

ZICHT OP LEERWINST

DEEL 1 INLEIDEND

1 AANLEIDING EN SITUERING

In verschillende onderwijssystemen woedt een aanhoudende discussie over de implementatie van systemen voor het monitoren van leerresultaten van leerlingen. Hoewel sommige landen doelbewust afzien van dergelijke initiatieven is toch een groeiende tendens waar te nemen van het aantal landen dat wél stappen in die richting onderneemt of waar het gestandaardiseerd in kaart brengen van leerresultaten en leerwinst van leerlingen reeds gemeengoed is geworden. Bovendien worden aan de uitkomsten van gestandaardiseerde toetsen steeds meer gevolgen gekoppeld (OECD, 2013). Naast het uitreiken van diploma's, die vaak de enige mogelijkheid zijn om leerlingen toegang tot hoger onderwijs te verschaffen, zijn in diverse regio's en landen een aantal andere consequenties verbonden aan de resultaten van een leer(winst)monitoringssysteem. Zo kunnen toetsresultaten de basis vormen voor het evalueren (en borgen of verbeteren) van de kwaliteit van het onderwijs, zowel op klas- en schoolniveau als op landelijk niveau (Cizek, 2005; Loeb, 2013). Deze resultaten kunnen ook leiden tot het opstellen van rankings van scholen, waarbij de toets gebruikt wordt als meetinstrument voor de kwaliteit van de scholen (de Wolf & Janssens, 2007; Dill & Soo, 2005). In sommige landen worden de uitkomsten van een leer(winst)monitoringssysteem eveneens gebruikt om leraren te evalueren en worden er financiële stimuli toegekend aan goed presterende scholen en/of leraren (Baker et al., 2010; Loeb, 2013). In de meest uiteenlopende onderwijscontexten geven gestandaardiseerde toetsen die deel uitmaken van een leer(winst)monitoringssysteem aanleiding tot reeds decennialang aanslepende discussies (Linn, 2001; Wang, Beckett, & Brown, 2006). Het aanhouden van deze discussie wordt verklaard door de inbedding ervan in een ruimere '*clash of philosophical positions*', waarbij gestandaardiseerde toetsen de uiting zijn van twee tegengestelde visies op onderwijs en kwaliteit in het algemeen. Aan de ene zijde van het continuüm staat een visie die pleit voor competitie en het belonen van verdiensten en prestaties. Deze visie, waarin gestandaardiseerde toetsen een belangrijke plaats innemen, omarmt het idee van individuele verantwoordelijkheid en erkent verschillen tussen individuen. Aan de andere zijde van het continuüm staat een visie waarbij competitie wordt ontmoedigd, waarbij inzet en goede processen worden beloond, waarbij collectieve verantwoordelijkheid de voorkeur krijgt boven individuele verantwoordelijkheid en waarbij gestreefd wordt naar gelijkheid. Het betreft daarnaast een '*clash of interests*', waarbij aan de ene kant wordt gepleit voor een transparante meting van de prestatie van elke openbare dienst, terwijl aan de andere kant wordt beweerd dat meting en transparantie enkel tot onwenselijke neveneffecten leiden (Phelps, 2005a).

Door de sterke geladenheid van het thema is er internationaal al bijzonder veel gepubliceerd omtrent de (veronderstelde) effecten en neveneffecten van een leer(winst)monitoringssysteem of van losstaande gestandaardiseerde toetsen. Sinds de opkomst van de eerste studies, in de jaren dertig van de vorige eeuw, is er een continue stroom van publicaties geweest rond dit thema (Phelps, 2005b). Het betreft evenwel talrijke opiniërende publicaties en slechts zelden goed onderbouwde

empirische studies. Vaak wordt er door beleidsmakers veel aangenomen inzake effectiviteit van gestandaardiseerde toetsen en zijn de onderliggende assumpties (hoe moeten de verschillende toetsen in het leer(winst)monitoringssysteem leiden tot een verbetering van de onderwijskwaliteit?) slechts matig geëxpliciteerd (Skedsmo, 2011).

Ook in Vlaanderen is het idee van het opzetten van systemen die leerprestaties van leerlingen op een gestandaardiseerde manier in kaart brengen steeds gecontesteerd (Shewbridge, Hulshof, Nusche, & Stoll, 2011). Er gaan verschillende stemmen op van stakeholders die het wenselijk achten een (meer) systematisch en gestandaardiseerd zicht te krijgen op de leerresultaten van leerlingen in het leerplichtonderwijs, waarbij deze resultaten zowel op micro-, meso- én macroniveau kunnen dienen. Daar staat tegenover dat er ook stakeholders van mening zijn dat op bestaande (buitenlandse) systemen om leerresultaten te monitoren gefundeerde kritieken te geven zijn en dat ongewenste neveneffecten zich (kunnen) voordoen (Hirtt, Nicaise, & De Zutter, 2007; Standaert, 2014). De discussie in Vlaanderen moet gezien worden in de context van het grondwettelijk vastgelegde principe van 'vrijheid van onderwijs' waarin bepaald wordt dat scholen een grote vrijheid hebben om een eigen invulling aan onderwijskwaliteit te geven (gegeven de decretale bepalingen). De visie wordt dan ook vaak gehuldigd dat verschillen tussen scholen niet enkel moeten worden toegelaten, maar zelfs moeten worden aangemoedigd. Uit onderzoek in het kader van een voorbereidende evaluatieschets voor een review van de OESO over evaluatie in het Vlaamse onderwijs blijkt dat de argumenten die door verschillende tegenstanders van centrale examens werden genoemd erg talrijk zijn (Vlaams Ministerie van Onderwijs en Vorming, 2010). Het betreft pedagogisch/onderwijskundige bezwaren (bijvoorbeeld dat onderwijskwaliteit niet mag worden verengd tot meetbare output), normatieve bezwaren (bijvoorbeeld dat een examen een statisch beeld weergeeft van leerlingen, en dat rankings kunnen worden opgemaakt van scholen), bezwaren tegen de examens zelf (bijvoorbeeld over beperkingen inzake de metingsbetrouwbaarheid), en praktische bezwaren (zoals de organisatorische en financiële kost). Desalniettemin is er toch vanuit het onderwijsveld alsook vanuit de Vlaamse Scholierenkoepel vraag naar het aanreiken van meer gestandaardiseerde tools die helpen bij beslissingen over de loopbaan van de leerlingen (Penninckx, Vanhoof, & Van Petegem, 2011).

Ondertussen kan men niet om de vaststelling heen dat in het Vlaamse onderwijs steeds meer aandacht wordt geschonken aan de resultaten van de onderwijsprocessen (op micro-, meso- én macroniveau). De tendens naar het aanspreken van onderwijsaanbieders (leraren, scholen en intermediairen) op de wijze waarop ze de autonomie die ze genieten, succesvol weten aan te wenden, is daar uiteraard niet vreemd aan. Bovendien groeit bij diverse stakeholders de overtuiging dat door het creëren van synergieën inzake het monitoren van leerresultaten op micro-, meso- en macroniveau krachtige instrumenten voor het bewaken en ontwikkelen van (aspecten van) onderwijskwaliteit kunnen opgezet worden (OECD, 2013). In het Vlaamse Regeerakkoord staat onder meer het voornemen om van een soort 'toolkit' die kan begrepen worden als een aanzet tot een leer(winst)monitoringssysteem werk te maken:

“Op het einde van het basisonderwijs worden alle leerlingen getoetst. De doelstelling is zowel feedback te kunnen geven in het kader van de interne kwaliteitscontrole van de scholen als gegevens te verzamelen op systeemniveau. Met de onderwijsverstrekkers wordt overeengekomen om de huidige OVSG-toetsen en de interdiocesane toetsen te valideren in samenspraak met het Steunpunt

Toetsontwikkeling, dat de peilingsproeven en paralleltoetsen ontwikkelt. Op die manier kan een toets toolkit ontstaan. Met de realisatie van dezelfde doelstelling voor ogen voeren we gesprekken met de onderwijsverstrekkers van het secundair onderwijs en bekijken we hoe de bestaande peilingsproeven daartoe een basis kunnen vormen” (Vlaamse Regering, 2014, p.96).

