

Eindrapport OBPWO 13.03

Scenario's voor leer(winst)monitoring in Vlaanderen: Een
ontwerponderzoek naar haalbaarheid en wenselijkheid

BELEIDSSAMENVATTING

Promotor-woordvoerder:
Prof. dr. Jan Vanhoof

Promotoren:
Prof. dr. Sven De Maeyer
Prof. dr. Peter Van Petegem

Projectmedewerkers:
Dr. Maarten Penninckx
Amy Quintelier

Vlaams Ministerie van Onderwijs en Vorming



BELEIDSSAMENVATTING

Het voorliggende rapport geeft een overzicht van de resultaten die geleverd zijn door de onderzoeksgroep Edubron (Departement Onderwijs- en Informatiewetenschappen - Faculteit Sociale Wetenschappen) van de Universiteit Antwerpen in het kader van de OBPWO-overeenkomst "Scenario's voor leer(winst)monitoring in Vlaanderen: Een ontwerp onderzoek naar haalbaarheid en wenselijkheid.

Dit onderzoeksrapport bestaat uit negen delen die werden samengebracht. In Deel 1 worden de probleemstelling en de onderzoeksvragen voorgesteld. Deel 2 schetst de onderzoeksmethodiek van de diverse luiken. De daaropvolgende delen rapporteren de resultaten van de drie te onderscheiden onderzoeksactiviteiten, namelijk de literatuurstudie (Deel 3), de eerste en tweede ronde van de Delphistudie (Deel 4) en de internationale studie (Deel 5). In Deel 6 wordt de ontwikkeling van acht afzonderlijke scenario's met een concreet uitgewerkt leer(winst)monitoringssysteem beschreven. Deel 7 beschrijft de resultaten afkomstig uit de derde ronde van de Delphi-studie en geeft een inschatting van de wenselijkheid en haalbaarheid van elk van de scenario's aan de hand van de standpunten van het expertenpanel. Deel 8 en Deel 9 sluiten ten slotte het rapport af met een algemeen besluit en enkele beleidsaanbevelingen. In deze beleidssamenvatting bespreken we kort de belangrijkste inhoud van elk van deze tekstdelen, in de wetenschap dat een samenvatting slechts een beperkte kijk kan geven op het ruimere geheel.

PROBLEEMSTELLING

In verschillende onderwijssystemen woedt een aanhoudende discussie omtrent 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).

In de meest uiteenlopende onderwijscontexten geven gestandaardiseerde toetsen die deel uitmaken van een leer(winst)monitoringssysteem aanleiding tot reeds decennia lang aanslepende discussies (Linn, 2001; L. 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).

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).

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 intermediären) 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 (Vlaamse Regering, 2014).

Het monitoren van leerresultaten/leerwinst is vanuit diverse perspectieven bekeken een complexe aangelegenheid (zowel politiek-maatschappelijk als onderwijskundig en wetenschappelijkmethodologisch) (Janssens, Rekens-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 monitoringsystemen 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.

ONDERZOEKSVRAGEN

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 onderwijscontext. Het betreft dus een beleidsevaluatie ex ante met als doel de beleidsvoorbereiding wetenschappelijk te ondersteunen. De volgende onderzoeksvragen worden vooropgesteld:

- OV1 - Welke zijn (volgens Vlaamse stakeholders) de karakteristieken van een optimaal leer(winst)monitoringsysteem 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?

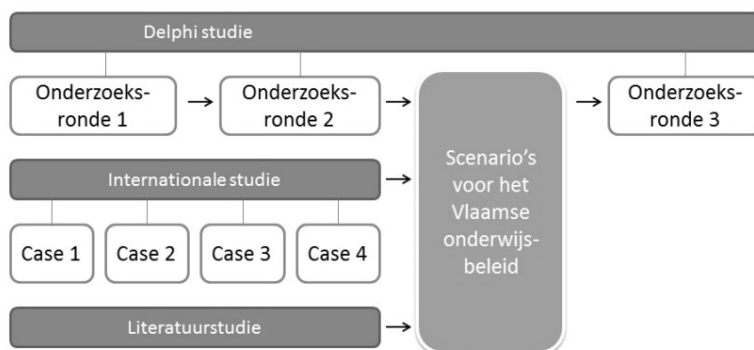
BEGRIPSAFBAKENING

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.

ONDERZOEKSDESIGN

Vanuit methodologisch oogpunt omvat deze studie meerdere deelstudies: een literatuuronderzoek, een internationale studie en een Delphi-studie (zie Figuur A). Elke deelstudie draagt op een unieke manier

bij tot de resultaten van dit onderzoek. Als gevolg hiervan wordt over elk van de studies individueel gerapporteerd in de resultatensectie.



Figuur A. Onderzoeksdesign

Het literatuuronderzoek had als voornaamste doel om relevante nationale en internationale opiniestukken en empirische evidentie in het kader van gestandaardiseerde toetsen en leerwinstmeting, op systematische wijze bijeen te brengen. Daarnaast verleenden de internationale case studies zich ertoe om informatie in te winnen over de kwaliteiten en valkuilen van buitenlandse monitoringssystemen. De beschrijving is gebaseerd op zowel schriftelijke bronnen als op gesprekken ter plaatse met betrokken actoren uit het onderwijsveld. Tot slot zorgden een eerste en tweede ronde van de Delphi-studie ervoor dat ook Vlaamse onderwijsexperten informatie en standpunten betreffende gestandaardiseerde toetsen konden meedelen.

Vanuit deze drie onderzoeksluiken werd vervolgens toegewerkt naar acht werksenario's voor het Vlaamse onderwijsbeleid. De wenselijkheid en de haalbaarheid van elk van deze scenario's werd nadien afgetoetst aan de opinies van de personen in het expertenpanel in een derde onderzoeksronde van de Delphi-studie.

Door middel van een systematisch literatuuronderzoek werd een inventaris gemaakt van de bestaande (inter)nationale kennisbasis omtrent effecten, processen en randvoorwaarden van leerwinstmonitoringsystemen. In tegenstelling tot bij een systematische review studie, werd zowel empirische als non-empirische (opiniërende en feitelijke) literatuur weerhouden met betrekking tot de thema's die centraal staan in dit rapport.

Het literatuuronderzoek startte in oktober 2014 met een zoektocht in de databases ERIC en Web of Science (specifiek het onderzoeksdomein education - educational research). Er werd de voorkeur gegeven aan peer-reviewed tekstmateriaal. Deze oefening werd herhaald in oktober 2015 om recente publicaties op te sporen. Door het grote aantal publicaties rond de (vermeende) effecten en neveneffecten van gestandaardiseerde toetsen diende een keuze te worden gemaakt inzake de te behandelen literatuur. Zoals algemeen wordt aanbevolen in het veld van sociale wetenschappen, blijft

het literatuuronderzoek niet enkel beperkt tot empirische studies (Grayson & Gomersall, 2003). De comprehensiviteit van de zoektocht maakt dat het belangrijkste nadeel van literatuurstudies als nietsystematische reviews (met name het mogelijke risico dat een bias in de onderzoeksresultaten kan voorkomen (Glass, McGaw, & Smith, 1981)) op een adequate wijze werd vermeden.

