement: that science could provide the answer to the problems of industrialization.   The belief in the utility of IQ tests corresponded with the idea of efficiency insofar as scientific measurement of ability was used to predict one's future role in life, thereby determining who was best suited for each vocation. The IQ test was perceived as the ultimate tool for avoiding *waste* in the educational system, since it would ostensibly guide the pupil to the best fitting educational program. This shows how assessment practices have the potential to shape society, which will be discussed more extensively in a later section (Shepard, 2000).

The ideas of behaviorism as a theory of learning were in time outweighed by the cognitivist theory of learning. Despite the changes in learning theory, the ideas concerning the measurement paradigm are still present in current educational practices, as Shepard (2000) explains:

> *"It is no coincidence that Thorndike was both the originator of associationist learning theory and the "father" of "scientific measurement," (…). Thorndike and his students fostered the development and dominance of the "objective" test, which has been the single most striking feature of achievement testing in the United States from the beginning of the century (20th) to the present day* (Shepard,2000, p.5).*"*

4.2.2. Cognitivist theory of learning

The cognitivist theory of learning arose alongside and often in reaction to the ideas of behaviorism. It was argued that not all learning occurs through the shaping and changing of behaviors (Valcke, 2008). The mindset of cognitivism must be seen in the same period in which computers made their ascent in daily life. Consequently, cognitivism refers to the study of the mind and how it obtains, processes and stores information. The concept of cognitivism is also influenced by the growth of neuroscience and brain research during the 1960s (James, 2006). As the reference to 'cognition' makes clear, this theory is interested in 'mind' as a function of 'brain'. This is in contrast to the theory of behaviorism where

the brain was seen as a 'black box'. The mind functions like a computer processor: information comes in as input, the mind processes the information and the information is stored in memory to be retrieved later. The differences between 'capable' and 'weaker' pupils is marked by the way the more capable pupils organize knowledge in structures that make it more retrievable. Their conceptual structures are more resilient. Cognitivists frame learning in terms of creating conceptual structures in pupils' minds and the use of strategies to shape these structures and to retrieve them fluently when the information is needed. Prior knowledge refers to an 'existing conceptual structure in a pupil's mind which can be elaborated by expanding it with new conceptual structures (Valcke, 2008); this is regarded as a powerful determinant of a pupil's capacity to learn new material. Processing strategies, such as deductive reasoning from principles and inductive reasoning from evidence, are important to stimulate the growth of the conceptual structures. Learning is thus shaped by acquiring learning strategies and prior knowledge.

This view on learning has implications for the view on instruction. According to the cognitive view of learning, the teacher still has a central role, being in charge of presenting new information in an organized manner with the goal of elaborating the pupils' conceptual structures. To avoid *cognitive overload* teachers will break down the new information into smaller parts, which is often referred to as '*chunking'* (Valcke, 2008). By splitting up the targets, pupils will not be overwhelmed with incoming new information. To facilitate learning, teachers will also elicit pupils' mental models to enable the connection between the new information and the existing conceptual structures. Although learning is still teacher-centered, this perspective sees the role of the pupil as more active than the behaviorist perspective does; acknowledging that learners do not function as a 'black box', this learning model requires more active engagement of them (Valcke, 2008).

Just as this theory of learning results in a particular view of instruction, it also consequently affects the corresponding view on assessment. Given the strong influence that prior learning has on new learning, teachers will activate pupils' prior knowledge through classroom dialogue, open-ended assignments, stimulating thinking-aloud, using mind-maps, …. to scaffold their understanding of knowledge structures. These practices correspond to the concept of 'formative assessment', since assessment is an integral element of pedagogic practice (James, 2006). In fact, teaching and assessment are both so necessary to achieve the targets and objectives that the two are almost indistinguishable (James, 2006). Note that according to the cognitivist point of view, 'assessment' – and not 'instruction'-- is seen as the starting point of the instructional process, which is in contrast with the behaviorist point of view where the process starts with instruction and ends with 'assessment'. 'Assessment' in the beginning

of the instructional process refers to the activation of pupils' prior knowledge, since it is serving a dominant condition to achieve learning.
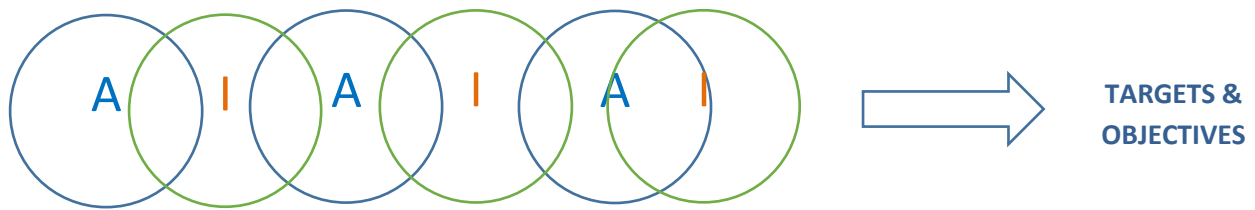

Fig. 3. Assessment in cognitivist theory of learning.

4.2.3. (Social-)constructivist theory of learning

The constructivist theory of learning emphasizes the role of the learner: learning is considered an active process. Consequently, the learner is seen as an active participant in the construction of knowledge: the learner takes in new information and gives meaning to it, using his or her own prior experiences, beliefs and attitudes as reference (Valcke, 2008). The Reworking of the theory of constructivism has led to some *sub*-theories, each with their own emphases, for example 'social cultural theory', 'metacognitivism' and '*social* constructivism' (Thurlings, Vermeulen, Bastiaens, & Stijnen, 2013). Palincsar (1998), who wrote a review study on constructivist perspectives on learning, stated that all sub-theories reject the idea that the locus of knowledge construction is situated in the individual – which is the case in the cognitivist theory of learning. Learning and understanding occurs in interaction between the individual and the social environment. Vygotsky, a psychologist who was influential for constructivism, compiled his ideas in his work *'Mind in Society',* a title that calls attention to the interaction with the environment (James, 2006). Constructivists place great emphasis on situated learning, or seeing learning as contextual. The purpose of learning is to achieve multi-contextual learning to ensure that pupils can broadly apply the information they possess.

The notion that learning is an activity in which cultural artifacts have a crucial role is essential in social constructivism. These artifacts can be understood as handbooks or other physical learning materials, but also as symbolic material, such as language. According to Vygotsky, language plays a central role in our capacity to think and is developed in relationships between people. Social relationships are therefore necessary for learning. For social constructivism, learning is by definition a social activity in which people develop their thinking together (Valcke, 2008).

The meaning of social constructivism for the practice of instruction is that the role of the teacher is to facilitate pupils' construction of knowledge, rather than supposing that they will accept the information from the instructor. As learning occurs in interaction with the environment, collaborative

learning should be stimulated by the teacher. Group work is not just an option but a necessary condition. Considering that learning takes place in a context, teachers should create meaningful learning environments with learning materials that include examples that relate to pupils, so they can make sense of the information. Teachers are in charge to guide the learning process, but the responsibility for learning is ascribed to the learner. There should be a form of guided discovery where learners are allowed to make decisions on learning goals (Valcke, 2008). Learners are getting opportunities to customize their curricula according to their interests, talents and strengths. As a result, constructivism does not favor the idea of a standardized curriculum.

According to the conclusion of Green and Gredler (2002) in their review and analysis of constructivism for school-based practices, constructivism is a recent educational movement with widespread intuitive appeal but no strong empirical base. The constructivist theory of learning is not yet well researched in terms of its implications for assessment practices (James, 2006). Nonetheless, some have maintained that constructivism is reflected in the concept of 'assessment *as* learning' insofar as it describes the role of the learner, the one who is assessed, as the *critical connector* between the assessment and learning process (Clark, 2012; Earl, 2003). The learner as 'critical connector' refers to the assessment as a personal affair in which the learner makes sense of new information and progresses in his/her learning. Assessment in this respect is embedded not only in the event of teaching, but also in the event of learning. Assessment and learning coincide in the role of the learner as assessment stimulates his/her metacognition and self-regulated learning. Assessment *as* learning focuses on the social contexts and on self and peer-assessment activities, since both types of assessment activities encourage metacognition and self-regulated learning.

The idea that 'assessment *as* learning' corresponds to a social-constructivist theory of learning is reinforced by Clark (2012), who attributes the following four goals to 'assessment *as* learning':

- Provide opportunities for students to become meta-cognitive and build knowledge of themselves as learners by encouraging them to evaluate and reflect on the quality and progress of their work.
- Create a non-comparative, productive environment, free of risks to self-esteem and founded upon cooperation and dialogue.
- Support students as they take more responsibility for their learning.
- Provide opportunities (for both teachers and learners) for frequent participation in the process of learning with their teacher as their advisor and with their peers in a climate of equality and mutuality (Clark, 2012).