Het monitoren van leerresultaten/leerwinst is vanuit diverse perspectieven bekeken een complexe aangelegenheid (zowel politiek-maatschappelijk als onderwijskundig en wetenschappelijk-methodologisch) (Janssens, Rekers-Mombarg, & Lacor, 2014). Als de Vlaamse overheid initiatieven tot monitoring van leerresultaten overweegt in te richten, te faciliteren of te stimuleren, dan is het aangewezen te leren uit bestaande (buitenlandse) systemen en uit bestaande literatuur. Heldere en gedragen doelen dienen vooropgesteld te worden, net zoals ook de haalbaarheid op voorhand moet worden afgetoetst. Bestaande monitoringssystemen zijn immers binnen een lokale traditie en cultuur gegroeid, waardoor hun determinanten niet automatisch gegeneraliseerd kunnen worden naar andere onderwijscontexten. Onderzoek naar de haalbaarheid en de mogelijkheden van leerwinstmeting in de Vlaamse context dringt zich dus op.

Deze studie onderbouwt en ontwikkelt mogelijke scenario's voor het monitoren van leerresultaten/leerwinst en gaat na of deze haalbaar en wenselijk zijn in de Vlaamse context. Het betreft dus een beleidsevaluatie ex ante met als doel de beleidsvoorbereiding wetenschappelijk te ondersteunen. De focus ligt op het leerplichtonderwijs (lager en secundair onderwijs) en op de leerdomeinen wiskunde, wetenschappen en Nederlands. De volgende onderzoeksvragen worden vooropgesteld:

- OV1 - Welke zijn (volgens Vlaamse stakeholders) de karakteristieken van een optimaal leer(winst)monitoringsstelsel in het Vlaamse leerplichtonderwijs?
- OV2 - Welke zijn de (volgens Vlaamse stakeholders) karakteristieken van een optimale conceptualisering en operationalisering van het concept 'leerwinst' in het Vlaamse leerplichtonderwijs?
- OV3 - Tot welke beleidsscenario's geven de ontwerpprincipes die uit OV1 en OV2 afgeleid kunnen worden aanleiding?
- OV4 - Hoe beoordelen Vlaamse stakeholders de haalbaarheid en wenselijkheid van deze beleidsscenario's?

2 BEGRIPSafbakening

Deze sectie heeft tot doel om enkele centrale concepten binnen dit boek af te bakenen. Zo worden de termen 'gestandaardiseerde toetsen', 'leerwinst' en 'toegevoegde waarde' kort besproken. Daarna worden functies en vormen van gebruik van gestandaardiseerde toetsen ingeleid.

2.1 Gestandaardiseerde toetsen

De focus van deze studie ligt op een leer(winst)monitoringsstelsel. Een leer(winst)-monitoringsstelsel bestaat per definitie uit één of meerdere gestandaardiseerde toetsen. De literatuur rond gestandaardiseerde toetsen richt zich zeer sterk op 'state tests': toetsen die door alle leerlingen op een van bovenaf opgelegd moment dienen te worden afgelegd. Vaak worden er ook

directe consequenties aan de resultaten van deze toetsen verbonden voor leerlingen, leraren of scholen.

De invulling van gestandaardiseerde toetsen in een leer(winst)monitoringsysteem, en dus ook in dit rapport, is evenwel breder. Volgens Sireci (2005, p. 113) betekent een gestandaardiseerde toets dat: *“the test content is equivalent across administrations and that the conditions under which the test is administered are the same for all the test takers”*.

Tabel 1 - Kenmerken en mate van standaardisatie

	Licht gestandaardiseerd	Medium gestandaardiseerd	Sterk gestandaardiseerd
Organisatie			
Verspreiding	Regionaal		Nationaal
Soort examen	Schriftelijk overheen hele gebied in aanvulling van door scholen ontwikkelde toetsen	Schriftelijke toetsen overheen het hele gebied, mondelinge toetsen door scholen	Schriftelijke en mondelinge toetsen overheen het hele gebied
Aantal leerdomeinen	1	4-5	Alle leerdomeinen
Belangrijkste bron van informatie	Toetsen van school meest belangrijk	Even belangrijk	Enkel eindexamens
Leerdomeinen	Enkel hoofdvakken	Hoofdvakken en specialisatie-vakken	Alle leerdomeinen
Keuze van leerdomeinen door leerlingen	Vrije keuze	Deels verplicht, deels vrije keuze	Alles verplicht
Keuze van moeilijkheidsgraad/ beheersingsniveau door leerlingen	Vrije keuze	Keuze automatisch gerelateerd aan keuze voor vakken	Geen keuze: moeilijkheidsgraad identiek voor alle leerlingen
Verbetering van toetsen			
Aantal fases in verbetering	1	2	3
Evaluatoren	Van de school zelf	Van andere scholen	Centrale commissie
Anonimiteit	Niet anoniem	Deels anoniem	Anoniem
Standaarden	Achteraf bepaald		Vooraf bepaald
Gebruik van de resultaten			
Aandeel in eindresultaat van de leerling	Minder dan 30%	50% of meer	100%
Publicatie van geaggregeerde resultaten op schoolniveau	Niet mogelijk	Mogelijk, maar wordt niet openbaar gemaakt	Openbaar gemaakt

Deze definitie houdt in dat een bepaalde inhoud overheen leerlingen en scholen dezelfde is en dat de toets in identieke condities door alle deelnemende leerlingen wordt afgelegd. Er zijn binnen deze

definitie geen vooropgezette bepalingen inzake verplichting of vrijwilligheid van afname noch inzake het doel van de toets, de consequenties, de frequentie van afname, de getoetste leerdomeinen, de leeftijd van leerlingen en de toetsvorm. Een gestandaardiseerde toets kan dus op zeer uiteenlopende manieren vorm worden gegeven.

Klein en Van Ackeren (2012) stelden een inventaris op om het concept gestandaardiseerde toetsen te operationaliseren en deelden toetsen op in 'licht gestandaardiseerd', 'medium gestandaardiseerd' en 'sterk gestandaardiseerd'. Om tot deze conceptualisering te komen, vertrokken zij vanuit drie kenmerken, namelijk (1) organisatorische kenmerken, (2) kenmerken inzake het verbeteren, en (3) kenmerken inzake het gebruik van de toetsresultaten (waarbij weliswaar vooral wordt uitgegaan van eerder summatieve eindtoetsen, die worden ingezet op een verantwoordingsgerichte manier op microniveau). Tabel 1 is een vertaling van de tabel uit deze studie (Klein & Van Ackeren, 2012, p. 182).

2.2 Leerwinst en toegevoegde waarde

De termen 'leerwinst' en 'toegevoegde waarde' (in de Angelsaksische literatuur worden termen als *learning gains*, *student growth* en *added value* aangewend) worden heel vaak door elkaar gebruikt. Uiteenlopende en soms zelfs tegenstrijdige invullingen - vooral wat de term 'added value' betreft - maken dat het vaak moeilijk is om correct te interpreteren wat door verschillende auteurs precies wordt bedoeld (Saunders, 1999).

In overeenstemming met een recent rapport over een pilootproject rond leerwinstmonitoring in Nederland (Janssens et al., 2014), wordt in deze studie over leerwinst gesproken wanneer het individuele leerlingen betreft (microniveau). De term toegevoegde waarde wordt gereserveerd voor de bijdrage van scholen aan het leren of aan de leerwinst van leerlingen. Toegevoegde waarde situeert zich dus op het mesoniveau (zie 2.3.2 voor een definitie van micro- en mesoniveau).

2.3 Uiteenlopende doelen van gestandaardiseerde toetsen

Een leer(winst)monitoringssysteem wordt vaak vanuit uiteenlopende doelen opgezet. Het gebruik van de resultaten van een leer(winst)monitoringssysteem kan gebeuren in ontwikkelingsgericht of een verantwoordingsgericht perspectief, en gericht zijn op micro-, meso- of macroniveau.

2.3.1 Ontwikkelingsgerichte en verantwoordingsgerichte doelen

Er kan onderscheid worden gemaakt tussen het ontwikkelingsgericht of verantwoordingsgericht perspectief wat het doel van een leer(winst)monitoringssysteem betreft. Overheden nemen immers initiatieven om scholen te ondersteunen bij hun ontwikkelingen, maar organiseren ook mechanismen waarbij scholen verantwoording dienen af te leggen. Dit kan op drie manieren: via onderzoek door experts (bijvoorbeeld inspectie), via administratieve monitoring of via gestandaardiseerde toetsen (Scheerens, 2014).

In het ontwikkelingsgericht perspectief dient een gestandaardiseerde toets een formatief doel. Of de resultaten ervan leiden tot consequenties hangt in de eerste plaats af van degene waarop de evaluatie betrokken is: een score van een leerling heeft in dit perspectief gevolgen die ofwel door de leerlingen zelf ofwel door de leraar na een dialoog met de betrokken leerling worden geïnitieerd.