In het internationale onderzoeksluik werd nagegaan of en op welke manier ervaringen met vier buitenlandse leer(winst)monitoringssystemen onderbouwend kunnen zijn en inspirerend kunnen werken voor Vlaanderen. De beschrijving van de verschillende leer(winst)monitoringssystemen vertrok allereerst vanuit een analyse van geschreven academische bronnen en werd daarna aangevuld met aan het systeem gelieerde praktijk- en beleidsgeoriënteerde literatuur. Vervolgens werden ook de resultaten verwerkt van de interviews die ter plaatse werden afgenomen met een aantal bevoorrechte getuigen (beleidsmakers, ontwikkelaars en gebruikers). Op die manier werd het mogelijk om verschillende effecten, standpunten en kritieken aangaande de bestaande systemen in kaart te brengen. Bij de selectie werd in twee fasen gewerkt. In een eerste fase werd op basis van beschikbare informatie via de OESO (met name de Country Background Reports (CBR) en OECD Review on Evaluation and Assessment Frameworks) en via een rapport in opdracht van de Europese Commissie (Eurydice, 2009) een eerste screening gemaakt van mogelijk interessante leer(winst)monitoringssystemen. Van elk van de overblijvende 16 leer(winst)monitoringssystemen werd in een tweede fase via verschillende kanalen op zoek gegaan naar bijkomende informatie om een zo duidelijk mogelijk beeld te krijgen van de huidige praktijk en recente ontwikkelingen. Daarnaast werden criteria omschreven waaraan de selectie van vier leer(winst)monitoringssystemen diende te voldoen. Elk van de leer(winst)monitoringssystemen moest uiteraard gestandaardiseerde toetsen aanbieden en moest gericht te zijn op het leerplichtonderwijs. Op basis van de in deze fase verzamelde data moest bovendien een redelijk adequaat beeld worden gevormd van het leer(winst)monitoringssysteem en de verschillende gestandaardiseerde toetsen, inclusief van de initiatiefnemers, de doelstellingen, de inzet van de toetsen, de getoetste leer domeinen, en van de wijze van toetsafname. Daarnaast werden criteria opgesteld voor de combinatie van monitoringssystemen. Op basis van de informatie die de hierboven omschreven methode opleverde, en mede na advies van de stuurgroep, werd besloten dat de volgende selectie de meeste kansen op een maximaal informatie-effect voor de Vlaamse onderwijscontext biedt: Australië, Engeland (met focus op het Centre of Evaluation and Monitoring), Nederland (met focus op het pilootproject 'Leerwinstmonitoring en toegevoegde waarde in het primair onderwijs') en Noorwegen (inclusief focus op de stad Oslo). In elk van de cases werden ter plaatse diepte-interviews afgenomen van minimaal acht en maximaal veertien respondenten.

De opinies van verschillende actoren in het Vlaamse onderwijs omtrent de wenselijkheid en haalbaarheid en mogelijkheden van gestandaardiseerde toetsen, leerwinstmonitoring en toegevoegde waarde, waren bij aanvang van deze studie nauwelijks op een systematische manier onderzocht. Deze opinies identificeren en tot één geheel kneden, is dan ook per definitie werk van exploratieve aard. Omwille van dit exploratieve karakter werd gekozen voor een Delphi-studie. In deze methode worden middels een iteratieve feedback-techniek doorheen verschillende onderzoeksrondes de standpunten van experts eerst verkend om ze vervolgens met elkaar te confronteren (Dalkey & Helmer, 1963; Day & Bobeva, 2005; Okoli & Pawlowski, 2004).

Rekening houdend met de functie en de expertise, en in overleg met de stuurgroep, werd een expertenpanel (n=24) samengesteld bestaande uit vier categorieën van stakeholders bepaald. De categorie 'beleidsmakers' (n=4) bevat ook een respondent uit de onderwijsinspectie. De categorie 'onderwijsverstrekkers' (n=6) bestaat uit respondenten van de koepelorganisaties en onderwijsnetten, inclusief pedagogische begeleidingsdiensten. De categorie 'scholen' (n=8) bestaat uit directies, leraren en een vertegenwoordiger van leerlingen, waarbij respondenten evenredig zijn verdeeld over basis- en secundair onderwijs. De categorie 'inhoudsdeskundigen' (n=6) bestaat uit respondenten met expertise in de ontwikkeling van gestandaardiseerde toetsen en academici met een belangrijke stem in dit veld. Concreet werd in deze studie gekozen om de Delphi-methode te operationaliseren middels drie onderzoeksrondes. Van de uitgestuurde vragenlijsten in de eerste ronde van de Delphi-studie werden 22 ingevulde vragenlijsten terugbezorgd aan het onderzoeksteam. De data van de eerste ronde werden geanalyseerd op deductieve wijze volgens de principes van Grounded Theory: er werd bij de codering en analyse niet uitgegaan van vooraf bepaalde categorieën (Strauss & Corbin, 1990). De eerste onderzoeksronde resulteerde in een op data 'gegronde theorie' omtrent effecten en neveneffecten van het monitoren van leerprestaties en leerwinst van leerlingen en de toegevoegde waarde van scholen, alsook omtrent de conceptualisering van de experten van deze begrippen. In de tweede onderzoeksronde werden de standpunten van elk van de respondenten verder verkend en verfijnd middels kwalitatieve diepte-interviews. Tevens werd de houding ten aanzien van stellingnames van andere experten afgetoetst. Op die manier werd de 'gegronde theorie' uit de eerste onderzoeksrunde verder uitgediept en werden belangrijke principes en twistpunten geïdentificeerd. Daarnaast werd in elk van de diepte-interviews toegewerkt naar een ideaal scenario inzake gestandaardiseerde toetsen, leerwinst en toegevoegde waarde voor het Vlaamse leerplichtonderwijs, volgens de opinies van de betreffende respondent. De interviews in de tweede ronde van de Delphistudie namen tussen 53 minuten en 118 minuten in beslag.

In een volgende stap werden de 24 scenario's geanalyseerd en werden gemeenschappelijke kenmerken en twistpunten in kaart gebracht, die de bouwstenen van acht verdere werkscenario's vormen. Op basis van het literatuuronderzoek en de resultaten van de internationale case studies werden deze bouwstenen verder afgebakend tot er uiteindelijk acht volwaardige werkscenario's ontwikkeld werden. Tot slot werden ook mogelijke implicaties per scenario in kaart gebracht. In de derde onderzoeksrunde van de Delphi-studie werden de acht scenario's ter beoordeling aan de 24 leden van het expertenpanel voorgelegd. De respondenten ontvingen deze scenario's enkele weken voor het eigenlijke interview, zodat iedere respondent de tijd had om zich inhoudelijk voor te bereiden. Daarnaast werd een schriftelijke vragenlijst bijgevoegd, met het verzoek om deze vragenlijst ingevuld te retourneren enkele dagen voorafgaand aan het gesprek. Tijdens het diepte-interview kregen de respondenten de kans om hun standpunten verder te verduidelijken en de gegeven scores te motiveren.

RESULTATEN LITERATUURSTUDIE

Het is niet evident om eenduidige conclusies te trekken uit de literatuurstudie. Zo zijn opinie en empirie vaak sterk verweven en moeilijk van elkaar te onderscheiden. Bovendien is de empirische kennisbasis sterk gericht op één bepaalde invulling van leer(winst)monitoringssystemen, met name de

implementatie van high-stakes eindexamens die beoordelen of een leerling een bepaald diploma krijgt, dan wel of een school aan bepaalde kwaliteitsvereisten voldoet. Bovendien is de kennisbasis bedreigd door de moeilijkheden die gepaard gaan met het meten van de effecten van leer(winst)monitoringssystemen.

De resultaten van de uitgebreide literatuurstudie zijn ingedeeld in drie thema's: de impact van gestandaardiseerde toetsen als basis van leer(winst)monitoringssystemen, de bouwstenen van effectieve leer(winst)monitoringssystemen, en de conceptualisering van de begrippen 'leerwinst' en 'toegevoegde waarde'.