These opportunities for mutuality are also included in the concept of *dynamic* assessment, which is referred to by Palincsar (1998) as an assessment practice in a social constructivist perspective. Appointing assessment as *dynamic* emphasizes the contrast with the *static* character of traditional assessments, as those yield information about an individual's actual level of development. Dynamic assessment, on the other hand, provides information of the potential level of development by indicating abilities that are developing and thus, is predictive of how a learner will perform independently in the future.

The relationship of assessment and instruction within a social (constructivist) theory of learning is portrayed in fig. 1.3. In contrast with the previous models of assessment in thebehaviorist and a cognitivist theories of learning, the activity of instruction and the activity of 'assessment' are much more related to each other. To contrast the view of behaviorism and cognitivism in which both activities are seen as respectively separate or interconnected activities, both lines in fig. 3 are synchronized/overlapping to emphasize how embedded learning is within the activities of 'assessment' and 'instruction'. The irregular shape of the line refers to the feature of assessment as a personal affair in which the learner makes sense of new information to progress in his/her learning. As assessment is a personal affair, it will lead to learning processes that evolve not always in a linear way, which is the case in the behaviorist and cognitivist point of view. Thus, the wobbly line in this model symbolizes the natural way of learning. The *route* of this line refers to a learning process in which assessment and instruction are not always clearly distinct activities, as assessment stimulates learners' meta-cognition and self-regulated learning.

TARGETS & OBJECTIVES
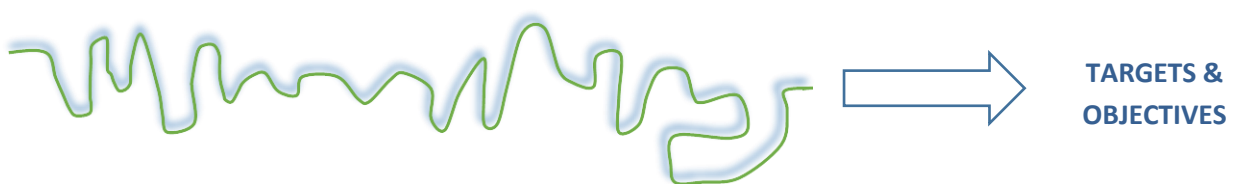
Fig. 4. Assessment in social-constructivist theory of learning.

4.2.4. Alignment of instruction and assessment

Looking into the concepts of different learning theories, it is evident that all the learning theories have their own importance and strengths. In practice, it means that teachers will act eclectically in making decisions about the instructional process. Behaviorist approaches, for example, seem to work perfectly well when the focus is on the development of basic skills, automatic actions or habitual behavior

(Valcke, 2008). Cognitivist approaches, on the other hand, are more appropriate when the focus is on deep understanding of conceptual structures. Constructivist approaches, finally, are recommended when complex problem solving is the purpose and when ownership of learning is required (Valcke, 2008). As a result, teachers make instructional decisions on the basis of 'fitness for purpose' (James, 2006) and an amalgam of different approaches in instruction will appear in classrooms.

The aforementioned changes in thinking about how learning occurs has implications for both the practice of instruction and assessment. The paradigm shift from assessment *of* learning to assessment *for* learning and subsequently to assessment *as* learning has challenged educational professionals to organize assessments that are all-purpose (Pattalitan Jr, 2016). According to James (2006), effective teaching requires consistency between assessment practices and the beliefs about learning. Changing ideas about learning require a change in assessment practices. A lack of alignment between instruction and assessment undermines effective teaching and is therefore detrimental to successful learning (James, 2006; Shepard, 2000).

Incongruence between assessment practices and beliefs about learning can be explained by the fact that assessment is subject to the belief systems of different stakeholders (James, 2006). Shepard (2000) described how traditional views of testing, corresponding to the ideas of a testing culture (cf. 1.1), are still present in the belief systems of stakeholders and function as the default framework affecting and driving current practices. It seems that the earlier described paradigm shift has not yet taken place at all levels of the process of learning and instruction (Shepard, 2000). Shepard explains that the dominance of objective tests in classroom practice has affected beliefs about the nature of evidence and the idea of fairness. The tradition of a 'testing culture' resulted in corresponding beliefs about evidence, objectivity and fairness, beliefs that still persist in teachers, parents and policymakers (Shepard 2000). According to teachers, parents and policymakers, assessment needs to be separate from instruction, and uniformly administered assessments are preferred to ensure fairness. Similar findings are detected by Hargreaves (2005), who conducted a survey among teachers that provides plenty of examples of teachers holding the objective model of testing. Teachers for the most part apparently believe that the objective model is the *correct model,* even if their own beliefs about learning and instruction do not match up with it.

The relationship between teachers' beliefs about learning and their assessment practices are likely affected by internal conflicts, contextual factors, and external pressures (Bonner, 2016), which will be explained in the following chapters.

4.3. The socio-cultural context plays a reciprocal role at individual & societal levels

Policies of assessment, systems of assessment and assessment practices (both development and use) do not occur in a vacuum. They are embedded in a socio-cultural context, and therefore, assessment policies and assessment tools are social constructs. They are never neutral nor objective (Shohamy, 2001). The role of the socio-cultural context occurs at both the individual level (of teachers and pupils) and the societal level. At the individual level we see that the socio-cultural context influences teachers' assessment practices. As described above, the learning theory out of which a teacher designs a learning environment has an impact on his/her instruction and his/her assessment practices. Furthermore, teachers sometimes modify their assessment practices to the pupils. For example, the ethnographic research of Stevens (2007) reveals that teachers in some cases lower their standards for pupils to avoid problems with these pupils or their parents. Also on the individual level, but on the side of the student, we see that assessment practices can have an influence on students' future life. The information produced by assessment instruments shapes decisions about the test taker, which impact his/her future educational career. The results of assessments can exclude pupils from certain disciplines and courses, and in so doing, also ultimately has an impact on one's opportunities on the labor market. Consequently, test results of a child are a deciding factor in his/her entire future life (Noam, 1996). These consequences of tests can foster test anxiety, which may have significant negative effects on individuals' test performance, as well as on their attitude towards assessments in general (Nemati, 2012). Test anxiety leads to difficulties in understanding the questions and in organizing one's thoughts; it can also lead to mental blocks. Test anxiety can be responsible for students' ability to perform at their maximum level. Cakici (2016) presents an interesting study on the relationship among pupils' characteristics, like age, gender, ethnic group and socio-economic background and amount of test anxiety.

More generally, the socio-cultural context plays a role on the level of society when it comes to the educational system, which is characterized by a particular point of view on learning and teaching and is thus related to a specific learning theory. The broader socio-cultural context also plays a role in how assessment is perceived and employed in that society. Ercikan and Solano-Flores (2016) describe that assessment practices are shaped by the characteristics of a society. Its values and social structures influence the content that gets assessed as well as who gets assessed. In other words, the culture and the society that generate assessments is reflected in how these assessments are created and to what end they are used. For example, the types of items and the form of language used in an assessment assume test takers' familiarity with a given set of everyday life experiences and views (Ercikan &

Solano-Flores, 2016). Consequently assessments and, more broadly, assessment systems could be seen as cultural products.Test takers who are unfamiliar with the culture in which these assessments are embedded have a clear disadvantage, which could lead to unfair assessment results.

Furthermore, the way assessment systems are perceived in a society can be determined by its culture. In East-Asian countries, for example, large-scale standardized examinations are perceived as being a mechanism to create equal opportunities for all students (Kennedy, 2016). These tests are viewed as the basis of meritocracy because they guarantee that all students are treated equally and fairly, free from family or other social influences. In China, for instance, the population has faith in the Chinese National College Entrance Exam, as it symbolizes the opportunity for upward social mobility (Kennedy, 2013). This view is rooted in the Confucian principle of treating all students equally, which stands in contrast with a major criticize of Anglo-Saxon countries: that such examinations are unfair (Kennedy, 2016). In the US for example, the introduction of 'education for all' in the 20th century was accompanied by the introduction of large-scale tests (Shohamy, 2001). Owing to industrialization and the resulting population growth, the American society changed from an ascribed society, where positions where predetermined, to an achieved society, in which individuals have the right to find their place regardless of their social background. The use of tests was seen as a procedure for fair selection and to provide equal opportunities to all pupils. But, as stated by Shohamy (2001), these selection tools turned out to be more of an illusion than a reality. This illusion is confirmed by a large body of research that examined how assessments could underestimate academic proficiency of students who are raised in a different language than the language of schooling (Ercikan & Solano-Flores, 2016). As shown by testing research, language proficiency of the test-taker has an impact on the test results, certainly in the case were a content test (for example maths) is administered by a second-language learner (Menken, 2010). Stigmatizing conclusions are unfairly drawn about pupils with a different home language based on these lower performances. The research of Menken (2006, 2008, 2010) affirms the illusion of equal opportunities by stating that the idea of equal opportunities does not apply to pupils of a second language. The results of high-stake tests showed that only 33% of pupils with English as a second language succeed, whereas 80% of the total population of students succeed. As a result, schools changed their curriculum and started to teach-to-the-test. A more devastating effect of these tests was that schools started to select pupils whose chances of succeeding on the high-stakes tests appeared favorable. Schools coordinated their policy for enrollment of new pupils in such a way that pupils with a different language could not enroll in their school. As a result, whole groups of pupils got excluded from schools. In the end, the idea of equal opportunities was nullified.