Doelen in een verantwoordingsgericht perspectief zijn daarentegen gericht op het maken van een summatieve beoordeling. De resultaten van het leer(winst)monitoringssysteem leiden tot consequenties die worden opgelegd aan degene die het onderwerp is van de evaluatie (Vanhoof & Van Petegem, 2007; Wang et al., 2006).

In recente decennia is er een geleidelijke verschuiving in vele onderwijssystemen geweest van een 'toetscultuur' naar een meer brede 'assessmentcultuur' waardoor meer toetsen met een eerder ontwikkelingsgericht doel worden afgenomen. Tegelijkertijd betekent de verhoogde autonomie van scholen in vele onderwijssystemen echter dat scholen ook meer verantwoording dienen af te leggen, wat onder meer heeft geresulteerd in een toename van gestandaardiseerde toetsen met een verantwoordingsgericht doel (OECD, 2013; Smith, 2014). Vanuit economisch standpunt kan inderdaad worden geredeneerd dat autonome scholen, bij gebrek aan rigoreuze doelstellingen en competitie, hun eigen belangen zullen nastreven (bijvoorbeeld het beperken van de werklast van personeelsleden) ten koste van de belangen van leerlingen (Bishop & Wössmann, 2004; Wössmann, 2005).

Vaak worden aan toetsen zowel ontwikkelingsgerichte als verantwoordingsgerichte doelen toegeschreven (Linn, 2000; Wang et al., 2006). Tevens geldt dikwijls de assumptie dat een toets met verantwoordingsgerichte doelen indirect zal leiden tot ontwikkeling (Looney, 2014). Andere auteurs wijzen evenwel op de gevaren van het combineren van ontwikkelingsgerichte en verantwoordingsgerichte doelen. De verantwoording zal immers maken dat er een *incentive* bestaat om resultaten op allerlei oneigenlijke manieren te verhogen, waardoor ze niet meer bruikbaar zijn in een ontwikkelingsperspectief (Saunders, 1999).

2.3.2 Macro-, meso- en/of microniveau

Een leer(winst)monitoringssysteem kan ingezet worden om gegevens te verzamelen op verschillende niveaus. Binnen elk van deze niveaus kan bovendien zowel het ontwikkelingsgerichte als het verantwoordingsgerichte perspectief worden gehanteerd. Van elk van de mogelijke combinaties worden in Tabel 2 enkele voorbeelden gegeven.

Zo kan op het microniveau het gebruik op niveau van de klas en de individuele leraar enerzijds en het gebruik door de individuele leerling anderzijds onderscheiden worden. Wanneer de resultaten worden weergegeven op klasniveau, krijgt elke leraar gegevens over zijn/haar professionele effectiviteit wat de getoetste leerdoelen betreft (Baker et al., 2010). De gegevens over de prestaties van verschillende leerlingen kunnen worden aangewend om op een meer efficiënte wijze differentiatie in te zetten (Collins, Reiss, & Stobart, 2010). Gestandaardiseerde toetsen kunnen ook worden gebruikt om leerlingen inzicht te helpen krijgen op hun eigen voortgang, en kunnen op die manier leiden tot toename van de motivatie (Wang et al., 2006). Daarnaast kunnen de resultaten van gestandaardiseerde toetsen in een leer(winst)monitoringssysteem ook ingezet worden binnen een verantwoordingsgericht perspectief. Voornamelijk in de Verenigde Staten is al geëxperimenteerd met verschillende vormen van *performance-based payments* waarbij de resultaten van gestandaardiseerde toetsen een sterke invloed hebben op de beoordeling van leraren (Fryer, 2013; Harris, 2011; Wiliam, 2010). Ten slotte kunnen gestandaardiseerde toetsen ook leiden tot summatieve oordelen over leerlingen, waarbij bijvoorbeeld de resultaten worden gebruikt om leerlingen al dan niet toe te laten tot bepaalde vormen van vervolgonderwijs (Acquah, 2013; Atkinson & Geiser, 2009).

Het mesoniveau betreft de scholen, aangezien resultaten van gestandaardiseerde toetsen in een leer(winst)monitoringssysteem op schoolniveau geaggregeerd kunnen worden. Op die manier kan informatie over sterktes en zwaktes in verschillende scholen worden gegenereerd (Eurydice, 2009). Deze informatie kan de school hulp bieden bij haar processen van interne kwaliteitszorg en kan een bron zijn om bijvoorbeeld professionele ontwikkeling van schoolteams effectiever in te zetten (Cizek, 2005; Loeb, 2013). Schooleffectiviteitsonderzoek wees uit dat scholen die veel data gebruiken, inderdaad op verschillende parameters meer succesvol zijn (Teddle & Reynolds, 2000; Wayman & Stringfield, 2006). Daarnaast kunnen de resultaten van een leer(winst)monitoringssysteem ook worden gebruikt om scholen verantwoording te laten afleggen over de mate waarin ze met het aan hen toegekende overheidsgeld efficiënt zijn omgegaan. Zo kunnen het leer(winst)monitoringssysteem één van de bronnen zijn bij de beoordeling van de kwaliteit van scholen. Het is een internationale trend dat steeds vaker de resultaten van gestandaardiseerde toetsen worden gebruikt in een verantwoordingsgerichte functie bij het evalueren van scholen (Smith, 2014), of openbaar worden gemaakt om marktmechanismen toe te laten waarbij ouders de scholen kiezen op basis van, onder meer, outputinformatie (Bishop & Wössmann, 2004). Scholen kunnen ten slotte ook de resultaten van een leer(winst)monitoringssysteem op een strategische manier gebruiken, bijvoorbeeld om zich aan ouders te verantwoorden voor de kwaliteit van hun onderwijs (Saunders, 1999).

Tabel 2 - Voorbeelden van doelen van een leer(winst)monitoringssysteem

Niveau		Ontwikkeling	Verantwoording
MICRO	Leraren	Zelfevaluatie van de eigen prestatie Beter inzetten van curriculum en van didactische methoden Differentiatie	Evaluatie van leraren Beslissingen inzake benoemingen en loopbaan
	Leerlingen	Zelfevaluatie, zicht op voortgang Leergerichte feedback	Behalen diploma Beslissingen inzake voortgang of zittenblijven
MESO	Scholen	Zelfevaluatie van scholen: interne kwaliteitszorg Betere inzet van pedagogische begeleidingsdiensten	Gebruik door onderwijsinspectie Gebruik voor schoolbesturen of scholengemeenschappen
MACRO	Onderwijssysteem	Beleidsontwikkeling	Beleidsevaluatie Maatschappelijke verantwoording

Op macroniveau kan een onderwijsverheid van de resultaten van een leer(winst)monitoringssysteem gebruik maken om het onderwijsbeleid van een informatierijke omgeving te voorzien en om haar eigen effectiviteit op een systematische manier te onderzoeken (Cizek, 2005; Loeb, 2013). Bovendien kan het onderwijsbeleid zich aan de hand van deze objectieve gegevens ook verantwoorden voor haar beleid of voor wijzigingen in haar beleid. Metingen en cijfers “maken een wezenlijk en zelfs erg bepalend deel uit van het ontwerpen, plannen, bijsturen en beheren van een onderwijssysteem” (Standaert, 2014, p. 16).

2.4 High-stakes en low-stakes leer(winst)monitoringssystemen

De impact van een leer(winst)monitoringssysteem hangt in zekere mate samen met de grootte van de eraan verbonden inzet (de 'stakes' van een toets). Toetsen binnen zulk systeem hebben een hoge inzet (*high-stakes*) wanneer er rechtstreeks gevolgen aan de resultaten zijn gekoppeld, zoals beslissingen over voortgang/zittenblijven van leerlingen, salarissen of bonussen voor personeelsleden of beloningen of sancties voor scholen (Au, 2007; Cizek, 2005; Madaus, 1988). Aanvullend dient te worden gesteld dat toetsen ook als high-stakes worden beschouwd indien een bepaald gebruik van de resultaten, bedoeld of onbedoeld, kan leiden tot beslissingen met ernstige gevolgen voor leerlingen, leraren of scholen (AERA, 2000). Voorbeelden daarvan zijn al dan niet doorstroom in de schoolloopbaan of al dan niet toegang verlenen tot vervolgonderwijs.