In het luik omtrent de impact van gestandaardiseerde toetsen worden (veronderstelde) effecten, alsook (veronderstelde) onwenselijke neveneffecten op micro-, meso- en macroniveau onderscheiden. Op micro-niveau werd op basis van empirische evidentie aangetoond dat leraren via gestandaardiseerde toetsen een beter beeld kunnen krijgen van de leerprestaties van leerlingen, en dat gestandaardiseerde toetsen op die manier een grotere garantie bieden dat alle leerlingen overheen de verschillende scholen op een gelijkmatige manier en volgens een gelijke maatstaf worden beoordeeld. Toch blijkt hier (net als op verschillende andere plaatsen) dat de empirische kennisbasis weinig eenduidig is. Zo wordt er door verschillende studies inderdaad een verschil vastgesteld tussen de prestaties van leerlingen op toetsen die door leraren zelf worden opgesteld, en de prestaties van diezelfde leerlingen op gestandaardiseerde toetsen. Er is evenwel geen eenduidige betekenis te geven aan dit verschil: zowel voor- als tegenstanders van gestandaardiseerde toetsen gebruiken het om de meerwaarde van de eigen meting te propageren. Enerzijds gaan stemmen op dat gestandaardiseerde toetsen leiden tot een 'meer gelijke lat voor alle leerlingen', anderzijds staat hier tegenover dat de gestandaardiseerde toets per definitie een beperking inhoudt ten opzichte van de vele elementen waarmee een leraar in zijn beoordeling rekening kan houden. Binnen dit laatste standpunt past de evidentie waarin wordt aangetoond dat de implementatie van gestandaardiseerde toetsen net tot het tegenovergestelde effect leidt, met name dat deze toetsen (door het gebruikte format) net ongelijkheid tussen leerlingen in de hand werken.

Een ander voorbeeld van incoherentie in de empirische kennisbasis is gerelateerd aan het veronderstelde effect dat de implementatie van gestandaardiseerde toetsen (in het onderzoek veelal geoperationaliseerd als gestandaardiseerde eindexamens) door uiteenlopende mechanismen zouden leiden tot betere leerprestaties. Hoewel er inderdaad verschillende studies deze thesis bevestigen, blijkt uit andere studies dat dit mogelijk kan verklaard worden door onwenselijke neveneffecten als 'teaching to the test' (en dus ten koste gaat van andere leerstofonderdelen of niet-getoetste kennis, vaardigheden of attitudes), terwijl andere studies aangeven dat de verhoogde leeruitkomsten gepaard gaan met een negatieve impact op de autonome motivatie van leerlingen.

Andere positieve effecten op micro-niveau waarvoor empirische evidentie werd verzameld, zijn het positieve effect op het lesgeven van leraren (al werden ook verschillende vormen van 'teaching to the test' empirisch gedocumenteerd), een positief effect op de doelmatigheidsbeleving van leerlingen (alhoewel we hierboven reeds tegenstrijdige evidentie aanhaalden) en een betere evaluatie door leraren zelf. Andere veronderstelde positieve resultaten (waarvoor geen empirische evidentie kon worden aangehaald) zijn de betere onderbouwing van de schoolkeuze en van de studiekeuze van leerlingen.

Het meest in het oog springende negatieve effect op micro-niveau is de zware emotionele impact die gestandaardiseerde toetsen kunnen hebben op leerlingen. Ook hierbij dient te worden gesteld dat de empirische evidentie voornamelijk gericht is op (high-stakes) eindexamens. Tevens werd empirische resultaten besproken met betrekking tot de negatieve impact van gestandaardiseerde toetsen op de studieloopbaan van leerlingen, en onwenselijk strategisch gedrag van leerlingen. Ten slotte werd de (veronderstelde) negatieve impact op de professionaliteit en de motivatie van leraren besproken.

Hoewel ook veel is geschreven over de (neven)effecten van leer(winst)monitoringssystemen op het mesoniveau en macroniveau, blijkt dat de empirische kennisbasis hieromtrent erg beperkt is. De resultatensectie in verband met deze niveaus bevat dan ook voornamelijk beleidsmatige literatuur of literatuur van opiniërende aard. Op mesoniveau wordt in de literatuur veel aandacht besteed aan het veronderstelde positieve effect op de interne kwaliteitszorg van scholen, en de opportuniteit om zowel professionele ontwikkeling van schoolteams als schoolinspecties meer effectief en efficiënt in te zetten. Daar staat echter tegenover dat er veel evidentie is omtrent twee onwenselijke neveneffecten van gestandaardiseerde toetsen, met name onwenselijk strategisch gedrag van scholen, en het ontstaan van rankings(al beschreven we ook de visie van auteurs die het ontstaan van rankings als een opportuniteit, eerder dan een bedreiging beschouwen). Beide onwenselijk geachte neveneffecten zijn vooral gedocumenteerd in een high-stakes toetsomgeving met openbaarheid van resultaten.

Op macro-niveau worden vier positieve effecten onderscheiden, met name: het onderwijsbeleid (en de verschillende stakeholders) die er een beter beeld door kunnen opbouwen van de kwaliteit van het onderwijssysteem, de geïnformeerde basis om het curriculum te evalueren, de beschikbaarheid van data voor onderzoeksdoeleinden, en de publieke tevredenheid. Daar staat tegenover dat verschillende auteurs aangeven dat door de implementatie van gestandaardiseerde toetsen een onderwijssysteem kan vervallen in een overmatige toetscultuur en dat een grotere segregatie tussen sterke en zwakke scholen zou kunnen ontstaan. Ten slotte wordt gewezen op de financiële impact van de ontwikkeling en implementatie van een leer(winst)monitoringssysteem.

Een rode draad doorheen heel deze beschrijving van effecten en neveneffecten is dat er sterke verschillen zijn tussen verschillende scholen inzake de effectiviteit van een leer(winst)monitoringssysteem, alsook inzake de globale effectiviteit van verschillende operationaliseringsmodi van een leer(winst)monitoringssysteem. Het tweede luik van de literatuurstudie onderscheidt daarom verschillende bouwstenen die in de literatuur worden aangegeven als (mogelijke) relevante factor die kan verklaren in welke mate bepaalde (neven)effecten zich voordoen. Deze bouwstenen omvatten de mogelijkheid tot de berekening van leerwinst en van toegevoegde waarde, het opzet van de gestandaardiseerde toetsen in een leer(winst)monitoringssysteem (waarbij de literatuur voornamelijk maar niet uitsluitend aandacht heeft voor het belang van psychometrische kenmerken van de toetsen, de inzet van de toetsen, en het openbare karakter van de resultaten van de toetsen), kenmerken van leraren, kenmerken van scholen, en kenmerken van het onderwijssysteem.

In het derde luik van de literatuurstudie worden de concepten 'leerwinst' en 'toegevoegde waarde' verder ontrafeld, en verschillende mogelijkheden tot operationalisering worden besproken. Hoewel de literatuur verschillende leerwinst-modellen onderscheidt waarbij leerwinst op basis van slechts één

meting wordt geoperationaliseerd, blijkt in de praktijk dat in zowat alle modellen twee of meer metingen worden gehanteerd. Bij gain-based models wordt uitgegaan van de groei van leerlingen op basis van twee of meerdere toetsmomenten. Specifiek aan deze modellen is dat er wordt uitgegaan van een gemeenschappelijke verticale schaal via Item Response Theory (IRT). Daardoor kunnen de resultaten van verschillende toetsen met elkaar worden vergeleken. In de context van gestandaardiseerde toetsen is de meest gebruikte vorm van IRT de inschaling van items volgens een Rasch-analyse. Rasch is een techniek die toelaat om de resultaten van twee toetsen met elkaar te vergelijken, omdat ze op een identieke schaal worden ingeschaald, zelfs wanneer ze verschillende items bevatten. Vaak wordt IRT ingezet om de resultaten van toetsen afgestemd op de leerstof van verschillende leerjaren toch met elkaar te kunnen vergelijken. De ontwikkeling van zulke verticale schaal is erg arbeidsintensief. Bovendien is het niet altijd mogelijk om een gemeenschappelijke schaal te vinden, bijvoorbeeld als twee leerstofdomeinen te ver uit elkaar liggen (bijvoorbeeld twee 'takken' van wiskunde). IRT is daarom makkelijker in te zetten bij niet-curriculum gerelateerde toetsen die een algemene ontwikkeling in kaart willen brengen. Binnen deze gain-based models wordt nog verder onderscheid gemaakt tussen gain score modellen, trajectory modellen, en categorical modellen. Naast de gain-based modellen worden in het rapport conditional status models besproken. Deze modellen beoordelen de prestatie van een leerling aan de hand van een vergelijking met een vooraf bepaalde verwachting. Deze verwachting kan op voorhand voor elke leerling op empirische wijze gesteld op basis van de eerdere prestaties van deze en andere leerlingen (en indien dat wenselijk wordt geacht ook op basis van andere factoren, bijvoorbeeld relevante factoren die buiten de invloedssfeer van de school liggen). Deze modellen vereisen geen gemeenschappelijke verticale schaal. Binnen de conditional status models wordt nog verder onderscheid gemaakt tussen residual gain modellen, projection modellen, en student growth percentile modellen.