International large-scale assessments foster and generalize the disgracing conclusions: the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science

Study (TIMSS) indicate that pupils who do not speak the language of schooling at home perform lower than pupils who master the language of school.

At this point, where assessments (whether large-scale or classroom-based) may underestimate the academic proficiency of the test-taker, the reciprocal relationship between assessment and the socio-cultural context becomes manifest. Hence, misjudgments of pupils with a different home-language, learning difficulties or other special educational needs may cause assessment practices which function as mechanisms of exclusion in education and, by extension, in society. In this respect, the influence of assessment on society can play a negative role as assessment can be used as a gatekeeper of access to education (Ercikan & Solano-Flores, 2016). The fact that unfair test results impact the schooling and thus the life and career of an individual can be extended to whole groups of society: assessment practices have the detrimental potential to maintain rather than change educational inequalities in a society (Ercikan & Solano-Flores, 2016).

5. The role of political-economic context for assessment practices

The role that the political-economic context plays in assessment policy and practices will be discussed in the following paragraphs. First, the impact of (inter)national assessment programs on policy will be discussed, recognizing that local politics of today are embedded in a context of internationalization and globalization (Standaert, 2008). In a second section, the power of tests as a disciplinary tool will be revealed, referring to theoretical assumptions and concrete examples. Similar to the socio-cultural context, reference will be made to the effects of standardized testing. The closing section will define this reciprocal character of the political context, the socio-cultural and educational context and ultimately its impact on assessment practices.
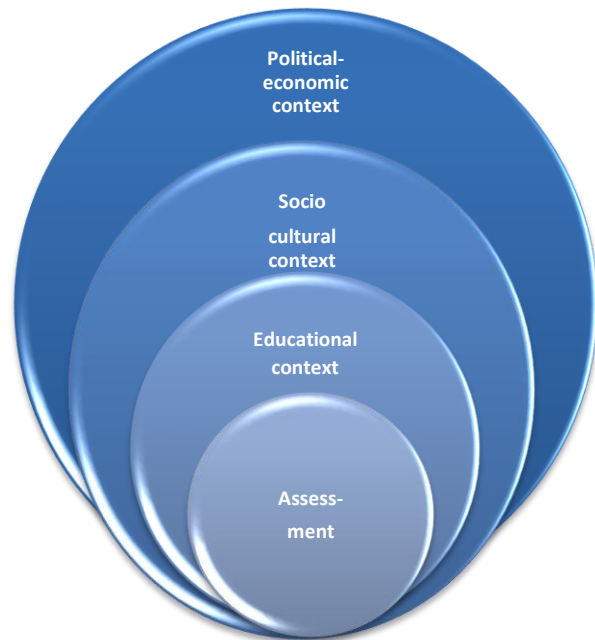
5.1. Impact of (inter)national assessment programs

International assessment programs are emerging in the context of globalization (Standaert, 2008). A more global economy has influenced the field of education since the 1960s – the connection between economy and education was actually put forward by educational economists. During that period, education was increasingly considered as an investment for economic growth. Research, based on a model of input-process-output, reinforced the idea of educational investment by revealing that one year of education results in 3-6% of economic growth of a nation (Standaert, 2008). Against this background, international organizations, like the *Organisation for Economic Co-operation and Development* (OECD), arose to share knowledge that could be used to make progress as a nation in the context of globalization (Standaert, 2008). In order to improve education policies and outcomes, the OECD published '*Education at a glance'* and '*Education policy analysis'*. Both publications present and interpret a whole range of indicators of the member states. Outcomes of educational programs in terms of students performances came in the picture in 2000 when the *Program for International Student Assessment* (PISA) was first performed by the OECD. A similar organization, *International*

Association for the Evaluation of Educational Achievement (IEA) introduces two assessment programs serving the same objective: '*Trends in International Mathematics and Science Study'* (TIMSS) and *'Progress in International Reading Literacy Study'* (PIRLS).

The data resulting from these international assessment programs are attracting great media interest in the participating countries, to the extent that the publication of the results contribute to the ranking of educational systems and to comparisons of educational policies. This sometimes leads to competition between member states (Standaert, 2008, 2014). Consequently, international assessment programs are considered as a powerful tool for policies to force educational reform. As an response to the alarming results of PISA, Germany implemented nation-wide standardized tests in the beginning of the 21st century to improve the quality of education (Ercikan & Solano-Flores, 2016; Klein, 2013; Standaert, 2014). Germany's response to the PISA results is exemplary for the fact that international organizations and their assessment programs disseminate coherent norms and policy models among their member states (Teltemann & Klieme, 2016).

Closer to home, the educational government in Flanders also refers to the PISA results to reinforce the reform of secondary education: although the results show that Flanders provides high-quality education, they also reveal that the socio-economic achievement gap is among the highest in Western countries (OECD, 2013). In order to bridge this socio-economic achievement gap, a reform of secondary education was announced to limit the tendency of social tracking, which is strongly apparent in the current education system (Vlaamse Regering, 2013).

Another noteworthy example can be found in the US, where in 2002 the Bush administration signed a school reform policy called the *No Child Left Behind* act (NCLB). The act required all public schools to administer a statewide standardized test in mathematics and in reading every year in grades 3 through 8 (Stiggins, 2002). The scores on these tests and the progress of schools through the years were used as legitimation to force schools with poor results to take up corrective actions or to engage in supplemental education service. These dispositions serve as a condition for schools to maintain federal funding. In the worst case, schools face decreased federal funding. This faith in assessment as a tool for school improvement at the political level is responsible for adventitious and large effects in the classroom, as schools and teachers were then held accountable for the results on statewide standardized exams. The increased accountability had an impact on the daily classroom practices, especially insofar as the year-end assessments led to the development of interim assessments. These assessments, sometimes referred to as benchmark assessments, were also standardized, though they were administered at school or district level throughout a school year in order to monitor the improvement of student outcomes in the year-end assessment (Schneider, Egan, & Julian, 2013). The

demands for accountability not only led to the *over*testing of pupils but also affected the curriculum and instruction practices: teachers began *teaching-to-the-test* in order to pursue high scores on the international tests (Standaert, 2008, 2014). Using results of standardized tests as the basis for decisions about the pupils' entire educational career invariably leads to a narrowing of the curriculum. Curricula then ceased to take into account the complex skills needed in the 21st century, simply because such skills were not represented in the test, and thus did not fit into the practice of teaching-to-the-test (Fullan, 2011). In the UK, the same trend is visible since the introduction of a National Curriculum, accompanied by the installation of National Assessments to measure pupils' attainment of the curriculum objectives (Hargreaves, 2005).

Although these policy decisions are made with the aim of improving the national educational system and the corresponding educational outcomes, these intentions run the risk of damaging educational systems rather than improving them. The effects of tests are often more far-reaching than intended, as they tend to be used for purposes beyond those envisioned at first (Shohamy, 2001). Opponents of nationwide standardized testing have strongly criticized its negative effects, citing that it leads to a narrowing of the curriculum because it forces teachers to *teach-to-the-test,* and that it leads to a *'blindness by numbers',* referring to teachers not seeing what else needs to be taught since they are preoccupied with achieving the right 'numbers' (Standaert, 2008, 2014). As described in the previous section, these standardized tests perpetuate the unequal educational opportunities (Menken, 2006, 2008, 2010) Moreover, the extent of the damage has to do with the degree of misinterpretation of test results or the use of tests in ways not intended by test developers or education systems (Ercikan & Solano-Flores, 2016; Standaert, 2014). The faith in standardized testing has sometimes outweighed the insight that competency is needed to interpret the results of standardized tests and to translate them into educational measurements set by policies (Campbell, 2013; Ercikan et al., 2015; Stiggins, 2002).

5.2.   Tests as a disciplinary tool

In the UK, faith in standardized (inter)national tests was so strong that it even cut through the tensions between opposing political ideologies. For example, the National Curriculum and the corresponding national assessments were supported by both left- and right-wing political parties in the UK. For the right wing, the national curriculum and the national tests are an answer to their needs for standardization and the principle of free market, considering that parents can inform their choice of school using the schools' published test results. As for the left wing, the new standards were seen as a

means to achieve educational equality, since all pupils were meant to achieve the same level; this resulted in the belief that the national curriculum would reduce the number of pupils *at risk* (Hargreaves, 2005).