Typische voorbeelden van high-stakes toetsen zijn eindexamens die bepalen of leerlingen al dan niet een diploma behalen. Ook leer(winst)monitoringssystemen waarbij de resultaten gebruikt worden om werkingsmiddelen toe te kennen of waarvan de resultaten openbaar gepubliceerd worden, zijn hier voorbeelden van. Dergelijke systemen floreren vooral in onderwijscontexten waar er een grote competitie tussen scholen bestaat en het marktmechanisme bij schoolkeuze wordt bevorderd (Sahlberg, 2011). High-stakes toetsen zijn dan ook meestal toetsen die door alle leerlingen in een heel onderwijssysteem moeten worden afgelegd, en die in een verantwoordingsgericht perspectief worden gebruikt (Marchant, 2007).

Een leer(winst)monitoringssysteem hoeft niet noodzakelijk voor elke betrokkene dezelfde inzet te hebben. Zo bestaat er in de Verenigde Staten een lange traditie van gestandaardiseerde toetsen die *low-stakes* zijn voor leerlingen, maar high-stakes voor leraren (Wiliam, 2010).

Wanneer de resultaten van een leer(winst)monitoringssysteem openbaar worden gemaakt, is er steeds een zekere inzet aan de toetsen verbonden (Jürges, Büchel, & Schneider, 2005). Daarom worden ze in dat geval door sommige auteurs per definitie als high-stakes bestempeld (Au, 2007; McNeil, 2000). Een leer(winst)monitoringssysteem heeft daarentegen een lage inzet (*low-stakes*) als er nauwelijks consequenties aan de resultaten zijn verbonden. High-stakes toetsen zijn vaak verbonden aan een verantwoordingsgericht perspectief (zie 3.1.1), terwijl *low-stakes* gestandaardiseerde toetsen eerder voor een ontwikkelingsgericht doel worden gebruikt. In de hoogte van de inzet bij verantwoordingsgerichte toetsen is nog een grote variatie mogelijk. Zo bestaan er bijvoorbeeld verschillende praktijken van gebruik van gestandaardiseerde toetsen om een diploma toe te kennen aan leerlingen: in sommige landen tellen de resultaten op het examen voor 100% (bijvoorbeeld Frankrijk), in andere landen 50% (bijvoorbeeld Nederland) of minder dan 50% (bijvoorbeeld Portugal), terwijl de rest van het eindcijfer dan wordt ingevuld door resultaten op interne evaluaties van de school (Klein & Van Ackeren, 2012).

De inzet van een leer(winst)monitoringssysteem is niet enkel sterk gerelateerd aan de impact of aan het gebruik van de toetsresultaten, maar hangt ook nauw samen met de betrouwbaarheid van de toetsafname. Bij toetsen met een hoge inzet zijn scholen, leraren en leerlingen immers meer geneigd om de resultaten op oneigenlijke of onwenselijke manieren te verhogen, waardoor bias ontstaat en de toetsresultaten geen getrouwe weergave van de werkelijkheid zijn (Harris, 2012). Tegelijk wordt geargumenteed dat de toetsbetrouwbaarheid net extra zorgvuldig moet worden bewaakt wanneer het high-stakes toetsen betreft (Elstad, 2009; Wiliam, 2010).

2.5 Verschillende criteria

Een leer(winst)monitoringssysteem levert gegevens op aan degene voor wie de resultaten beschikbaar zijn. Mogelijke gebruikers zijn de: overheid, schooldirecties, leraren, leerlingen, ouders, inspectie, andere onderwijsbetrokkenen of het ruimere publiek. Deze gegevens worden evenwel maar zinnige informatie op het moment dat ze binnen een bepaald referentiekader kunnen worden afgetoetst. Dat referentiekader wordt ook het 'criterium' genoemd dat wordt gebruikt om aan de gegevens betekenis te geven (Neumann, Trautwein, & Nagy, 2011). Er zijn drie soorten van criteria mogelijk: een groepsnorm-criterium, een standaard-criterium en een zelf-criterium (Lysne, 1984).

Bij een groepsnorm-criterium wordt een bepaalde score afgewogen ten opzichte van de scores van anderen. De score van een leerling voor een wiskundetoets wordt bijvoorbeeld in perspectief geplaatst door een overzicht van de prestaties van zijn klasgenoten of door te vergelijken met het klasgemiddelde. In klassituaties is dit het meest gebruikte criterium (Neumann et al., 2011). Hoewel dit criterium vaak een goed kader geeft voor de interpretatie van een bepaald resultaat kunnen ongelijkheden in de groep waarmee wordt vergeleken tot foutieve conclusies leiden: zo zal een leerling in een 'zwakke klas' schijnbaar een beter resultaat halen dan een leerling met dezelfde prestatie in een 'sterke klas' wanneer de gegevens enkel binnen de eigen klas worden vergeleken (Willingham, Pollack, & Lewis, 2002). Onderzoek wees zelfs uit dat identieke prestaties in een sterkere context een lager waardeoordeel krijgen dan in een minder sterke context (Südkamp & Möller, 2009 in Neumann et al., 2011). Gestandaardiseerde toetsen hanteren vaak een groepsnorm-criterium, maar breiden de vergelijkingsbasis uit naar alle leerlingen in een bepaald land. Daardoor wordt duidelijk gemaakt hoe een leerling presteert in vergelijking met alle andere leeftijdsgenoten, wat een beoordeling van zijn/haar prestaties kan informeren (Bishop & Wössmann, 2004). Het groepsnorm-criterium kan uiteraard ook op mesoniveau worden gehanteerd, waarbij het vaak *benchmarking* wordt genoemd (Kelly, 2001). In deze methode wordt de gemiddelde leerprestatie van leerlingen in scholen vergeleken met de gemiddelde leerprestatie in andere scholen.

Bij een standaard-criterium wordt een score vergeleken met een absoluut doel, bijvoorbeeld (voor leerlingen) een eindterm of ontwikkelingsdoel. Meer dan bij het groepsnorm-criterium, krijgt men hier een duidelijk beeld van wat en hoeveel iemand geleerd heeft, en of dit als voldoende kan worden beschouwd (Sireci, 2005). Dit criterium wordt best gebruikt wanneer een evaluatie moet worden gemaakt over de vraag of een doelstelling van een bepaald proces al dan niet werd bereikt (Neumann et al., 2011).

Bij een zelf-criterium wordt de score van een leerling, klas, school of onderwijssysteem vergeleken met een score in het verleden. Dit criterium lijkt het meest effectief wanneer een motiverend effect wordt beoogd (Neumann et al., 2011). Bij de berekening van 'leerwinst als groei' of 'leerwinst op basis van eerder leerprestaties' wordt per definitie een zelf-criterium gebruikt.

Vaak worden bij de interpretatie van gestandaardiseerde toetsen in een leer(winst)-monitoringssysteem verschillende van deze criteria gecombineerd (McMillan, Myran, & Workman, 2002; Willingham et al., 2002). Bij grootschalige gestandaardiseerde toetsen met een verantwoordingsgericht doel gaat men meestal in twee fasen te werk. Eerst bepaalt men een absoluut doel (standaard) dat bereikt moet zijn om te 'slagen'. Na de toetsafname kan dit doel worden bijgesteld, op basis van de gemiddelde scores op verschillende items (Klein & Van Ackeren, 2012). Daarbij wordt dus een groepsnorm-criterium toegevoegd aan het standaard-criterium.

3 HISTORIEK

De aard en de bestaansredenen van leer(winst)monitoringssystemen en gestandaardiseerde toetsen zijn zo uiteenlopend, waardoor het moeilijk is om een globaal beeld te krijgen van het totale aanbod (Standaert, 2014). In wat volgt, wordt een beknopt chronologisch overzicht gegeven van het ontstaan van gestandaardiseerde toetsen binnen diverse Europese onderwijssystemen. Vervolgens wordt ook geschetst waarom gestandaardiseerde evaluatiemechanismen in Vlaanderen tot op heden beperkt zijn gebleven.

3.1 Historiek in Europees perspectief

In verschillende buitenlandse onderwijssystemen, en in het bijzonder in Angelsaksische landen, is er reeds een lange traditie van gestandaardiseerde toetsen. Zo opperde de *Royal Commission* in 1861 bij het overzien van het onderwijs in Engeland en Wales dat alle leerlingen in lagere scholen aan een mondeling examen zouden worden onderworpen, dat door een inspecteur zou worden afgenomen. Dit examen werd geïmplementeerd om de gelijkheid tussen scholen te borgen met betrekking tot zowel het geboden onderwijs als de evaluatiepraktijken van scholen. Officieel diende dit initiatief om de goede besteding van overheidsmiddelen aan scholen te verzekeren, al is het mogelijk dat ook een politiek motief (met name het uitbouwen van de zaak van publiek onderwijs als tegengewicht voor confessioneel onderwijs) ten grondslag lag aan deze eerste initiatieven rond gestandaardiseerde toetsen (Wiliam, 2010). In Ierland volgde reeds in 1879 een verplicht eindexamen geïmplementeerd om veralgemeende standaarden te verzekeren in het secundair onderwijs (Klein & Van Ackeren, 2012).