Ondanks de vele aandacht voor de verschillende conceptualisering, geeft de literatuur geen aanwijzing omtrent welke conceptualisering in een bepaalde situatie de voorkeur geniet. Door deze lacune leidt de literatuurstudie niet tot een eenduidige conclusie omtrent het meest wenselijke model van leerwinstberekening voor de Vlaamse onderwijscontext.

De term 'toegevoegde waarde' wordt in dit rapport gebruikt op meso-niveau: de bijdrage van de school aan het leren (of de leerwinst) van leerlingen. De literatuurstudie identificeert verschillende vormen van 'statusmodellen' van toegevoegde waarde, waarbij de leerprestaties van leerlingen (op een enkele meting) met elkaar worden vergeleken. Wanneer toegevoegde waarde wordt gebaseerd op de leerwinst van de leerlingen, is de berekening van de toegevoegde waarde afhankelijk van de berekeningswijze die werd gebruikt voor het bepalen van de leerwinst. Daarnaast zijn er verschillende multivariate modellen bekend waarbij meer precies de bijdrage van een leraar of een school aan het leerresultaat van een leerling kan worden bepaald. Doordat meer factoren in de berekening worden opgenomen, discrimineren multivariate modellen vaak sterker tussen scholen: de verschillen tussen scholen zijn groter dan kan worden aangetoond met status- of groeimodellen. Bovendien leveren de multivariate modellen door de veelheid aan factoren meer robuuste en consistente resultaten op. Een nadeel van elk van deze multivariate modellen is evenwel dat het moeilijker is om betekenis te geven aan de resultaten van deze modellen. Het wordt daarom gesteld dat de keuze voor een model van toegevoegde

waarde een trade off is tussen technische accuraatheid en complexiteit enerzijds, en het kunnen afleiden van begrijpelijke en praktisch hanteerbare interferenties anderzijds. Buiten bovenstaande evenwichtsoefening kan evenwel ook inzake toegevoegde waarde op basis van de literatuurstudie geen eenduidige conclusie voor de Vlaamse onderwijscontext worden getrokken, laat staan één bepaald model naar voren worden geschoven.

RESULTATEN DELPHI-STUDIE (EERSTE EN TWEEDE ONDERZOEKSRONDE)

Er is een relatief grote consensus tussen de 24 respondenten over de te verwachte positieve effecten en onwenselijke neveneffecten op micro-, meso- en macroniveau. Soms zijn er respondenten die de opinie van anderen in twijfel trekken, maar dit gebeurt slechts eerder uitzonderlijk. De meest in het oog springende effecten zijn sterk in lijn met de (veronderstelde) effecten die in de literatuurstudie worden onderscheiden. Hetzelfde geldt tevens voor de veronderstelde neveneffecten van een leer(winst)monitoringssysteem, waarbij voornamelijk neveneffecten als 'teaching to the test', rankings van scholen en de emotionele impact op leerlingen veel aandacht krijgen.

Toch blijken de verschillende respondenten tot uiteenlopende ideaal-scenario's te komen aan het einde van de tweede onderzoeksrunde. Dat heeft dan vooral te maken met het gewicht dat aan de verschillende effecten en neveneffecten wordt toegekend: sommige respondenten menen dat de som van de positieve effecten zwaarder weegt dan de som van de onwenselijke neveneffecten, terwijl andere respondenten menen dat de balans in de andere richting overhelst. Daarnaast zijn er verschillen inzake het gewicht toegekend aan effecten en neveneffecten op micro-, meso-, en macroniveau, waardoor sommige respondenten menen dat een leer(winst)monitoringssysteem zich in de Vlaamse context eerder op het micro-, of eerder op het meso- of macroniveau dient te richten.

21 van de 24 respondenten in de Delphi-studie vinden na twee onderzoeksrondes dat het huidige Vlaamse onderwijssysteem baat zou hebben bij de invoering van een leer(winst)monitoringssysteem. In het perspectief van bijna elke respondent werd meer dan één doel gekoppeld aan het leer(winst)monitoringssysteem.

In 20 van deze 21 scenario's die door respondenten werden uitgetekend in het kader van leer(winst)monitoring staat de interne kwaliteitszorg van de scholen - een ontwikkelingsgericht doel op mesoniveau - als één van de prominente doelen centraal. Daarnaast worden in 18 van deze 20 scenario's (alsook in het enige andere scenario) gesteld dat het leer(winst)monitoringssysteem bijkomende doelen heeft die zijn gericht op het microniveau van individuele leerlingen en leraren, tevens in een ontwikkelingsgericht perspectief.

Naast bovenstaande doelstellingen, hebben acht respondenten ook doelstellingen op macroniveau toegevoegd (bijvoorbeeld het nagaan van de validiteit van de eindtermen of ontwikkelingsdoelen). In geen enkel scenario staat het macroniveau evenwel centraal.

Door elf respondenten worden ook één of meerdere doelen in een verantwoordingsgericht perspectief toegevoegd. Meestal is dat op microniveau (een impact op beslissingen over het al dan niet slagen van

leerlingen), soms ook op mesoniveau (de externe beoordeling van scholen). Op één uitzondering na, staan deze verantwoordingsgerichte doelen echter niet centraal in de door respondenten ontwikkelde scenario's.

Er zijn duidelijke verschillen tussen de categorieën van respondenten. Bij de 'beleidsmakers' kennen twee van de vier respondenten enkel ontwikkelingsgerichte doelstellingen toe aan het leer(winst)monitoringssysteem; bij een derde en vierde respondent staat ook verantwoording op meso- en macroniveau centraal. Slechts twee van de respondenten bij de 'onderwijsverstrekkers' zien heil in een leer(winst)monitoringssysteem met een verantwoordingsgericht perspectief (respectievelijk op mesoniveau van scholen en op microniveau van leerlingen), maar dan enkel als secundair doel. In de eerste plaats gaat het scenario van elk van de onderwijsverstrekkers echter uit van een ontwikkelingsgericht perspectief, en dan voornamelijk op mesoniveau (interne kwaliteitszorg van de scholen).

Deze verscheidenheid aan invalshoeken betekent dat, ondanks (of net omwille van) de rijkheid aan data die binnen de uitklaring van de verschillende standpunten werd bereikt, uit deze deelstudie moeilijk eenduidige conclusies kunnen worden getrokken. Dit impliceert dat bij het opmaken van scenario's ten behoeve van de derde ronde van de Delphi-studie een breed perspectief zal moeten worden gehanteerd waarbij uiteenlopende doelen aan bod komen.