As politicians have been appealing to (inter)national assessment practices in the debate about education reform, the use of assessments got transferred to other areas in which political and ideological agendas play a role (Ercikan & Solano-Flores, 2016). Governments have increasingly come to realize the potential of tests as a mechanism of state control (Broadfoot & Black, 2004). Michel Foucault wrote on this topic in 1979, stating that examinations can legitimize decisions of a government; they have built-in features that allow them to be used for exercising power and control. Australia, for example, introduced two language tests '*Australian Assessment of Communicative English Skills*' and '*Special Test of English Proficiency*' in order to reduce the number of immigrants. The results of the tests were used to decide whether someone was eligible for permanent residence (Shohamy, 2001).

In Flanders the now abolished language test at the beginning of primary education shows how a government can drive parents' choice to enroll their children in pre-primary education even though education is only compulsory from the beginning of elementary school onwards (Mertens, Ysenbaert, Vanderlinde, & Van Avermaet, 2015). In 2011 a language test was introduced for pupils starting primary education. This test was obligatory for children who attended less than 220 half days in the last year of preprimary education and nevertheless wanted to enroll in primary education. By introducing this test, the government wanted to stimulate as many parents as possible to send their children to pre-primary education, despite it not being compulsory. The measurement must be seen in the framework of a broader policy that focused on the importance of language proficiency. In 2015 the obligatory language test was abolished because opponents disputed that repeating the last year of pre-primary education because of language delay damages children's progress in other developmental areas (Vanhoof et al., 2016). Instead of this language test, schools are nowadays obliged to screen the language proficiency – which is a type of assessment – of children who enroll for the first time in primary or secondary education, in order to gain insight into their language proficiency and to take measurements to stimulate language acquisition. Schools are free to choose an instrument for this screening of children's language proficiency. In order to support schools, the government offered a free 'Toolkit for Alternative Assessment' which supports schools in establishing an assessment program for the language competences for Dutch (Philips, I., Seghers, M., Versteden, P., & Ysenbaert, J., 2013).

In *'The power of tests,'* Shohamy (2001) comprehensively describes the interrelatedness between assessment policy and assessment practices, on the one hand, and politics, on the other hand. According to Shohamy (2001), tests serve as disciplinary tools that enable those in power to emphasize and carry out their policies. The power of tests lies in their capacity to influence and change the behavior of the test-taker in the direction that the test-givers believe to be important.

5.3.   Reciprocal character of the political context

Assessments and the role they play are embedded in educational, socio-cultural and political contexts. The relationship between assessment and the political context is reciprocal. On the one hand, assessments and its results induce governmental decisions to reform education. For example, the German National policy was guided by international assessment programs to adjust their policy concerning educational practice, curriculum and instruction. On the other hand, as outlined in the previous section, the political context uses assessment as a tool to pursue political and ideological motives (Ercikan & Solano-Flores, 2016). Moreover, the role of the political context cannot be separated from the socio-cultural and educational contexts, since these contexts continually interact with and consequently influence each other. In other words, the political context in education cannot be underestimated, as the purposes of assessment and the use of information obtained from assessments are modified by a nation's social organization and culture (socio-cultural context) as well as its institutions and legislation (political context) (Ercikan & Solano-Flores, 2016).

6. Assessment practices in Flanders

6.1. Educational organization

In Flanders, the constitutional *'freedom of education'* has a wide impact on evaluation and assessment frameworks. 'Freedom of education' has a twofold implication. First, 'freedom of organization' refers to the fact that every natural or legal person is free to start a school. Second, every parent can choose freely in which school to enroll his or her child. The 'freedom of organization' allows each school to develop its own educational policies, including its own pedagogical plan, teaching methods, curriculum. As a result, schools are free to develop their own approaches in evaluation and assessment (Penninckx, Vanhoof, & Van Petegem, 2011).

For a good understanding of evaluation and assessment practices and policies, it is important to have knowledge of the framework of quality assurance in the Flemish education system. The framework is often referred to as the 'Triangle of Quality', which is built around three major pillars; for each pillar an institution bears responsibility to ensure that schools deliver good quality outcomes (Vlaamse Regering, 2009; Shewbridge, Hulshof, Nusche, & Stoll, 2011):

- **Government-determined minimum learning objectives for students**

In 2009 all services related to quality improvement of education fell under the jurisdiction of the 'Agency for higher education, adult education, qualifications and study grants' (AHOVOKS). AHOVOKS is one of the agencies of The Ministry of Education and Training and has the task of developing minimum standards for quality in primary and secondary education. After approval by the Flemish government, the developed standards are consolidated by a decree. The minimum standards are divided in two types shaping the core curriculum for compulsory education in Flanders: (1) attainment targets that schools must ensure that students meet at a certain point in time and (2) developmental objectives to be pursued by schools; schools need to account for their efforts to reach these objectives (Shewbridge et al., 2011). Other tasks involving quality assurance are ensuring clear processes for certification; ensuring the quality of educational institutions; ensuring coherence between the attainment targets and developmental objectives of compulsory schooling and competences of teachers taught in initial teacher education programs; and organizing the National Assessment Program (Penninckx et al., 2011). It is important to notice that the National Assessment Program in Flanders, called *peilingen,* is of a different nature than the high-stakes standardized tests in the Anglo-Saxon tradition. Through the National Assessment

Program, the government researches a representative sample of schools and pupils in order to measure to what extent pupils do reach the attainment targets. Schools participate voluntarily and the results are not made public at the level of the schools (Vlaams Ministerie voor Onderwijs en Vorming, 2017). Recent research reveals that there is no support from the educational field in Flanders to move towards high-stakes standardized testing (Vanhoof et al., 2016). On the contrary, there is a wish for a comprehensive system to measure learning progress. Such a system could collect a broad range of standardized tests in diverse subjects into a database, from which teachers and schools could select the most appropriate ones according to their needs (Vanhoof et al., 2016).

- **External controlling by the Inspectorate of schools' implementation of the centrally set learning objectives and their quality assurance**

The responsibility for the second pillar in the 'triangle of quality' is reserved for the Flemish Inspectorate, which has the task of formulating recommendations in order for schools to gain certification. This certification allows a school to award recognized diplomas and certificates. It is also a precondition for a school to receive public funding. In other words, the Inspectorate inspects the school's curriculum to make certain that it addresses the centrally-set attainment targets and/or developmental objectives. As a result, inspection is compulsory for every public funded school (Shewbridge et al., 2011).

Recently, the framework to inspect the schools has been reformed into a framework in which quality is the central focus. Many different stakeholders (pupils, teachers, school leaders, umbrella-organisations, researchers, ….) were involved in the development of the new framework for educational quality (Vlaams Ministerie voor Onderwijs en Vorming, 2016), which was designed to enhance internal quality control in schools and to stimulate cooperation between schools and pedagogical advisory services. The new framework also functions as the basis for the Inspectorate to review schools in order to formulate recommendations towards certification.

Concerning 'assessment' it is remarkable that assessment is now solidly embedded in the context of supporting, supervising and monitoring pupils' learning processes, whereas in the past, the process of 'assessment' and the process of 'supporting the learning process' were seen as distinct processes. The new framework lays out the function of assessment by specific quality expectations that emphasize the importance of feedback, alternative assessments and instructional adjustments based on assessment information.

- **The offer of support by Pedagogical Advisory Services to schools for developing and assuring their quality**

In Flanders, most school boards are attached to an 'umbrella organization', a representative association of school boards that acts as a partner for schools in policy. School boards may surrender some of their autonomy to their umbrella organizations, for example, by following a curriculum (that meets the centrally set attainment targets and development objectives) developed by the umbrella organization. Each umbrella organization has its own Pedagogical Advisory Service (PBD), an important partner of schools in quality assurance (Penninckx et al., 2011). Consequently, these 'umbrella organizations' can have a significant influence on school evaluation, as they may support their schools by offering student assessments. Of course, schools are free to choose whether or not they use these offered curricula or tests, but in reality many of them do (Shewbridge et al., 2011). In some cases, schools can be obliged to accept support from the PBD and this is where the 'triangle of quality' becomes visible: when a school receives a negative report from the Inspectorate, a school can be obliged – in case of insufficient policy-making capacity – to implement a successful improvement plan with the help of the Pedagogical Advisory Service (Shewbridge et al., 2011).

The 'triangle of quality' ascribes the major responsibility for providing good quality education to the schools. The decree of 2009 states that all schools are obliged to produce results for the centrally-set attainment targets and to account for their efforts to pursue the centrally-set developmental objectives (Vlaamse Regering, 2009). The educational freedom of schools, which is described in the Constitution, was recently further emphasized in an educational policy document for the period of 2014-2019. The document emphasized that the government decides *what* schools have to achieve without specifying *how* to achieve the attainment targets. Schools do have autonomy with regard to their assessment policy and practices – this is part of the '*how'* (Crevits, 2014)*.