Aanvankelijk werden nationale examens ingezet om beslissingen te kunnen nemen over individuele schoolloopbanen van leerlingen (Eurydice, 2009). Zo is IJsland het eerste land dat in 1946 nationale examens introduceerde om beslissingen te nemen over de doorstroom van leerlingen naar het volgende schooljaar. Kort daarna volgden Portugal (1947) en het Verenigd Koninkrijk (1947) waar ook certificering op basis van de resultaten mogelijk werd gemaakt. In de jaren '60 en '70 van de vorige eeuw volgden ook nog onder meer Luxemburg, Nederland, Malta en Denemarken. In deze periode ging het bijna uitsluitend om eindexamens (Eurydice, 2009). In landen met een afsluitend eindexamen hebben scholen een duidelijk idee van het eindniveau per onderwijsvorm- en richting. Dat scholen hierdoor een deel aan autonomie inboeten om zelf inhoud en doelen te bepalen, leek in deze onderwijssystemen niet op te wegen tegen de gepercipieerde voordelen (Fuchs & Wössmann, 2008).

Toch waren er reeds in die periode onderwijssystemen waarbij de gestandaardiseerde toetsen verder werden uitgebreid met achterliggende doelstellingen (Eurydice, 2009). Zweden, Ierland, Frankrijk, het Verenigd Koninkrijk en Hongarije maakten het mogelijk om meer te analyseren dan louter de individuele schoolloopbanen van leerlingen. Zweedse leraren konden in 1962 de resultaten van hun leerlingen vergelijken met de nationale standaard. Ook in Ierland (1972) werden gestandaardiseerde leestoetsen ontwikkeld als antwoord op een maatschappelijk debat omtrent dit topic. Frankrijk implementeerde in 1977 een gestandaardiseerd meetinstrument dat het mogelijk maakte om verbeterpunten van het onderwijssysteem te identificeren. In het Verenigd Koninkrijk bracht men de leeromstandigheden van leerlingen in kaart en kende men op basis van de resultaten extra werkmiddelen toe aan bepaalde scholen. Tot slot implementeerde Hongarije in 1986 een

meetinstrument dat het monitoren, evalueren en controleren van de onderwijskwaliteit mogelijk maakte (Eurydice, 2009).

In de 21^e eeuw nam het aantal landen of regio's toe waarin gestandaardiseerde toetsen werden gebruikt. Tegelijk met de internationale tendens naar meer autonomie voor scholen, hebben overheden zich steeds vaker toegelegd op het bepalen van minimumvereisten (*core curriculum of education standards*) waaraan (leerlingen in) scholen dienen te voldoen (Buisson-Fenet & Pons, 2011; Eurydice, 2007; OECD, 2013). Door het vooropstellen van minimumdoelen - een vorm van *outcomes-based governance* (E. D. Klein, 2013) - worden uiteenlopende doelstellingen beoogd, zoals de gelijkwaardigheid tussen scholen (Mathis, 2012), de gerichtheid van het onderwijs op minimale maatschappelijke verwachtingen (Florestal & Cooper, 1997; Schoenfeld, 2002), de geïnformeerde ontwikkeling en verantwoording van het onderwijssysteem (Elmore, 2002), en de contractuele verantwoording van scholen ten aanzien van de overheid (Faubert, 2009; OECD, 2013).

De toetsen die in de 21^e eeuw bijkomend worden ontwikkeld, zijn vaak gebaseerd op deze kerncurricula. Ze zijn tevens vaak van een andere aard dan de traditionele examens aan het eind van een onderwijsfase. Een voorbeeld van dit soort toetsen zijn de grootschalige toetsafnames bij een representatieve groep leerlingen in een steekproef van scholen. Deze toetsen hebben geen summatieve functie op systeemniveau en worden afgenomen om een overzicht te krijgen van de onderwijskwaliteit (op macroniveau). In verschillende landen/regio's, zoals Frankrijk, Estland, Finland, Schotland en Spanje, is deze vorm van gestandaardiseerde toetsen intussen geïntegreerd in het onderwijssysteem. Meer en meer zijn leer(winst)monitoringssystemen ontstaan met toetsen die ook ontwikkeld zijn om scholen te ondersteunen in hun interne kwaliteitszorg en beleidsvoerend vermogen. Daarnaast zijn leer(winst)monitoringssystemen met een meer diagnostische doelstelling, op het niveau van de individuele leerlingen, in gebruik in verschillende Europese onderwijssystemen, zoals Denemarken, Ierland, Frankrijk, Engeland en Schotland. In Cyprus, bijvoorbeeld, bestaat sedert 2007 een landelijke toets die als doel heeft om leerlingen met verhoogde kans op analfabetisme zorg op maat te kunnen aanbieden (Eurydice, 2009). Tegelijkertijd zijn er evenwel onderwijssystemen die ook recent formele nationale examens hebben geïmplementeerd. Zo zijn verschillende deelstaten in Duitsland (als gevolg van de 'PISA-shock' aan het begin van de 21^e eeuw) begonnen met een centraal eindexamen (E. D. Klein, 2013).

Tevens hebben technologische ontwikkelingen de tendens tot gestandaardiseerd toetsen versterkt, maar ook de vorm van toetsen sterk beïnvloed. De grotere beschikbaarheid van computers in scholen leidde niet alleen tot een toename van gestandaardiseerde toetsing, maar ook tot een toename van toetsvormen die geschikt zijn voor de computer (Wang et al., 2006; Zenisky & Sireci, 2002)

Momenteel zijn er slechts een beperkt aantal Europese onderwijssystemen waarin geen enkele vorm van gestandaardiseerde toetsen op grote schaal hebben geïmplementeerd. Voorbeelden hiervan zijn Duitstalig België, Tsjechië, Griekenland en Liechtenstein (Eurydice, 2009).

3.2 Historiek in Vlaanderen

In Vlaanderen is de traditie inzake leer(winst)monitoringssystemen of gestandaardiseerde toetsen relatief klein. Eén van de redenen hiervoor is het grondwettelijke principe van vrijheid van onderwijs.

Deze vrijheid van onderwijs impliceert immers onder meer een pedagogische vrijheid. Dat wil zeggen dat een school recht heeft om eigen waarden, attitudes en overtuigingen bij te brengen. Toch moeten scholen die erkende diploma's willen uitreiken en overheidssubsidies willen ontvangen, voldoen aan wettelijke en statutaire voorschriften (Van Petegem & Imbrecht, 2012). Tegenstanders van gestandaardiseerde toetsen hebben evenwel altijd volgehouden dat die toetsen een inbreuk zouden betekenen op het principe van vrijheid van onderwijs. Dit is een opmerkelijke situatie, want in andere landen waar scholen een vrij grote autonomie toebedeeld krijgen, gaat deze autonomie vaak hand in hand met externe gestandaardiseerde toetsen die aan scholen worden opgelegd (OECD, 2013). Zoals vermeld in de probleemstelling zijn de bezwaren ten aanzien van gestandaardiseerde toetsen evenwel steeds talrijk en gevarieerd geweest. De terughoudendheid ten aanzien van gestandaardiseerde toetsen in het algemeen, en verplichte centrale examens in het bijzonder, stoelt op een lange traditie. In tegenstelling tot de economische benadering van onderwijs, waarin scholen worden beschouwd als instellingen die allemaal dezelfde (meetbare) resultaten dienen te produceren, wordt in Vlaanderen doorgaans de visie gehuldigd dat verschillen tussen scholen niet enkel mogelijk maar ook wenselijk zijn, bijvoorbeeld in het creëren van een rijk onderwijsaanbod (Penninckx et al., 2011; Standaert, 2000).

Als gevolg hiervan is het steeds zo geweest in het Vlaamse onderwijs dat, hoewel de overheid scholen verplicht om hun leerlingen op regelmatige basis te evalueren, nauwelijks eisen worden gesteld aan de leerlingenevaluatie. Leraren in Vlaamse scholen ontwikkelen hun eigen evaluatiemateriaal en nemen deze ook zelf af bij hun leerlingen. De beoordeling van het werk van de leerlingen is in handen van de leraar waardoor de rollen van leraar en evaluator in Vlaanderen volledig samenvallen (Van Petegem & Imbrecht, 2012).