Daarnaast worden in de eerste en tweede ronde van de Delphi-studie enkele cruciale bouwstenen geïdentificeerd, waarbij verschillende respondenten soms een verschillende invulling geven. Op microniveau meent een meerderheid in elk van de vier categorieën van respondenten dat het van belang is dat de gestandaardiseerde toetsen die erin worden aangereikt zo breed mogelijk zijn. Dat wil zeggen dat volgens deze respondenten naast kennis ook vaardigheden, attitudes en (beroeps)competenties moeten geëvalueerd worden. Bovendien bevestigen elf respondenten dat het toetsaanbod in het leer(winst)monitoringssysteem niet beperkt mag worden tot de leerdomeinen die meest frequent aan bod komen in andere leer(winst)monitoringssystemen, zoals rekenen en taal. Zeven respondenten - waarvan vier uit de categorie 'scholen' - opperen het idee van een toetsenbank, waarin uiteenlopende gestandaardiseerde toetsen in een breed gamma aan leerdomeinen kunnen ontleend worden. Tegelijkertijd erkennen de meeste respondenten dat uiteenlopende vaardigheden, attitudes en competenties zeer moeilijk via gestandaardiseerde toetsen te evalueren zijn. Alle respondenten zijn het er daarnaast over eens dat, indien er geïnvesteerd wordt in de ontwikkeling van een leer(winst)monitoringssysteem, de gestandaardiseerde toetsen die hiervan deel uitmaken gebaseerd moeten zijn op een vooropgesteld curriculum - de eindtermen en ontwikkelingsdoelen, of de leerplannen van de onderwijskoepels. Daarnaast acht men het van belang dat een leer(winst)monitoringssysteem snelle en duidelijke feedback oplevert, waarbij individuele, digitale schoolfeedbackrapporten de voorkeur krijgen. Tot slot stellen verschillende respondenten dat onafhankelijk van de kracht van een leer(winst)-monitoringssysteem, de effectiviteit ervan op het microniveau beperkt zal blijven indien niet wordt geïnvesteerd in een cultuuromslag in vele scholen, waarbij sterker wordt ingezet op differentiatie tussen leerlingen, en op het toekennen van verantwoordelijkheid in het leerproces aan de leerlingen zelf.

Om de beoogde effecten op mesoniveau te bewerkstelligen, worden tevens een aantal vereiste bouwstenen van een leer(winst)monitoringssysteem genoemd. De mate waarin onwenselijke neveneffecten op mesoniveau zich voordoen hangt sterk samen met de beslissing of de resultaten van het leer(winst)monitoringssysteem al dan niet publiek openbaar zullen worden gemaakt. Deze openbaarheid blijkt een sterk overheersende factor in de discussie tussen de 24 respondenten - al kan er in deze fase van de Delphi-studie geen consensus worden bereikt. Vier respondenten in het expertenpanel, verdeeld over de categorieën 'scholen' en 'inhoudsdeskundigen', zijn van mening dat de resultaten van een leer(winst)monitoringssysteem, geaggregeerd op schoolniveau, openbaar mogen worden gemaakt. Bijkomend meent één respondent dat de resultaten niet globaal mogen worden gepubliceerd, maar dat ouders (bijvoorbeeld via de oudergeleding in de schoolraad) wel recht hebben op deze informatie. Eén respondent wil hierin geen standpunt innemen, terwijl de achttien andere respondenten uitgesproken tegenstander zijn van openbaarheid van resultaten. De grootste vrees van de tegenstanders van openbaarheid (naast verschillende andere veronderstelde neveneffecten) is dat rankings zullen worden opgemaakt die op basis van beperkte gegevens een onterecht als betrouwbaar beschouwde (want op een beperkt deel van het hele takenpakket van de school) uitspraak zullen doen over de kwaliteit van de school.

Een andere bouwsteen op mesoniveau situeert zich binnen de school. Het aanwenden van data als een instrument om het beleid van scholen vorm te geven, is volgens de respondenten geen vanzelfsprekendheid. De mate waarin een leer(winst)monitoringssysteem een impact kan hebben op de interne kwaliteitszorg van de scholen, alsook de mate waarin scholen zich laten verleiden tot ongewenst strategisch gedrag, is volgens respondenten uit verschillende categorieën sterk gelinkt aan het beleidsvoerend vermogen van scholen. De veronderstelling luidt dat hoe sterker dit beleidsvoerend vermogen is ontwikkeld, hoe groter de positieve impact zal zijn. Bovendien wordt verondersteld dat scholen met een sterk ontwikkeld beleidsvoerend vermogen minder onwenselijke neveneffecten zullen vertonen, omdat de aanvaarding van een leer(winst)monitoringssysteem wellicht groter is, alsook omdat er in sterkere mate een cultuur heerst van vertrouwen dat de resultaten binnen een ontwikkelingsgericht perspectief zullen worden gebruikt en kunnen leiden tot kwaliteitsverbetering ten bate van de leerlingen. Scholen met een beperkte datageletterdheid kunnen verkeerde conclusies trekken en daardoor ondoordachte acties op het getouw zetten. De meest genoemde relevante dragers van beleidsvoerend vermogen in dit kader zijn het 'reflectief vermogen' van scholen, alsook het 'innovatief vermogen'.

Ten slotte worden verschillende bouwstenen op macroniveau genoemd. Elf respondenten - waarvan zes in de categorie 'scholen' - geven aan dat het effectief gebruik van een leer(winst)monitoringssysteem een cultuuromslag in het onderwijsveld vereist waarin 'vertrouwen' centraal staat opdat scholen effectief van de resultaten van een leer(winst)monitoringssysteem gebruik kunnen maken. Scholen moeten aanvoelen dat ze in vrijheid kunnen werken en dat de overheid hen vertrouwt dat ze op een zinvolle manier met het leer(winst)monitoringssysteem aan de slag gaan. De gevolgtrekking uit deze vraag naar vertrouwen is evenwel niet gelijklopend: voor sommige respondenten betekent dit een argument om geen verantwoordingsgericht perspectief toe te kennen aan de resultaten van een leer(winst)monitoringssysteem. Andere respondenten (allen uit de categorie

'scholen') menen integendeel dat scholen zich door middel van de resultaten van een leer(winst)monitoringssysteem kunnen verantwoorden voor het gegeven vertrouwen.

Tien respondenten, verspreid over elk van de vier categorieën, wijzen op het belang van de aanvaarding van het leer(winst)monitoringssysteem door uiteenlopende stakeholders in het Vlaamse onderwijsveld. De overheid wordt op haar taak gewezen om bij de beslissing tot invoering van een leer(winst)monitoringssysteem expliciet een campagne op te zetten met onder meer publicaties en vormingsdagen die de meerwaarde hiervan aantoont voor de verschillende betrokkenen. Uiteenlopende respondenten schatten dat er makkelijker een draagvlak zal gevonden worden indien louter op ontwikkelingsgerichte perspectieven wordt ingezet en waarbij de scholen een soort 'eigenaarschap' kunnen ontwikkelen van het leer(winst)monitoringssysteem. Daarnaast zijn er een aantal respondenten die aangeven dat het relevant kan zijn om van start te gaan met een beperkte set toetsen, gericht op hoofdvakken en op de ontwikkeling van scholen, om daarna de lijn verder door te trekken naar andere vakken en om eventueel ook verantwoordingsgerichte doelen te integreren.

RESULTATEN INTERNATIONALE STUDIE

In elk van de vier bestudeerde buitenlandse leer(winst)monitoringssystemen werd een uitgebreide analyse gemaakt van de verschillende gestandaardiseerde toetsen die deel uitmaken van het betreffende monitoringssysteem en van de manier waarop de verschillende toetsen met elkaar interageren. De resultaten van deze analyses werden nadien vertaald naar aandachtspunten voor de ontwikkeling en implementatie van een leer(winst)monitoringssysteem in de Vlaamse onderwijscontext. Deze aandachtspunten zijn gerelateerd aan (1) de implementatie van een leer(winst)monitoringssysteem, (2) de eigenschappen waaraan zulk monitoringssysteem dient te voldoen, (3) de gegevens over leerwinst en (4) de toegevoegde waarde die dit monitoringssysteem zou moeten genereren, (5) de rapportering van de resultaten, (6) de vereiste ondersteuning voor scholen, en ten slotte (7) de financiële impact van een leer(winst)monitoringssysteem.