Given that schools obtain autonomy, the regulations set by the government for schools concerning their assessment organization are minimal. At the level of the school, minimal guidelines should be followed:

- In **elementary education** the agreements concerning pupils' evaluation and assessment should be communicated to the parents through the school regulations. The agreement of the parents to these regulations concerning pupils' evaluation and assessment is a prerequisite for enrolling their child.
- The government regulates the procedure for the assignment of a primary school certificate. At the end of elementary education the *class council* autonomously decides whether a certificate of elementary education will be issued or not. The class council judges whether the pupil has

sufficiently reached the objectives developed by the umbrella organization, which aim to meet the centrally set attainment targets and development objectives. In the case that a pupil does not receive the certificate, he/she is entitled to a written motivation of this decision and working points for his/her further school career. Parents have the possibility to appeal this decision of the class council.

- During elementary education, pupils are assessed by means of tests, exams or continuous assessment organized by the individual teacher under the final responsibility of the school. However, the final assessment determining the pupil's progression is a joint decision by the class deliberating council.

Corresponding with the regulations in elementary education, secondary schools also obtain autonomy when it comes to assessment organization. Consequently, the regulations concerning the organization of schools' assessment policy and practices are limited to the following:

- Schools have a class council which acts as the central assessment body in secondary education. There are three types of class councils: guidance class council, deliberation class council and admission class council. The *guidance class council* consists of the school's headmaster and all the members of the teaching staff who teach a particular pupil in a particular grade. They may be assisted by the members of staff providing psycho-social or pedagogical counseling. These people have an advisory voice. The guidance class council has the delicate task of carrying out an evaluation of each pupil at regular intervals and of drawing its pedagogical conclusions from these evaluations. In this respect, the guidance class council has the responsibility to follow up the progression of a pupil and to suggest help and, if necessary, to support a (re)orientation. The *deliberating class council* consists of the same members as the guidance class council, but its task is different: at the end of the school year they decide whether or not the pupil progresses to the next grade. The decision results in a certification with three options: an A certificate if the pupil has completed the grade successfully and can move on to the next grade; a B certificate if the pupil is admitted to the next grade but certain branches of education and/or courses of study are excluded; a C certificate which means that the pupil must repeat another year in the same grade. In contrast to the guidance and deliberation class councils, the *admissions class council* does not in all cases have to be composed of all the pupil's subject teachers. The admissions class council decides whether a pupil is allowed to enroll in a certain educational track.

- There are no mandatory rules and regulations that the class council needs to follow in deciding how pupils progress in their school careers. The decisions are generally made taking the

following factors into account: pupils' previous school career; the interim results of daily work, tests and examinations; the information from the Pupil Guidance Centre, and if necessary, discussions with parents and pupils.

- The class council's decision may also be deferred and may be made conditional to the pupil passing a re-examination.

- The maximum number of days that may be spent on the assessment of pupils has been capped at 30 (= 60 half days); schools that have introduced the system of continuous assessments and education forms 1, 2 and 3 of special secondary education cannot spend more than 9 (= 18 half) days on the assessment of pupils. The integrated test and the skills test, imposed by the authorities other than the school for certain technical and vocational courses of study (e.g. the exam for lorry drivers), are not included in the maximum number of assessment days.

6.2. Assessment policies and practices in Flanders

According to the constitutional '*freedom of education'*, only a few regulations are set up for schools concerning their assessment policies and practices. In view of this freedom, a wide variety in practices could be expected. In this section, the current state of affairs concerning assessment practices will be reported by giving an overview of findings from the Flemish Inspectorate and other, unfortunately less recent reports. For all that, the following section will describe which incentives are offered by the government and the *umbrella organizations* to support schools in their assessment policy and practices.

6.2.1. Inventory of (standardized) assessments in Flemish education

There is a wide variety of forms and types of assessments: assessment tasks, simulations, portfolios, performance assessments. Besides the different forms and types, there is a wide range of origins of tests and assessments: generally teachers develop the assessments and tests themselves, but teachers and schools can also choose to make use of ready-made assessments or to participate in tests which are conducted by umbrella organizations. At this point it becomes important to take into account the purpose of assessment: a lot of assessments are developed not only to give an overview of student outcomes, but also to stimulate schools' internal quality control in achieving the attainment targets.

Table 3 presents a summary of the assessments available in Flemish compulsory education categorized by the level to which they deliver useful information. The following paragraphs outline for which levels and purposes the tests and assessments are intended.

	Test
Macro	– Peiling
Meso	– Parallel – Pupil Monitoring System (LVS) – School Feedback Project – Toolkit Alternative Assessment
Micro	– Pupil Monitoring System (LVS) – School Feedback Project – SALTO – Toolkit Alternative Assessment – Tests and assessments provided by educational publishers – Tests and assessments developed by teachers

Table 3: Summary of available assessments in Flemish compulsory education.

Nationwide standardized testing or examinations of all pupils at certain moments in their educational career are not mandated in Flanders, but information on student outcomes is available from a number of externally designed tests available to help schools measure their outcomes. The inventory consists of tests provided by the National Assessment Programme, other nationally developed tests and tests developed by umbrella organizations.

The national assessment program in Flanders was introduced in 2002 and aims to provideinsight into how well the attainment targets have been implemented at the level of the Flemish school system. Students in a representative sample of schools are assessed and their schools receive feedback on their students' performance, which can inform school self-evaluation activities. Schools in Flanders that are not selected for the representative sample for the *peiling* can decide to use a parallel version of the

peiling (Vanhoof et al., 2016). These *parallel tests* are similar to the test used for the national assessment program and are offered by the Ministry of Education and Training's website 'Tests for Schools'. Schools conducting these *parallel tests* receive feedback about how many students have reached attainment targets, contextual added results in relation to background characteristics, comparative data from similar schools in Flanders, as well as comparative information from the *peiling* representative sample.

On the website 'tests for schools' (http://onderwijs.vlaanderen.be/nl/toetsen-voor-scholen), a pupil-monitoring system (LVS, leerlingvolgsysteem voor Vlaanderen) is offered for all registered primary schools. Schools can use this to monitor student progress in the Dutch language and mathematics skills at different stages of their primary education. A pupil-monitoring system has a twofold aim. First, at the level of the teacher, the learning progress and learning results of an individual student can be used to adjust the educational approach and process. Second, at the level of the school, the results of a pupil monitoring system can be used to coordinate a systematic approach to improve their quality of education. In other words, a pupil-monitoring system can be used as an instrument for a school's self-evaluation (Muys, 2016).

Another example of tests that have the aim to manage the internal quality of a school is provided by three universities cooperating in the 'School Feedback Project'. This project has the ambition to provide feedback to schools concerning three areas: pupils' learning outcomes, the added value of schools and the effect of innovations on students' learning outcomes. The School Feedback Project provides tests for primary and secondary education. In primary education there are tests to assess pupils in the subjects of mathematics, comprehensive reading, spelling and technical reading. The tests are conducted both at the beginning and at the end of a grade in primary education. In secondary education, tests are provided for mathematics and comprehensive reading in every grade except for grade 5 and 6 (Schoolfeedbackproject, 2015).

There is also a tool to monitor children's Dutch language ability at the start of primary education (SALTO). The tool is used as an instrument to screen the academic language proficiency of incoming pupils so as to identity those with a language deficiency and assess the extent of the deficiency. There is no feedback report offered for the SALTO test, but schools can calculate where their school performs in relation to other schools on the basis of norm data (Ramaut, Roppe, Verhelst, & Heymans, 2008).

Furthermore, concerning the assessment of the language competences for Dutch, the Flemish Ministry of Education provides a toolkit that supports schools in establishing an assessment program for the language competences for Dutch in primary and secondary education. Initially, the Ministry of

Education wanted to introduce a compulsory standardized language test at the beginning of secondary education in order to identify students at risk of failing academically as well as to raise the language proficiency of every student. An expert committee that was consulted on this issue advised against the idea of one standardized test, which is in line with the aforementioned absence of support in the educational field for an introduction of standardized tests (Van Avermaet, Pulinx, & Mondt, 2011). Instead, the experts recommended to develop an assessment toolkit that would support schools in setting up a language assessment policy. The Toolkit Alternative Assessment brings together all the language competences that students need to function academically, as mentioned in the attainment targets and developing objectives. Furthermore, all language assessment instruments (observation tools, tests, ...) available in Flemish education were inventoried and placed within the framework of language competences (Philips, I. et al., 2013).