Toch werd met de implementatie van de ontwikkelingsdoelen en eindtermen sinds 1997 een eerste stap richting standaardisatie van de minimale leerresultaten gezet. Bovendien hebben de voornoemde bezwaren tegen gestandaardiseerde toetsen verschillende stakeholders, waaronder beleidsmakers, niet weerhouden om voorstellen uit te werken om leerprestaties en leerwinst in kaart te brengen. Zo schreef een werkgroep binnen de toenmalige Dienst voor Onderwijsontwikkeling (DVO) reeds in 1998 en 1999 twee opeenvolgende nota's waarin een meting van leerprestaties en leerwinst werd voorgesteld met zowel ontwikkelings- als verantwoordingsgerichte doelstellingen (Ver Eecke, 2004). Ook recentelijk laaide de discussie weer sterk op, onder meer door het rapport van de OESO over de evaluatiemechanismen in het Vlaamse onderwijs (Shewbridge et al., 2011).

In tegenstelling tot de beperkte ontplooiing van eigen initiatieven om de leerprestaties van leerlingen te monitoren, heeft Vlaanderen wel een lange traditie van deelname aan internationaal vergelijkend onderwijsonderzoek waarbij deze prestaties worden nagegaan. België (onderwijs was toen nog een federale bevoegdheid) was één van de eerste deelnemers in de First International Mathematics Study (FIMS), die in opdracht van de OESO werd uitgevoerd tussen 1963 en 1969 (Penninckx et al., 2011).

4 DE HUIDIGE SITUATIE: GESTANDAARDISEERDE TOETSEN IN HET VLAAMSE ONDERWIJS

Kwaliteitszorg in het Vlaamse onderwijssysteem wordt traditioneel sterk toevertrouwd aan de individuele scholen. Met het Decreet betreffende de kwaliteit van onderwijs werd dit in 2009 nog

sterker uitgesproken. De Inspectie heeft de taak om de interne kwaliteitszorg te controleren en de pedagogische begeleidingsdiensten ondersteunen scholen bij de ontwikkeling ervan. Meer en meer worden gestandaardiseerde toetsen aan scholen ter beschikking gesteld als bron van informatie bij deze interne kwaliteitszorg op microniveau (individuele leerlingen en leraren) en mesoniveau (scholen). Om de kwaliteit van het onderwijssysteem systematisch na te gaan, zijn sinds 2002 peilingen ontwikkeld (Janssen, 2009). Elk van deze gestandaardiseerde toetsen wordt hieronder besproken. Tabel 3 geeft een overzicht van de onderstaande toetsen en hun doelstellingen.

Tabel 3 - Overzicht van gestandaardiseerde toetsen volgens doelstelling

	Niveau	Ontwikkeling	Verantwoording
MACRO	Onderwijssysteem		Peilingen
MESO	Scholen	Parelleltoetsen Toetsen van OVSG en Katholiek Onderwijs Vlaanderen Schoolfeedback systeem	
MICRO	Leraren ----- Leerlingen	Leerlingvolgsystemen Oriënteringstoetsen SALTO	----- Centrale Examencommissie Ingangstoets lager onderwijs

4.1 Peilingen en paralleltoetsen

Tot in 2002 was Vlaanderen één van de weinige regio's binnen de Europese Unie die niet over landelijke data beschikte inzake leerlingenprestaties. Hoewel Vlaanderen wel deelnam aan internationaal vergelijkend onderzoek naar de kwaliteit van het onderwijs richtten deze meetinstrumenten zich niet specifiek op het Vlaamse curriculum. Hierdoor kon geen antwoord gegeven worden op de vraag of Vlaamse leerlingen de vooropgestelde minimumdoelen behaalden (Penninckx et al., 2011).

Door jaarlijks grootschalige toetsafnames bij een representatieve steekproef van Vlaamse scholen en leerlingen te organiseren, kan onderzocht worden in welke mate leerlingen bepaalde eindtermen of ontwikkelingsdoelen bereiken. De resultaten van deze peilingen, die telkens afgenomen worden bij ongeveer 1500 leerlingen, kunnen belangrijke parameters opleveren voor het Vlaamse Ministerie van Onderwijs en Vorming om de kwaliteit van het onderwijs na te gaan en te verbeteren (Standaert, 2000).

De peilingen hebben als doel om op een betrouwbare en valide manier zowel sterktes als zwaktes in het Vlaams onderwijs in kaart te brengen. Via IRT-analyse (Itemresponstheorie) worden items op een gemeenschappelijke ladder ingeschaald, zodat vergelijkingen doorheen jaren mogelijk worden gemaakt. Niet alleen wordt nagegaan hoeveel procent leerlingen de eindtermen behalen, maar ook kan de haalbaarheid en wenselijkheid van de eindtermen worden nagegaan (Steunpunt Toetsontwikkeling en Peilingen, 2015). Daarnaast maken peilingen het eveneens mogelijk om te onderzoeken of er systematische verschillen zijn in prestaties tussen scholen, klassen en leerlingen en of deze verschillen samenhangen met bepaalde school-, klas- of leerlingenkenmerken (Standaert, 2014). Er worden immers ook achtergrondkenmerken verzameld bij leerlingen, ouders,

leraren en directies. Deze informatie stelt beleidsmakers in staat om de kwaliteit van het onderwijs gericht te evalueren, en indien nodig te verbeteren. Deelnemende scholen ontvangen ook een individueel, niet publiekelijk feedbackrapport over de gemiddelde prestaties van hun leerlingen (Steunpunt Toetsontwikkeling en Peilingen, 2015).

Om tegemoet te komen aan de vraag van scholen die niet deelnemen aan de peilingen worden evenwaardige parallelversies van de peilingen ontwikkeld. Deze toetsen hebben een niet-verplichtend karakter en worden gratis ter beschikking gesteld aan alle basisscholen en secundaire scholen, die de resultaten ervan dan zelf kunnen interpreteren, rekening houdend met de context van de school (Steunpunt Toetsontwikkeling en Peilingen, 2015). Scholen kunnen dan informatie verzamelen over zowel de prestaties van hun leerlingen op eindtermen en ontwikkelingsdoelen als over de positie van de school in vergelijking met vergelijkbare scholen. De paralleltoetsen zijn dus een instrument voor de interne kwaliteitszorg van scholen (Standaert, 2014; Verhaeghe, 2010).

4.2 Toetsen van Katholiek Onderwijs Vlaanderen en OVSG

Ook de koepelorganisaties Katholiek Onderwijs Vlaanderen en OVSG stellen gestandaardiseerde toetsen ter beschikking aan scholen die de netgebonden leerplannen volgen. Deze toetsen worden jaarlijks vernieuwd en hebben - in tegenstelling tot de peilingen en paralleltoetsen - geen gemeenschappelijke schaal zodat resultaten niet over de jaren heen vergeleken kunnen worden. Bij beide vormen van toetsen ligt het doel voornamelijk op het mesoniveau: scholen bijkomende informatie verschaffen over de kwaliteit van hun onderwijsaanbod door de leerprestaties van leerlingen in kaart te brengen.

Bij het ontwikkelen van interdiocesane proeven baseert het Katholiek Onderwijs Vlaanderen zich op het eigen curriculum. Deze schooloverstijgende toetsen worden afgenomen in het vierde en/of zesde leerjaar. Naast de toetsen voor wiskunde en Nederlands wordt ook het leerplan wereldoriëntatie geëvalueerd. Doordat elk jaar een andere focus gekozen wordt, komen niet alle eindtermen jaarlijks aan bod. De toetsen gaan na in welke mate de leerplandoelen bij de leerlingen zijn gerealiseerd. Het inrichten van interdiocesane proeven verschaft de school in de eerste plaats informatie op schoolniveau, en heeft niet als doel om individuele leerlingen te beoordelen. Het feedbackrapport kan vooral worden aangewend om de interne werking van de school te beoordelen. De schoolresultaten kunnen vergeleken worden met de gemiddelde resultaten van de Vlaamse katholieke basisscholen en kunnen worden gespiegeld aan referentiegroepen op basis van SES-kenmerken (Katholiek Onderwijs Vlaanderen, z.j.; Verhaeghe, 2010).