Beslissing over de implementatie van een eventueel leer(winst)monitoringssysteem

Voor begonnen wordt met de ontwikkeling van een leer(winst)monitoringssysteem, blijkt het raadzaam aan elke gestandaardiseerde toets een duidelijk en eenduidig doel moet worden gesteld, waarbij de verschillende toetsen in een transparante rationale toeleiden naar een coherent geheel van nagestreefde doelen. Uit de case Noorwegen komt de boodschap dat een duidelijke rationale waarin helder wordt gesteld welke doelen met de implementatie van een leer(winst)monitoringssysteem worden nagestreefd, op welke manier dit systeem geacht wordt te leiden tot de vooropgestelde doelen, en welke randvoorwaarden er ingevuld moeten zijn om deze doelen te bereiken, een belangrijke voorwaarde is die de ontwikkeling van gestandaardiseerde toetsen of een toetsensysteem moet voorafgaan. Zonder deze rationale bestaat het gevaar dat verschillende toetsen onvoldoende op elkaar zijn afgesteld of dat teveel doelen aan één toets worden gesteld. Ook in de andere cases wordt door alle respondenten met ervaring in toetsontwikkeling benadrukt hoezeer het van belang is om eerst vast te stellen welk doel er aan de gestandaardiseerde toetsen wordt gesteld. Gevraagd naar het

ervaringsdeskundig advies van respondenten voor de Vlaamse context, is vaak het eerste antwoord een vraag: “what is the problem you are hoping to fix?” In bovengenoemde rationale moet onder meer duidelijk blijken of het doel van de gestandaardiseerde toetsen ontwikkelingsgericht, dan wel verantwoordingsgericht is. Het samengaan van beide perspectieven bleek ook in verschillende cases ertoe bij te dragen dat de ontwikkelingsgerichte doelstelling slechts in beperkte mate werd bereikt. Bovendien kunnen we uit de case studies ook concluderen dat bij de ontwikkeling van gestandaardiseerde toetsen best gericht wordt op één niveau (micro-, meso- of macroniveau).

Verder kan worden opgemaakt dat de formatieve toetsen door scholen sterk gewaardeerd worden en dat ze hen richting geven om hun kwaliteit verder uit te bouwen, zowel op micro- als op mesoniveau. In zowel Noorwegen als Australië zien we bevestigd dat hoe hoger de (gepercipieerde) inzet van een toets is, hoe meer onwenselijke neveneffecten verwacht kunnen worden. In elk van de case studies blijkt dat wanneer toetsen een verantwoordingsgericht karakter hebben, de gebruikte toetsitems meteen na de toetsafname worden neergeschreven en gebruikt worden bij de voorbereiding van volgende toetsen. Bij een verantwoordingsgericht gebruik of een competitief element verdient het dus aanbeveling om telkens een volledig nieuwe toets te ontwikkelen. Dit is in veel mindere mate het geval bij toetsen die enkel in een ontwikkelingsgericht perspectief worden gebruikt. Toch komt uit de cases ook een waarschuwing voor low-stakes toetsen die op geen enkele manier van belang zijn voor de leerlingen. Uit enkele voorbeelden blijkt dat een hogere inzet van de toets voor de leerlingen een positief effect kan hebben op de validiteit van de toetsresultaten. Het roept de vraag op of louter formatieve toetsen, die geen enkele verdere betekenis hebben voor de leerlingen, een betrouwbare maatstaf kunnen vormen voor conclusies op meso- of macroniveau.

Bovendien concluderen we uit de internationale cases dat een ontwikkelingsgerichte impact op microniveau niet makkelijk te realiseren is, en wordt door de respondenten aangeraden om bij een beslissing over een eventuele implementatie van ontwikkelingsgerichte toetsen niet enkel uit te gaan van wat de actuele behoeften zijn van leraren in Vlaamse scholen, maar dat een behoefte aangetoond moet worden: “Persuade people by doing”. In elk van de onderzochte onderwijssystemen zien we dat de aanvaarding van elk leer(winst)monitoringssysteem doorheen de jaren gegroeid is - alhoewel er in elk systeem nog steeds fervente tegenstanders zijn. Initiële twijfelaars onder de stakeholders kunnen worden overtuigd van de meerwaarde van een leer(winst)monitoringssysteem, maar dan is het bereiken van een positieve impact, voornamelijk op micro- en mesoniveau, uiteraard een voorwaarde.

Eigenschappen van een leer(winst)monitoringssysteem

In verband met de eigenschappen waaraan zulk leer(winst)monitoringssysteem dient te voldoen, komen de aanbeveling naar boven om te investeren in digitale toetsen om snelle feedback te voorzien. De snelheid van de feedback wordt als essentieel ervaren om een impact te genereren, zeker wat betreft het microniveau. Toch zijn digitale toetsen niet evident, zeker niet voor toetsen met een hogere inzet die door alle leerlingen simultaan moeten worden afgenomen. High-stakes toetsen worden dan ook voornamelijk op papier afgenomen. Enkel in Noorwegen zagen we een voorbeeld van digitale toetsen met een eerder high-stakes karakter op vrij grote schaal.

Ook adaptieve toetsen worden gezien als een grote meerwaarde, omdat ze meer gedetailleerde informatie van leerlingen kunnen weergeven, omdat ze motiverender zijn voor leerlingen die de toets afnemen, en omdat ze (zeker inzake hoog- en laagpresteerders) meer betrouwbare informatie voor leraren en scholen kunnen opleveren. Adaptieve toetsing is volgens bepaalde respondenten dan weer minder aangewezen bij high-stakes toetsen met gevolgen op de schoolloopbaan van leerlingen of de evaluatie van leraren, omdat de hoge consequenties op microniveau een meting zouden vereisen waarbij alle leerlingen dezelfde items aangeboden krijgen. Naar het voorbeeld van bepaalde toetsen van ACER en van de plannen voor NAPLAN kunnen toetsen ook getrapt adaptief worden aangeboden ('branching').

We kunnen geen eenduidige conclusie trekken over de vraag of de gestandaardiseerde toetsen gebaseerd moeten zijn op het bestaande curriculum, dan wel eerder curriculumonafhankelijk moeten zijn. In Australië en in Nederland hecht men er sterk aan om een leer(winst)monitoringssysteem te baseren op het curriculum. Daartegenover staat dat zowel bij CEM in Engeland als in Noorwegen sterk wordt geïnvesteerd in toetsen die niet meteen gelinkt zijn aan het curriculum. In Noorwegen uit zich dat in toetsen rond vaardigheden die men ziet als essentiële einddoelen van het leerplichtonderwijs.

De inzet van curriculum-gebaseerde toetsen leidt evenwel tot enkele problemen van praktische en organisatorische aard. Zo is er het probleem dat scholen niet allemaal op hetzelfde tempo dezelfde leerstof aanbieden en dat niet alle curriculumdoelen per trimester of leerjaar zijn gedefinieerd. De ontwikkeling van curriculum-gerelateerde toetsen houdt bovendien in dat de toetsitems bij elke mogelijke verandering in curricula moeten nagekeken worden op inhoudsvaliditeit. Bovendien bestaan er (in het secundair onderwijs) uiteenlopende curricula voor verschillende studierichtingen. De curriculum-gerelateerde gestandaardiseerde toetsen voor het secundair onderwijs in Noorwegen worden enkel mogelijk geacht door het sterk comprehensieve karakter van het Noorse onderwijs.

Uit deze internationale studie kan ook geen eenduidig advies worden gehaald over het vraagstuk of best alle domeinen, dan wel enkel de kernvakken getoetst worden. Bij een high-stakes toets lijkt het aangewezen om alle leerdomeinen als gelijkwaardig te beschouwen en dus voor elk vak een toets aan te bieden. Daar staat tegenover dat de meeste leer(winst)monitoringsystemen slechts enkele vakken omvatten, meestal reken- en leesvaardigheid.