The umbrella organizations *'Katholiek Onderwijs Vlaanderen'* and *'Onderwijs voor Vlaamse Steden en Gemeenten (OVSG)'* provide standardized tests to these schools following their specific curriculum. These tests have the aim of collecting the results of pupils' performance to gather information about the quality of education at that school. Similar to the *peiling tests,* the aim of the tests of the umbrella organizations is to support internal quality assurance with respect to the successful implementation of the curriculum set by the umbrella organization, instead of the curriculum set by the attainment targets and developing objectives (Vanhoof et al., 2016).

The tests of '*Katholiek Onderwijs Vlaanderen',* also referred to as '*interdiocesane toetsen'*, are classical paper-and-pencil tests conducted in grade 4 and/or in grade 6 of primary education for the subjects mathematics, Dutch and 'world orientation'. The *OVSG-tests* are administered at the end of grade 6 and consist of written tests about all subjects in primary education: mathematics, Dutch, French, and *world orientation*. Beside the written tests, there are six performance tests in the following subjects: Dutch and French speaking and listening, physical education, technology, artistic education and road safety (Vanhoof et al., 2016). The combination of written tests and performance tests creates a variation in what gets assessed: both knowledge and competences. It is important to notice that the OVSG-tests measure the attainment of the curriculum set by the OVSG-umbrella organization, but the link with the centrally set attainment targets and developing objectives is also made clear by referring to these attainment targets. The similarities between the goals of the curriculum and the centrally set objectives are made explicit in these tests.

Furthermore, the commercial world is responding to the need of schools to manage internal quality. Educational publishers, for instance, now provide teaching material for teachers and pupils that is in alignment with the centrally set attainment targets and developing objectives. The same educational

publishers also provide tests and assessment material in the packages they offer, responding to the (international) trend towards more standardized tests. For example: publishing office 'Van In' has launched a digital learning platform for primary (Bingel) and secondary (Diddit) education. Both platforms provide adaptive tests whereby both underachievers and top-achievers receive exercises adapted to the degree of difficulty they can handle. Another publishing office 'Abimo' is renowned for its normed tests with regard to technical reading (Vanhoof et al., 2016).

6.2.2. Schools' assessment policies and teachers' assessment practices

Research about assessment practices in primary and secondary education in Flanders is rather scarce. Previous research dates from the beginning of 2000. To gain insights in the current situation, we rely on the reports of the Flemish Inspectorate.

Previous research from 2002, conducted in secondary education, has shown that pupils are generally evaluated through written tests and examinations. This study reveals that the testing culture has long been the dominating paradigm in Flemish education, in practice the assessment culture was delayed as only 18% of the teachers insisted to assess regularly the process of pupils (Verhoeven, Devos, Bruylant, & Warmoes, 2002).

The annual reports of the Flemish inspectorate reveal a lack of alignment between the assessment policy and the practices of schools. The report of 2006 shows that in elementary education the developing objectives and attainment targets set by the government are seen as a source for information or inspiration rather than as a framework to keep an overview on educational and learning processes. With regard to the development of evaluation or assessment instruments, schools do not accommodate them in alignment with the centrally set curriculum. Instead, they tend to base them on the curriculum set by their *umbrella organization*, which in turn is based on the centrally set objectives (Onderwijsinspectie, 2006). As a result, there is evidence that schools set higher expectations for students than is defined by the centrally set objectives and targets, because the *umbrella organizations* extend and amplify the minimum curriculum.

In addition, it appears that in elementary education, assessment practices focus much more on attainment targets and developmental goals, which are related to a specific subject. Cross-curricular attainment targets and developmental goals which focus on attitudes receive much less attention in assessment. Schools do register assessment information at the level of pupils and classes, but the collection of information at the school level is rather exceptional. As a result, few schools attempt to

use the information as part of their internal quality assurance (Onderwijsinspectie, 2006). An analysis of the results of the Flemish Inspectorate dating from 2003 to 2006 reveals that the assessment practices of pupils in secondary education is characterized by some deficiencies. Assessment is considered as a teacher-related matter whereby a shared vision within the school and an equal orientation among teachers even in the same department is noticeably absent. There is a striking contrast between the amount of time that goes to the assessment of pupils and its return for the educational process: assessment seems to be perceived as part of the educational and learning process only on a very limited level (Onderwijsinspectie, 2006). Little attention is given to the alignment of assessment practices with the pupils' profiles. A survey conducted by the Flemish Inspectorate concerning the competences of novice teachers revealed that they consider building up appropriate assessment and evaluation as particularly complex and difficult to realize (Onderwijsinspectie, 2007)

The recent reports of 2015 and 2016 confirm earlier findings concerning weaknesses in evaluation and assessment practices in both primary and secondary education (Onderwijsinspectie, 2015). Assessment practices being in line with the school's vision on assessment is only the case in 50% of the schools. Only 26% of the schools have an efficient assessment practice, efficient meaning that the assessments measure whether pupils achieve the attainment targets (according to the expected level) and whether teachers know how to measure this appropriately. Schools do not engage in any reflection with respect to their assessment policy and practices; consequently the information gained from assessment is not utilized to make possible improvements to their practices (Onderwijsinspectie, 2015).

In 2015 assessment practices in the region of Brussels were the focus of a study carried out by the Flemish Inspectorate. This research revealed some obstacles that may be generalized for the region of Flanders. In primary education it seems that the assessment practices of language competences are inadequate, in particular for reading and writing competences. A similar difficulty occurs in the field of *world orientation,* where the focus is too much on the reproduction of knowledge. The assessment of attitudes and competences is too limited, and therefore it is difficult to get insight into the talents and the progress of pupils. The cause of these inadequate assessment practices lies in the limited knowledge of the curriculum and the various options for alternative assessment (Onderwijsinspectie, 2015). At the level of the school and the teachers, the report shows that the results and information gathered by assessments are hardly a stimulus to reflect on the assessment practices (Onderwijsinspectie, 2015). In secondary education, similar findings have been observed. A positive evolution is noticed at the level of assessment policy in secondary education: school leaders show willingness to change their policy and to reflect on the results and information of assessments, but also

indicate that the task of adjusting their educational processes and their assessment practices is very complex and time consuming. One third of the schools have difficulties in finding the balance between creating learning opportunities for pupils and checking via valid assessments if these pupils attain the targets (Onderwijsinspectie, 2015). In secondary education, the focus of assessment practices is often on the reproduction of knowledge, whereas during apprenticeships and integrated assignments the evaluation of the process and the different levels of competences get more attention (Onderwijsinspectie, 2012).

Besides the reports from the Flemish Inspectorate, other research confirms the absence of a shared, written vision about assessment. Looking at the schoollevel, the majority of schools in Flemish secondary education do not have explicit agreements concerning assessment (Struyf, 2000). At the same time, research from Verhoeven et al. (2002) indicates that teachers as well as principals in Flemish secondary education are convinced that there is an assessment policy in their school (Verhoeven et al., 2002). In practice, the assessment policy of a school consists mainly of non-committal, unwritten and implicit agreements (Struyf, 2000). This reveals that the nature of assessment policy must be derived from a study of the practices of teachers, principals and all stakeholders. Even if there is a formal assessment policy , its effect on assessment practices is neither guaranteed, nor consistent.

6.3.   The impact of the different contexts on the assessment practices

The previous chapters have described how assessment, and the broader sphere of educators' beliefs about assessment, the assessment practices and the assessment policies of schools, cannot be seen in a vacuum. Assessment is embedded in multiple contexts which is represented in the model with circles: the educational, socio-cultural and political-economic contexts.

As described earlier, these contexts are interacting with each other and are subject to reciprocal influences. This model of interacting circles will be used as a leading framework to research assessment policy and practices in Flemish education, which is characterized by increasing diversity. Little is known yet about how these practices are related to
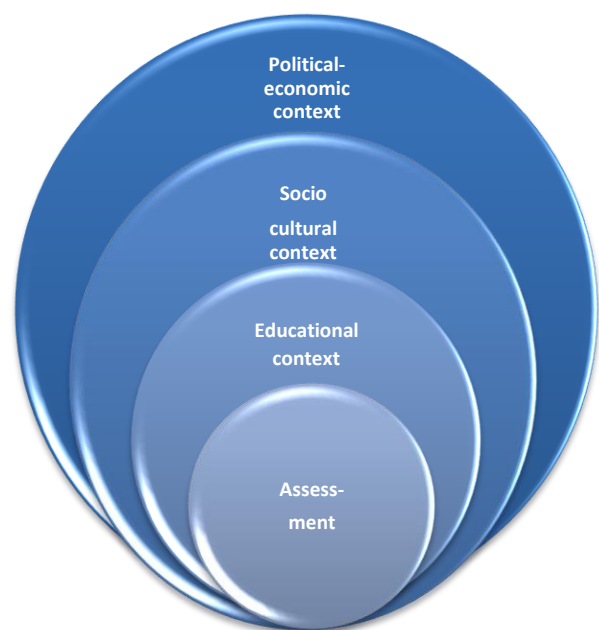
and being modified vis-à-vis the increasing diversity. The changing educational context challenges teachers and principals to move towards assessment policies and practices taking into account a diverse population of pupils. These changes demand new competencies of educational practitioners. The question is how different stakeholders in education act and handle in this changing educational context and to what extent this has an impact on assessment policy and practices. Moreover, because of the constitutionalized freedom of education, a wide variety in practices can be expected. In order to obtain an overview of the current policy and practices concerning assessment in Flemish education, the following research questions will be central:

1. What do assessment policies and practices look like in Flemish compulsory education?
2. What factors determine assessment policies and practices and which processes are influencing them?
3. Which beliefs, system features and cultural features have an impact on schools' assessment policies and practices?
4. Can a typology in practices and policies concerning assessment be identified?