De OVSG-toets baseert zich inhoudelijk op het curriculum van het zesde leerjaar van het stedelijk en gemeentelijk onderwijs, hoewel ook andere scholen (waaronder GO! basisscholen) van deze proeven gebruik maken. De OVSG-toets toetst alle leergebieden en domeinen van de basisschool op schriftelijke wijze, zoals wiskunde, Nederlands, Frans, wereldoriëntatie en artistieke opvoeding, en richt daarnaast ook zes praktische toetsen in: Nederlandse en Franse spraakvaardigheid, lichamelijke opvoeding, technologie, artistieke opvoeding en verkeersveiligheid. De toets levert niet enkel betrouwbare informatie over de competenties van hun leerlingen, maar brengt ook informatie om de onderwijskwaliteit te bewaken (OVSG, z.j.).

4.3 Leerlingvolgsystemen

Bij een leerlingvolgsysteem worden verschillende toetsen voorzien die op vlak van de getoetste inhoud en moeilijkheidsgraad van elkaar kunnen onderscheiden worden en waarbij de voortgang van individuele leerlingen in kaart kan worden gebracht. De definitie die werd gegeven door het pilootproject Leerwinstmonitoring en Toegevoegde waarde in primair onderwijs in Nederland wordt hier gevolgd:

“Onder een leerlingvolgsysteem wordt een systeem verstaan dat bestaat uit:

- *een serie toetsen die gebaseerd is op longitudinale leerlijnen uitgezet over meerdere aaneensluitende leerjaren;*
- *een registratiesysteem waarin de vorderingen per leerling en per groep kunnen worden vastgelegd;*
- *normen voor de beoordeling van de vorderingen”* (Janssens et al., 2014, p.9).

Een leerlingvolgsysteem is een instrument dat zowel voor de leraar als voor de school nut heeft (Colpin et al., 2006). Enerzijds brengt een leerlingvolgsysteem de leervorderingen en leeropbrengst van de individuele leerling in beeld. De resultaten van deze metingen kunnen worden aangewend om zorg op maat te bieden door zowel de pedagogische aanpak en didactische werkvormen toe te spitsen op de behoefte van de leerling als door het leer- en onderwijsproces van de leerling bij te sturen (Vlaams Ministerie van Onderwijs en Vorming, z.j.).

Anderzijds ondersteunt een leerlingvolgsysteem ook de school zelf in het coördineren van een planmatige aanpak. Een inzicht in de resultaten op schoolniveau geeft schoolleiders en leraren een beeld van de leer- en onderwijsprocessen en onderwijskwaliteit van de eigen school. Een goed leerlingvolgsysteem houdt rekening met de ontwikkelingsdoelen en eindtermen en kan daardoor worden ingezet als hulpinstrument voor een zelfevaluatie. Doordat het systeem de vorderingen van alle leerlingen in kaart brengt, rekening houdend met de achtergrondgegevens van de leerlingen, kunnen scholen nagaan of het geboden onderwijs voor alle leerlingen even effectief is (Vlaams Ministerie van Onderwijs en Vorming, z.j.).

Er zijn verschillende voorbeelden van leerlingvolgsystemen beschikbaar voor de scholen. De website Toetsen voor Scholen van het Vlaamse Ministerie van Onderwijs en Vorming biedt een Leerlingvolgsysteem voor Vlaanderen (LVS) aan dat verschillende toetsen omvat die de kennis van spelling en de vaardigheid technisch lezen registreren. Ook de Vrije Centra voor Leerlingenbegeleiding in Vlaanderen hebben een eigen leerlingvolgsysteem ontwikkeld bestaande uit toetsen en analyse-instrumenten die specifieke problemen helpen achterhalen.

Omdat een leerlingvolgsysteem het doel heeft om vaststellingen te doen over mogelijke vorderingen tussen twee meetmomenten, is het noodzakelijk dat de resultaten die op die verschillende momenten verzameld zijn wel met elkaar vergeleken kunnen worden. Hierdoor is het noodzakelijk om met meetinstrumenten te werken die over de hele periode hetzelfde blijven (Voeten, 1990). Hoewel enkele leerdomeinen zich hiertoe makkelijk lenen, blijkt dit niet voor alle toetsen het geval te zijn. Vele opeenvolgende toetsen zijn immers vaak zowel onvergelijkbaar inzake inhoud, moeilijkheidsgraad als aantal opgaven.

Om aan dit probleem tegemoet te komen, zijn leerlingvolgsystemen vaak gebaseerd op de itemresponsetheorie (IRT) die middels een uni-dimensionale schaal de ontwikkeling van leerlingen

over meerdere leerjaren kan volgen. Deze schaal kan worden voorgesteld als een meetlat met een constante meeteenheid waarop zowel de vaardigheid van de leerlingen als de moeilijkheidsgraad van de opgaven kunnen worden afgezet. Door de resultaten van leerlingen op verschillende meetmomenten te vertalen naar een positie op de schaal wordt hun vooruitgang in kaart gebracht (Moelands, Mommers, & Oud, 1990).

4.4 Examencommissie van de Vlaamse Gemeenschap

Leerlingen die niet in het bezit zijn van een getuigschrift basisonderwijs, getuigschrift eerste en tweede graad of geen diploma secundair onderwijs hebben, kunnen terecht bij de Examencommissie van de Vlaamse Gemeenschap (EVG).

Het getuigschrift basisonderwijs kan behaald worden door alle leerlingen vanaf het jaar waarin de leerling de leeftijd van negen jaar bereikt. In elke Vlaamse provincie wordt minstens één school van het officieel onderwijs en één school van het vrij onderwijs aangeduid door de Vlaamse regering. Deze scholen, waar de kandidaat zelf uit kan kiezen, hebben de taak om de examens zowel te organiseren als op te stellen (Vlaams Ministerie van Onderwijs en Vorming, z.j.). Wel kunnen kandidaten voor een getuigschrift of diploma secundair onderwijs kiezen tussen examens die gebaseerd zijn op het curriculum van de EVG zelf of op het curriculum van een bepaalde school. Anders dan bij de examens voor het getuigschrift basisonderwijs, is de EVG zelf verantwoordelijk voor de organisatie en het opstellen van de examens voor secundair onderwijs (Penninckx et al., 2011).

Om de kwaliteit van de examens te kunnen borgen, worden de toetsitems voor een bepaald domein in het secundair onderwijs binnen een vakgroep ontwikkeld. Afhankelijk van de leerinhoud bestaan er op dit onderwijsniveau mondelinge en schriftelijke examens, maar ook praktijkexamens en assessments kunnen een vorm van evaluatie zijn. Per vak wordt nagegaan in welke mate de leerling de eindtermen of ontwikkelingsdoelen beheerst. Tevens is de Examencommissie een examenvragendatabank aan het uitbouwen, waarop toetsvragen digitaal ter beschikking worden gesteld (Vlaams Ministerie van Onderwijs en Vorming, z.j.).

4.5 Schoolfeedbacksysteem

Naar aanleiding van IWT-Strategisch Basisonderzoek 'Schoolfeedbackproject' van drie Vlaamse universiteiten (KULeuven, Universiteit Antwerpen en Universiteit Gent) werd in 2011 het Centrum voor Schoolfeedback opgericht (Schoolfeedbackproject, 2015).

Deze instelling biedt scholen ondersteuning aan voor hun interne kwaliteitszorg aan de hand van toetsen die scholen kunnen afnemen bij hun leerlingen. Op basis van de resultaten wordt dan een feedbackrapport gegenereerd dat scholen kunnen aanwenden voor een zelfevaluatie van de eigen werking en het verbeteren van de interne kwaliteitszorg.

Scholen krijgen zowel resultaten op schoolniveau als op individueel leerlingniveau te zien. Zo wordt in eerste instantie informatie gegeven over de toegevoegde waarde van een school. Indien mogelijk wordt er ook rekening gehouden met de toegevoegde waarde betreffende diverse leerlingencategorieën (moedertaal, welvaart gezinnen, ...). Scholen die hun leerlingen op meerdere momenten bevragen, krijgen feedback over zowel de gemiddelde leerwinst als individuele

vaardigheidsscores. Doordat ook achtergrondkenmerken van scholen mee worden opgenomen in de analyses kunnen resultaten van de school vergeleken worden met de resultaten van scholen met een vergelijkbaar leerlingenpubliek (Schoolfeedbackproject, 2015).