In elke case studie wordt uitvoerig gedocumenteerd welke procedures er worden gevolgd om de betrouwbaarheid van de toetsen te garanderen. Gestandaardiseerde toetsen in een leer(winst)monitoringssysteem worden uitvoerig ge-pre-toetst en kennen een stevige statistische onderbouw. Deze statistische basis wordt door de scholen sterk gewaardeerd, helpt bij de aanvaarding van de toetsen door leraren, en draagt bijgevolg bij tot de mate waarin men in scholen van de toetsen gebruik maakt binnen hun eigen kwaliteitszorg op meso- of microniveau. RASCHmodellen of IRT-modellen vinden we in elk van de bezochte case studies.

Het is tevens opvallend dat voornamelijk bij high-stakes toetsen werk wordt gemaakt van strikte procedures voor de afname en de correctie van de gestandaardiseerde toetsen. De nauwgezet

omschreven procedures die worden gehanteerd bij toetsen met een hoge inzet worden ervaren als een cruciaal element binnen de breed verspreide aanvaarding van deze gestandaardiseerde toetsen.

Bij de toetsontwikkeling kan het betrekken van veel actoren helpen bij de aanvaarding van de toetsen (en de perceptie van betrouwbaarheid en validiteit). Dit betekent evenwel niet dat het proces van toetsontwikkeling daardoor vlotter verloopt. Bij de ontwikkeling van toetsen is een goede samenwerking tussen test-inhoudelijke ontwikkelaars en de technische ontwikkelaars een cruciale succesfactor.

Leerwinst

Verschillende systemen stellen zich voor als leerwinstmonitoringssystemen. Toch zijn systemen geïnitieerd door een overheid die leerwinst meten uiterst zeldzaam. Dit zien we eigenlijk enkel bij het pilootproject leerwinst en toegevoegde waarde in Nederland - en zelfs daar maakt men gebruik van eerder ontwikkelde toetsen (door CITO). Toetssystemen van private aanbieders (buiten de overheden) voorzien wel vaak een leerwinstmonitoring.

Uit de cases in Nederland, Noorwegen, en Australië blijkt duidelijk dat wanneer scholen informatie over de leerwinst van leerlingen kunnen verkrijgen, hier meestal gretig gebruik van wordt gemaakt. Kritiek op bestaande monitoringssystemen zonder deze mogelijkheid, is vaak gericht op het ontbreken van leerwinstinformatie wat door verschillende respondenten als een 'gemiste kans' wordt omschreven. Uit de verschillende cases blijkt ten slotte ook dat leerwinstmeting met verschillende toetsen doorheen de schoolloopbaan neveneffecten als 'teaching the test' kan voorkomen.

Het verdient verder aanbeveling om in de Vlaamse context goed na te gaan wat scholen wensen, en aan welke vereisten een leerwinstmonitoringssysteem moet voldoen om de kans op optimaal gebruik te maximaliseren. Daarbij moet onder meer worden ingezet op het vinden van een goede balans tussen enerzijds betrouwbaarheid van de berekening van leerwinst en anderzijds de relatieve eenvoud en transparantie van de concepten en berekeningswijzen en de gebruiksvriendelijkheid van de rapportage. Alle leerwinstmonitoringssystemen die werden beschreven, gebruiken een gemeenschappelijke vaardigheidsschaal die aan verschillende toetsen doorheen de schoolloopbaan ten grondslag ligt. Dit is overeenkomstig de eerste aanbeveling uit het Nederlandse pilootproject Leerwinst en Toegevoegde Waarde in het primair onderwijs. Opvallend is dat leerwinstmonitoringssystemen zich bijna steeds richten tot het basisonderwijs, en dan nog meestal beperkt worden tot basisvaardigheden als lezen en rekenen. Voor meer complexe leerdomen is het moeilijker om een gemeenschappelijke vaardigheidsschaal te ontwikkelen.

Toegevoegde waarde

De case studies resulteerden in geen voorbeelden van toegevoegde waarde-metingen die verder gaan dan het vergelijken van resultaten tussen scholen. Naast statusmodellen, werd in Australië en Noorwegen ook een vergelijking gemaakt tussen de behaalde leerwinst in scholen. Dit zorgt voor een meer genuanceerd beeld tussen de scholen, waarbij relevante kenmerken van de leerlinginput van de school een minder grote impact hebben op de berekening van toegevoegde waarde. Door de

Nederlandse projectgroep Leerwinst en Toegevoegde Waarde in het primair onderwijs werd dan wel besloten dat toegevoegde waarde van scholen meten mogelijk is. Toch blijkt dat er nog onderzoek nodig is om de bestaande technieken te verfijnen. Bovendien blijken scholen in verschillende case studies voor hun interne kwaliteitszorg meest gebaat bij relatief eenvoudige vergelijkingen tussen de leerprestaties of leerwinst in de eigen school, en de vergelijkbare indicatoren in andere (vergelijkbare) scholen.

Rapportering

Een brede en gedetailleerde rapportering is belangrijk. De rapportering moet toelaten om informatie te geven over de prestatie op schoolniveau en van individuele leerlingen en klassen in vele subdomeinen van de toets. Daarnaast is het belangrijk dat de informatie snel ter beschikking staat van de leraren en de school. Interactieve rapportering wordt in nagenoeg alle beschreven leer(winst)-monitoringssystemen gebruikt. De opgeleverde data sluiten op die manier zo dicht mogelijk aan bij de individuele doelstellingen van de gebruikers.

In elk van de bezochte onderwijssystemen is er informatie over de geaggregeerde prestaties van leerlingen op gestandaardiseerde toetsen openbaar beschikbaar. Toch is deze publieke openbaarheid van resultaten in elk van de landen gecontesteerd, en werden verschillende onwenselijke neveneffecten aan deze openbaarheid toegeschreven.

Ondersteuning van scholen

Een aanbod voor scholen doorheen het gehele proces van bij aanvang, afname, interpretatie van de opgeleverde data en de koppeling met actieplannen, wordt in elke van de case studies als een belangrijke succesvoorwaarde beschouwd. Dat aanbod gebeurt best geïndividualiseerd en op maat van de school. Het is opvallend dat de noodzaak aan begeleiding op maat van scholen ook zo sterk wordt aanvoeld in de Nederlandse, Engelse en Australische onderwijscontext, waar scholen al decennia lang ervaring hebben met gestandaardiseerde toetsen. Dit begeleidingsaanbod wordt bij voorkeur aangeboden vanuit de organisatie die verantwoordelijk is voor de toetsontwikkeling.

De ondersteuning is vereist bij de afname van de gestandaardiseerde toetsen, en bij het omzetten van de verzamelde informatie naar actieplannen. Doorheen de verschillende cases worden uiteenlopende aspecten van het beleidsvoerend vermogen van scholen door verschillende respondenten vernoemd als cruciale randvoorwaarden. Het gaat daarbij onder meer om de gezamenlijke doelgerichtheid, het reflectief vermogen en het innovatief vermogen van scholen. We zagen daarnaast enkele voorbeelden waarin het maken van de link tussen de geïnterpreteerde data en een actieplan een inherent deel uitmaakt van het leer(winst)monitoringssysteem.

Financiële inschatting is bijzonder moeilijk

Bij elk monitoringssysteem werd gepeild naar de financiële kost ervan. Geen enkele respondent kon een inzage geven in de totale kost. Alle gegevens duiden er evenwel op dat de ontwikkeling van een

leer(winst)monitoringssysteem een kost- en tijdsintensief werk is waarbij verschillende professionals gedurende een langere periode moeten worden betrokken.

ONTWIKKELING VAN WERKSCENARIO'S

Uit de eerste en tweede ronde van de Delphi-studie blijkt enerzijds de relatief grote consensus tussen de 24 respondenten over de te verwachte positieve effecten en onwenselijke neveneffecten, alsook over de bouwstenen die de effectiviteit van een leer(winst)monitoringssysteem in gunstige of ongunstige zin kunnen beïnvloeden; anderzijds blijkt dat toch zeer uiteenlopende scenario's worden uitgewerkt door de verschillende respondenten door verschillen in het gewicht dat aan de verschillende effecten en neveneffecten wordt toegekend. Deze verscheidenheid aan invalshoeken betekent dat, ondanks (of net omwille van) de rijkheid aan data die binnen de uitklaring van de verschillende standpunten werd bereikt, moeilijk eenduidige conclusies kunnen worden getrokken. Dit impliceert dat bij het opmaken van werksценario's ten behoeve van de derde ronde van de Delphistudie een breed perspectief zal moeten worden gehanteerd waarbij uiteenlopende doelen aan bod komen.