These research questions will be answered taking into account the educational, socio-cultural and political-economic context. It is important to note is that Flemish education has no tradition in centrally administered high-stakes standardized tests. Even low-stakes tests have never been part of the system. In comparison to all other Western countries, Flanders occupies a unique position when it comes to standardized testing. As standardized testing leads to teaching-to-the-test and to a narrowing of the curriculum, assessment practices of teachers and, more widely assessment policies of schools, are developing in convergence, all striving for high achievement on the tests. Education in Flanders, on the contrary, does not suffer the pressure of nationwide tests. Moreover, the regulations for testing in Flanders are minimal, which is reinforced by the constitutional freedom of education. The role of political context is rather minimal, in comparison to those countries where standardized tests are used. One can expect to find a wide range of variety in assessment policies and practices just because of this freedom and the absence of nationwide tests. In contrast to the convergence of assessment practices in countries with nationwide tests, the overall assessment policies and practices in Flanders can be characterized by divergence because of the freedom in which they can act.

This divergence can be explained by the fact that assessment practices and policies are also embedded in a socio-cultural and educational context. Depending on the processes and their reciprocal influences, differing assessment patterns can be shaped by the specific features of a school, for example, the its composition or its pedagogical vision on learning and consequently on assessment.

Bibliography

Agodi. (2017). Cijfermateriaal- Onthaalonderwijs anderstalige nieuwkomers. Retrieved 16 february, 2017, from http://www.agodi.be/cijfermateriaal-onthaalonderwijs-anderstalige-nieuwkomers

Allal, L., & Lopez, L. M. (2005). Formative assessment: Improving learning in secondary classrooms In J. Looney (Ed.), *Formative assessment of learning: A review of publications in French* (pp. 241-264). Paris, France: Organisation for Economic Cooperation and Development.

Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., & Young, D. B. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education, 21*, 315-334.

Banks, J., Cochran-Smith, M., Moll, L., Richert, A., Zeichner, K., LePage, P., Darling-Hammond, L., Duffy, H., & McDonals, M. (2007). Teaching Diverse Learners. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing Teachers for a Changing World*. San Fransisco.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Birenbaum, M. (2014). Conceptualizing assessment culture in school. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *The enabling power of assessment.* (Vol. 1, pp. 285-302). Dordrecht: Springer.

Birenbaum, M. (2016). Assessment Culture Versus Testing Culture: The Impact on Assessment for Learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation.* Dordrecht: Springer.

Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation, 37*(1), 35-48.

Black, P. (2013). Formative and summative aspects of assessment: theoretical and research foundations in the context of pedagogy. In J. H. Mc Millan (Ed.), *Research on Classroom Assessment*. London: Sage Publications.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bonner, S. M. (2016). Teachers' Perceptions About Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. London: Routledge.

Boxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment & Evaluation in Higher Education, 35*(3), 291-300.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education, 11*(1), 7-26.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education, 8*(2), 153--169.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers college record, 106*(3), 429-458.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. Mc Millan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43-62). New York: Teachers College Press.

Cakici, D. (2016). The Correlation among EFL Learners' Test Anxiety, Foreign Language Anxiety and Language Achievement. *English Language Teaching, 9*(8), 190-203.

Campbell, C. (2013). Research on Teacher Competency in Classroom Assessment. In J. H. Mc Millan (Ed.), *Handbook of Research on Classroom Assessment* (pp. 71-84). London: SAGE Pulications.

Cauley, K. M., & Mc Millan, J. H. (2009). Formative assessment techniques to support student motivation and achievement. *Clearing House, 83*(1), 1-6.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(1), 1-6.

CGVS. (2015). Asielstatistieken: Jaaroverzicht 2015. Retrieved 12 February, 2017, from http://www.cgvs.be/nl/actueel/asielstatistieken-overzicht-2015

Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-44.

Clark, I. (2010). Formative assessment:'There is nothing so practical as a good theory'. *Australian Journal of Education, 54*(3), 341-352.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249.

Cochren-Smith, M. (2014). *A U.S. Persepective on Teacher Preparation Research: Past, Present and Future.* Paper presented at the ECER 2014, Porto.

Crevits, H. (2014). *Beleidsnota 2014-2019 onderwijs*.

Crooks, T. J. (1987). The impact of classroom evaluation practices on students. *Review of educational research, 58*(4), 438-481.

De Backer, F., Van Avermaet, P., & Slembrouck, S. (2016). Schools as laboratories for exploring multilingual assessment policies and practices. *Language and Education*, 1-14.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*(4), 307-329.

Earl, L. M. (2003). *Assessment As Learning: Using Classroom Assessment to Maximize Student Learning.* London: Sage Publications.

Ercikan, K., Roth, W. M., & Asil, M. (2015). Cautions about Inferences from International Assessments: The Case of PISA 2009. *Teachers college record, 117*(1), n1.

Ercikan, K., & Solano-Flores, G. (2016). Section Discussion: Asessment and Sociocultural Context: A Bidirectional Relationship. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment.* London: Routledge.

Falchikov, N. (2005). *Improving assessment through student involvement: practical solutions for aiding learning in higher and further education.* London: RoutledgeFalmer.

Foucault, M. (1979). *Discipline and Punish*. New York: Vintage Books.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics, 18*(3), 402-423.

Fullan, M. (2011). *Choosing the wrong drivers for whole system reform*: Centre for Strategic Education Melbourne.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the Fidelity of Implementing Embedded Formative Assessments and Its Relation to Student Learning. *Applied Measurement in Education, 21*(4), 360-389.

Green, S. K., & Gredler, M. E. (2002). A review and analysis of constructivism for school-based practice. *School Psychology Review, 31*(1), 53.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213-224.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Hayward, L., & Spencer, E. (2010). The complexities of change: formative assessment in Scotland. *Curriculum Journal, 21*(2), 161-177.

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.

Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395-420.

Ikeda, M. (2011). When students repeat grades or are transferred out of school: What does it mean for education systems? *PISA in focus No. 6.* Retrieved 12 February, 2017 from http://www.oecd.org/dataoecd/35/58/48363440.pdf

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 47-60). London: Sage.

Kennedy, K. (2013). *High stakes school based assessment and cultural values: Beyond issues of validity. .* Paper presented at the Cambridghe Horizon's Seminar, 'School-based Assessment: Prospects and Realities in Asian Contexts'. Kuala Lumpur, Malaysia. http://www.cambridgeassessment.org.uk/Images/139719-sba-seminar-papers.pdf

Kennedy, K. (2016). Exploring the Influence of Culture on Assessment. The case of Teachers' Conceptions of Assessment in Confucian Heritage Cultures. . In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. London: Routledge.

Kind & Gezin. (2013). Het kind in Vlaanderen. Brussel: Kind & Gezin. Retrieved 12 February, 2017 from http://www.kindengezin.be/img/kind-in-Vlaanderen-2013.pdf

Kingston, N., & Nash, B. (2011). Formative Assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*, 28-37.

Klein, E. D. (2013). *Statewide Exit Exams, Governance, and School Development. An International Comparison.* (Vol. 18). Münster: Waxmann.

Mc Millan, J. H. (2013). Why we need research on Classroom Assessment. In J. H. Mc Millan (Ed.), *Research on Classroom Assessment*. London: SAGE Publications.

McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing, 18*(4), 333-349.

McNamee, G. D., & Chen, J. Q. (2005). Dissolving the Line between Assessment and Teaching. *Educational Leadership, 63*(3), 72-76.

Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal, 30*(2), 521-546.

Menken, K. (2008). *English learners left behind. Standardized testing as language policy.* Clevedon, Buffalo, Toronto: Multilingual Matters.

Menken, K. (2010). NCLB and English Language Learners: Challenges and Consequences. *Theory Into Practice, 49*(2), 121-128.