4.6 Toetsen voor oriëntering van leerlingen

In het basisonderwijs zijn geen gestandaardiseerde oriënteringstoetsen terug te vinden. Wel biedt de website www.onderwijskiezer.be, een overkoepelend initiatief van centra voor leerlingenbegeleiding van het gemeenschapsonderwijs, het vrije, het provinciaal, het stedelijk en het gemeentelijk onderwijs, de belangstellingsproef *I-like* aan.

Een belangstellingsproef is een hulpmiddel om de belangstelling voor bepaalde studiedomeinen in kaart te brengen. De *I-like* belangstellingsproef richt zich voornamelijk naar leerlingen vanaf het zesde leerjaar basisonderwijs tot en met de tweede graad van het secundair onderwijs. De vragen toetsen de mate van interesse in het uitoefenen van beroepen uit verschillende beroepssectoren op een zevenpuntschaal. Nadat alle vragen beantwoord zijn, verschijnt een grafiek met de resultaten per belangstellingsdomein en bijhorende studierichtingen (Onderwijskiezer, 2015).

In het secundair onderwijs bestaat de *I-Prefer* belangstellingsproef voor laatstejaarsstudenten. Dit meetinstrument tracht de interesse voor opleidingen van het hoger onderwijs weer te geven. De *I-Prefer* is een genormeerde proef waardoor de student zijn of haar resultaten kan vergelijken met die van leeftijdsgenoten. Hij houdt evenwel geen rekening met de competenties van de leerlingen.

In het huidige regeerakkoord is sprake van het implementeren van een verplichte niet-bindende oriënteringsproef voor leerlingen aan het einde van het secundair onderwijs. Deze toelatingsproef beoogt een betere ondersteuning van leerlingen bij hun studiekeuze in het hoger onderwijs. Doordat de slaagcijfers van eerstejaarsstudenten redelijk laag zijn, hoopt de Vlaamse regering dat de resultaten van deze proef jongeren bewustmaken voor de gevolgen van een verkeerde studiekeuze (Vlaamse Regering, 2014). In het academiejaar 2016-2017 zou gestart worden met een toelatingsproef in alle STEM-richtingen en in de lerarenopleiding (Oriëntatieproef, 2014).

Een aanzienlijk deel van de informatie die beschikbaar is voor systeemevaluatie wordt verzameld door het vergelijken ('benchmarken') van voornamelijk leerlingenprestaties. Vergelijkende data van internationale studies creëren een raamwerk voor het Vlaams Ministerie van Onderwijs en Vorming, voor verschillende onderwijsactoren en voor de maatschappij om de sterktes, zwaktes, items en kansen voor verdere ontwikkeling te bepalen.

4.7 Deelname aan internationaal vergelijkende studies van leerprestaties

De Vlaamse Gemeenschap neemt daarnaast deel aan verschillende internationale vergelijkende studies die gericht zijn op de meting van leerlingenprestaties: het Programme for International Student Assessment (PISA; meest recente onderzoek in 2015), de Progress in International Reading Literacy Study (PIRLS; 2016 (lopend)), de Trends in International Mathematics and Science Study (TIMSS, 2015), de International Civic and Citizenship in Education Study (ICCS, 2016 (lopend)) en de European Survey on Language Competences (ESLC 2011). PIRLS, TIMSS en ICCS worden georganiseerd door de International Association for the Evaluation of Educational Achievement (IEA). ESLC tenslotte is een onderzoeksproject van de Europese Commissie.

4.8 Andere toetsen

4.8.1 Taalvaardigheidstoetsen

Het Vlaams Ministerie van Onderwijs en Vorming biedt op haar website toetsinstrumenten aan die de taalvaardigheid van leerlingen aan het begin van het lager onderwijs (SALTO-toets) of het secundair onderwijs (TAS, TASAN en TIST) evalueren. Deze toetsen hebben geen summatief karakter, maar geven de leraren de mogelijkheid om de schoolse taalvaardigheid bij instromende leerlingen te beoordelen (Verhelst, 2010). Deze normgerichte screening zorgt ervoor dat leraren de prestatie van hun leerlingen kunnen inschatten en vergelijken met de prestaties van een referentiegroep. Hierdoor worden niet enkel de leerlingen met een taalachterstand geïdentificeerd, maar kan ook duidelijkheid worden geschapen over de omvang van de achterstand door middel van een berekening van de kloof tussen de norm (verwachte prestaties) en de prestatie van de individuele leerling. Op die manier kunnen zowel de individuele begeleiding als de stimuleringsmaatregelen efficiënter gestuurd worden zodat de taalontwikkeling van deze leerlingen kan bevorderd worden.

4.8.2 Afgeschafte taalproef - instap lagere school

De verplichte taalproef Nederlands die in het schooljaar 2010-2011 zijn intrede deed in het basisonderwijs, had als doel om kinderen die onvoldoende naar school waren gegaan in het kleuteronderwijs, toch de mogelijkheid te geven om in te stromen in het lager onderwijs bij aanvang van de leerplicht. Het onderliggende doel was om ouders te stimuleren hun kinderen ook voor de leerplicht zo vaak mogelijk naar school te sturen. Deze taalproef was verplicht voor leerlingen die minder dan 220 halve dagen aanwezig waren in de derde kleuterklas en die de overstap naar het eerste leerjaar in een Nederlandstalige school wilden maken. De resultaten van deze taalproef wezen uit of de leerlingen over voldoende kennis van de Nederlandse taal beschikken om de doorstart naar het lager onderwijs te kunnen maken. Kinderen die vaak thuisblijven, kampen mogelijk met een taalachterstand wat kan resulteren in een leerachterstand. Tegenstanders van de taalproef waren van mening dat het overdoen van de derde kleuterklas enkel en alleen omwille van taalachterstand de ontwikkeling van andere domeinen in het gedrang brengt. Mede door dit debat werd deze taalproef in het schooljaar 2014 - 2015 afgeschaft.

Momenteel is er wel een taalscreening beschikbaar voor scholen om de beginsituatie van leerlingen in kaart te brengen bij instroom in het eerste leerjaar. Scholen kunnen op basis daarvan beslissen om leerlingen met een grote taalachterstand een remediëringstraject aan te bieden (Vlaams Ministerie van Onderwijs en Vorming, z.j.).

4.8.3 Betrokkenheid en welbevinden van leerlingen

Daarnaast zijn er een aantal meetinstrumenten beschikbaar die de betrokkenheid en het welbevinden van leerlingen in de klas meten. Het Expertisecentrum Ervaringsgericht Onderwijs (ECEGO) ontwikkelde een meetinstrument dat zowel de betrokkenheid als het welbevinden van de leerling situeert op een vijfpuntenschaal op basis van observatie door de leraar (Laevers, Bertrands, Declercq, & Daems, 2004). Een alternatief instrument om het welbevinden van leerlingen te meten is beschikbaar op de website van de onderwijsinspectie (De Volder & De Lee, 2009).

4.8.4 Prodiagnostiek

De website www.prodiagnostiek.be is een initiatief van de onderwijskoepels en CLB-centrumnetten met de steun van de Vlaamse overheid. Het project heeft als uitgangspunt het realiseren van gelijkgerichte, kwaliteitsvolle en wetenschappelijk verantwoorde diagnostiek in leerlingenbegeleiding. Protocollen vormen de leidraad hoe men diagnostische praktijken moet structureren en welke meetinstrumenten moeten gebruikt worden. Ook de interpretatie- en beslissingscriteria werden op voorhand vastgelegd.

4.8.5 Toetsen van educatieve uitgevers

Tal van uitgeverijen springen op de kar van gestandaardiseerde toetsen. Zo beschikt Uitgeverij Van In over twee digitale leerplatformen 'Bingel' voor leerlingen uit het basisonderwijs en 'Diddit' voor leerlingen uit het secundair onderwijs. Beide systemen richten zich op differentiatie en worden aangestuurd door middel van een digitale verbeteringstool die het niveau en de prestaties van de leerlingen evalueert. Onder- en bovenpresteerders krijgen automatisch oefeningen met de juiste moeilijkheidsgraad.

Uitgeverij Abimo is in Vlaanderen vooral gekend om zijn publicaties inzake de technische leesontwikkelingen bij leerlingen. Met behulp van genormeerde toetsen 'Technisch Lezen' kan de leraar het technisch leesniveau van zijn of haar leerlingen bepalen. Wie nood heeft aan meer diagnostisch informatie kan leerlingen toetsen met behulp van de 'Drie-Minuten-Toets' en de 'AVI-leeskaarten'. AVI staat voor 'Analyse van Individualiseringsvormen'. Daarna kan gestart worden met de leesboeken die het juiste AVI-niveau hanteren.