Mertens, C., Ysenbaert, J., Vanderlinde, R., & Van Avermaet, P. (2015). Evaluatie van de engagementsverklaring: bevraging van relevante stakeholders. *Tijdschrift voor onderwijsrecht en onderwijsbeleid, 3*, 15-27.

Muys, M. (2016). LVS Leerlingvolgsysteem CITO-Vlaamse versie. Retrieved 10 november, 2016, from http://www.ond.vlaanderen.be/toetsenvoorscholen/LVS/index.asp?

Nemati, A. (2012). On the dimensions of test anxiety and foreign language learners. *International Journal of English and Literature, 3*(4), 97-102.

Nuffield Foundation. (2017). The assessment reform group. Retrieved 01 January, 2017, from http://www.nuffieldfoundation.org/assessment-reform-group

Nicaise, I.,Spruyt, B., Van Houtte, M. & Kavadias, D. (2014). Het onderwijsdebat: een tussenstand. In I. Nicaise, B. Spruyt, M. Van Houtte, & D. Kavadias (Eds.), *Het onderwijsdebat: waarom de hervorming van het secundair broodnodig is.* Berchem: EPO.

Nicol, D. J., & Mcfarlane-Dick, D. (2006). Formative assessment and selfregulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199-218.

Noam, G. (1996). Assessment at a Crossroads: Conversation. *Harvard Educational Review, 66*, 631-657.

Noppe, J., & Lodewijckx, E. (2013). De Gekleurde Samenleving. Personenen van Vreemde Herkomst in Vlaanderen. Brussel: Studiedienst Vlaamse Regering.

OECD. (2013). PISA 2012 results in Focus. What 15-year-olds know and what they can do with what they know. Organisation for Economic Co-operation and Development (OECD).

Onderwijsinspectie. (2006). Onderwijsspiegel: Verslag over de toestand van het onderwijs schooljaar 2005-2006. Brussel: Vlaams Ministerie van Onderwijs en Vorming

Onderwijsinspectie. (2007). Onderwijsspiegel 2006-2007. Brussel: Vlaams Ministerie van Onderwijs en Vorming.

Onderwijsinspectie. (2012). *Onderwijsspiegel 2012*. Brussel: Vlaams Ministerie van onderwijs en Vorming.

Onderwijsinspectie. (2015). Onderwijsspiegel 2015: jaarlijks rapport van de onderwijsinspectie. Brussel: Vlaams Ministerie van Onderwijs en Vorming.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual review of psychology, 49*(1), 345-375.

Pattalitan Jr, A. P. (2016). The Implications of Learning Theories to Assessment and Instructional Scaffolding Techniques. *American Journal of Educational Research, 4*(9), 695-700.

Penninckx, M., Vanhoof, J., & Van Petegem, P. (2011). *Evaluatie in het Vlaamse onderwijs. Beleid en praktijk van leerling tot overheid.* Antwerpen-Apeldoorn: Garant.

Philips, I., Seghers, M., Versteden, P., & Ysenbaert, J. (2013).Toolkit Competenties Nederlands Breed Evalueren Lager Onderwijs. Gent: Steunpunt Diversiteit & Leren, Leuven: Centrum voor Taal en Onderwijs, Brussel: Vlaams Ministerie van Onderwijs en Vorming. Retrieved 10 November 2016 from https://onderwijs.vlaanderen.be/sites/default/files/atoms/files/Toolkit%20breed%20evalueren%20competenties%20Nederlands%20in%20het%20lager%20onderwijs.pdf

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review Of Education, 34*(1), 1-20.

Ramaut, G., Roppe, S., Verhelst, M., & Heymans, R. (2008). SALTO Screeningsinstrument Aanvang Lager Onderwijs Taalvaardigheid. Retrieved 10 november 2016, from http://www.ond.vlaanderen.be/toetsenvoorscholen/taaltoets/index.asp?uid=&pwd=

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics, 16*(2), 163-188.

Ruiz-Primo, M. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation, 37*(1), 15-24.

Schneider, C. M., Egan, K. L., & Julian, M. W. (2013). Classroom Assessment in the Context of High-Stakes Testing. In McMillan (Ed.), *Handbook of Research on Classroom Assessment*. London: Sage.

Schoolfeedbackproject. (2015). Schoolfeedback. Retrieved 10 november 2016, from http://www.schoolfeedback.be/schoolfeedback/

Schraepen, B., Lebeer, J., Vanpeperstraete, L., & Hancké, T. (2010). Draagkracht ten aanzien van diversiteit en inclusief onderwijs voor lagere scholen in de provincie Antwerpen. Antwerpen: Plantijn Hogeschool.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally Shavelson R.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E., Ruiz-Primo, M., & Yue, Y. (2008). On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers. *Applied Measurement in Education, 21*(4), 295-314.

Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher, 29*(7), 4-14.

Shewbridge, C., Hulshof, M., Nusche, D., & Stoll, L. (2011). School Evaluation in the Flemish Community of Belgium *OECD Reviews of Evaluation and Assessment in Education* (pp. 139): OECD.

Shohamy, E. (2001). *The Power of Tests. A critical perspective on the uses of language tests.* Harlow, England: Longman.

Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen. Den Haag: NWO-PROO.

Standaert, R. (2008). *Globalisering van onderwijs in contexten.* Leuven: Acco.

Standaert, R. (2014). *De becijferde school. Meetcultus en meetcultuur.* Leuven: Acco.

Stevens, P. A. (2007). Exploring the importance of teachers' institutional structure on the development of teachers' standards of assessment in Belgium. *Sociology of Education, 80*(4), 314-329.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758-765.

Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. . *Phi Delta Kappan, 87*, 324-328.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*: Routledge.

Struyf, E. (2000). *Evalueren: een leerkans voor leraren en leerlingen?* (Vol. 26). Leuven: Acco.

Taras, M. (2009). Summative assessment: the missing link for formative assessment. J*ournal Of Further and Higher Education, 33*(1), 57-69.

Teltemann, J., & Klieme, E. (2016). The Impact of International Testing Projects on Policy and Practice. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. London: Routledge.

Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational research review, 9*, 1-15.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom.* Buckingham: Open University Press.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631.

UN. (2007). Convention on the Rights of Persons with Disabilities. Retrieved 10 november 2016 from https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html

Valcke, M. (2008). *Onderwijskunde als ontwerpwetenschap.* Gent: Academia Press.

Van Avermaet, P., Pulinx, R., & Mondt, K. (2011). Bruikbaarheidsstudie screeningsinstrument Nederlands aanvang secundair onderwijs. Gent: Universiteit Gent.

Van Avermaet, P., & Sierens, S. (2010). Diversiteit is de norm. Er mee leren omgaan de uitdaging. Een referentiekader voor omgaan met diversiteit in onderwijs. *De Coen, D. ea (Red.), Handboek beleidsvoerend vermogen (Doelgerichte visie, Aflevering 4)*, 1-48.

Van Praag, L., Boone, S., Stevens, P., & Van Houtte, M. (2015). De paradox van het watervalsysteem: wanneer het groeperen van studenten in homogene groepen tot meer heterogeniteit leidt in het beroepsonderwijs. . *Sociologos*.

Vanhoof, J., De Maeyer, S., Van Petegem, P., Penninckx, M., & Quintelier, A. (2016). Scenario's voor leer(winst)monitoring in Vlaanderen: Een ontwerponderzoek naar haalbaarheid en wenselijkheid (pp. 521). Brussel: Vlaams Ministerie van Onderwijs en Vorming

Verhaeghe, P. P., Van der Bracht, K., & Van de Putte, B. (2012). *Migrant zkt toekomst: Gent op een keerpunt tussen oude en nieuwe migratie.* Antwerpen: Garant.

Verhoeven, C., Devos, G., Bruylant, W., & Warmoes, V. (2002). *Evalueren in Vlaamse secundaire scholen. Een onderzoek bij leraren en directeurs.* Antwerpen-Apeldoorn: Garant.

Vlaamse Regering (2009). *Decreet betreffende de kwaliteit van onderwijs*. Brussel.

Vlaamse Regering. (2013). *Masterplan hervorming S.O.* Brussel. Retrieved 10 november 2016 from http://onderwijs.vlaanderen.be/sites/default/files/atoms/files/Masterplan-hervorming-secundair.pdf.

Vlaams Ministerie voor Onderwijs en Vorming (2017). Peilingen - Algemene informatie - Wat en waarom? Retrieved 16 February, 2017, from http://www.ond.vlaanderen.be/curriculum/peilingen/algemeen/index.htm#wat-peilingen

Vlaams Ministerie voor Onderwijs en Vorming (2016). Mijn school is OK: ontdek hier het referentiekader voor onderwijskwaliteit. Retrieved 16 February, 2017, from http://mijnschoolisok.be/

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-64.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education, 45*(4), 477-501.