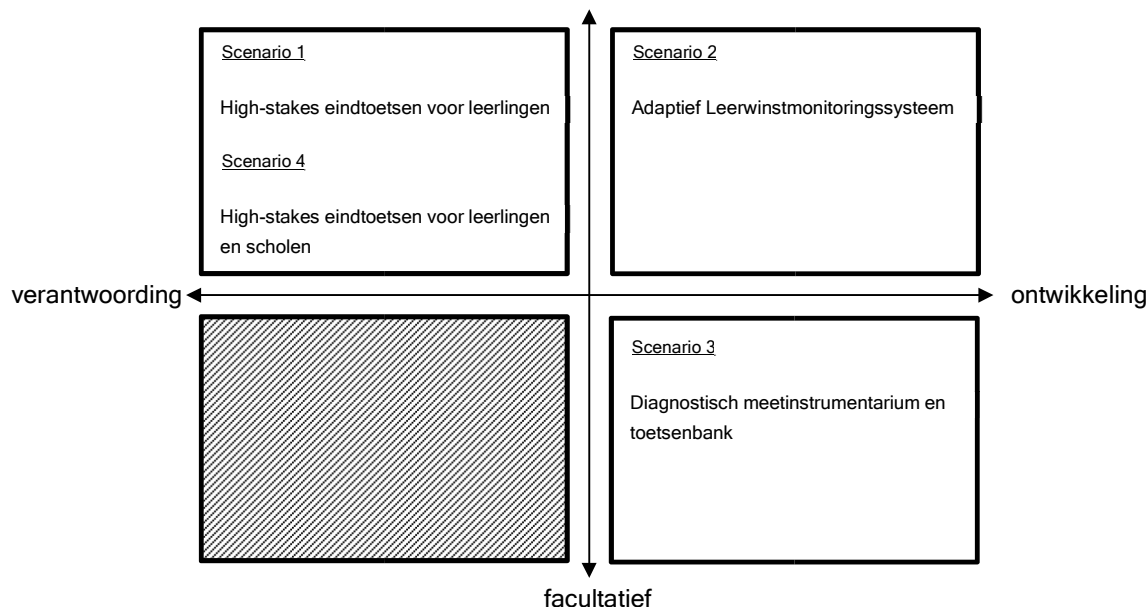
Globaal kan worden gesteld dat een ruime meerderheid van de respondenten in de eerste twee rondes van de Delphi studie (met name 21 van de 24 respondenten) menen dat het huidige Vlaamse onderwijssysteem baat zou hebben bij de invoering van een leer(winst)monitoringssysteem. De aandacht gaat daarbij voornamelijk uit naar ontwikkelingsgerichte doelen op micro- en mesoniveau, maar ook verantwoordingsgerichte doelen, alsook doelen op macroniveau worden door verschillende respondenten genoemd.

Verder kunnen op basis van de resultaten in de hierboven besproken deelstudies een groot aantal bouwstenen worden geïdentificeerd. De 'primaire bouwstenen' zijn de bouwstenen die meer bepalend zijn dan andere bouwstenen, omdat de invulling ervan ook consequenties heeft voor de invulling van de andere (secundaire) bouwstenen. Uit het internationale onderzoeksluik werd geconcludeerd dat het van groot belang is om eerst het doel van het leer(winst)monitoringssysteem te bepalen. Dit doel (enerzijds gerichtheid op micro-, meso-, of macroniveau; anderzijds het ontwikkelingsgericht of verantwoordingsgericht perspectief) wordt dan ook als een primaire bouwsteen beschouwd.

Daarnaast blijkt ook het verplichtend (dan wel facultatief) karakter van het leer(winst)monitoringssysteem een factor die cruciaal is bij de invulling van andere bouwstenen. Daarom wordt ook de vraag of de toetsen binnen een bepaald scenario door alle leerlingen (van een bepaalde cohorte) in elke school moeten worden afgelegd, als een primaire bouwsteen beschouwd.

Op microniveau werden in totaal vier scenario's ontwikkeld (zie figuur B).

verplicht



Figuur B. Overzicht van de verschillende scenario's op microniveau

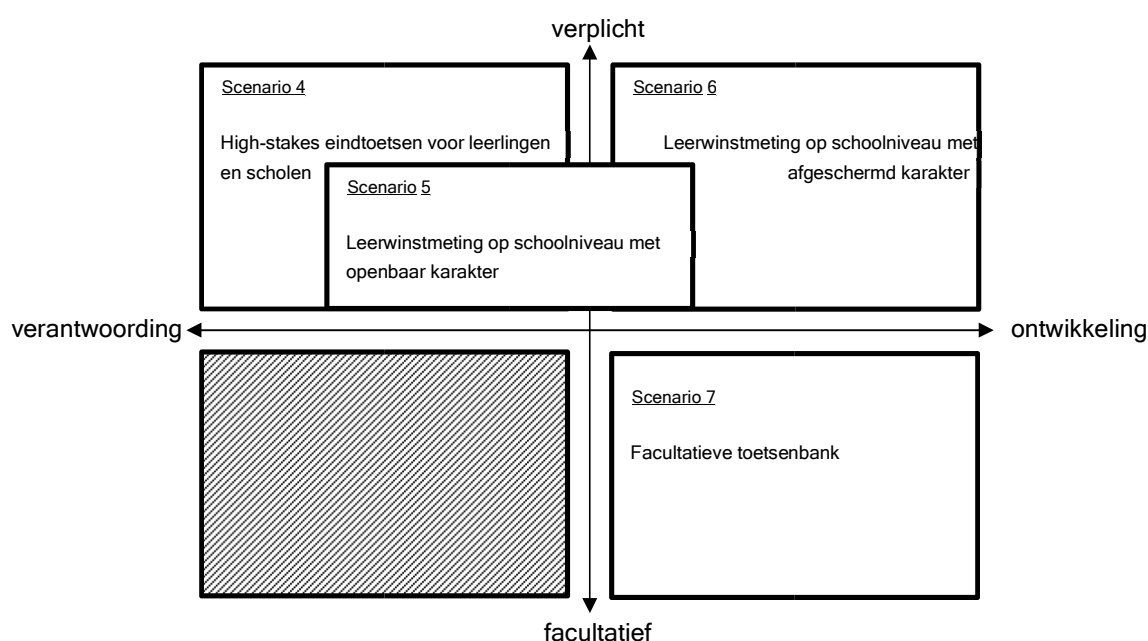
In het eerste scenario 'high-stakes eindtoetsen voor leerlingen' worden gestandaardiseerde toetsen ontwikkeld die als eindtoets worden ingezet bij elk van de onderwijsniveaus in het leerplichtonderwijs. De resultaten van deze verplichte eindtoetsen beïnvloeden in progressief opgebouwde mate of de leerling een getuigschrift basisonderwijs, een oriënteringsattest van de eerste of tweede graad secundair onderwijs of een diploma secundair onderwijs behaalt. Het verantwoordingsgerichte perspectief wordt dus expliciet op het microniveau van leerlingen gericht: het zijn leerlingen die door middel van de gestandaardiseerde toetsen dienen te bewijzen dat ze een getuigschrift, oriënteringsattest of diploma verdienen. De opeenvolging van deze eindtoetsen vormt tevens de basis voor een leerwinstmonitoringssysteem.

Het tweede scenario met de titel 'adaptief leerwinstmonitoringssysteem' bevindt zich in het kwadrant ontwikkelingsgericht/verplicht. Binnen dit leerwinstmonitoringssysteem worden een reeks op elkaar volgende digitale adaptieve toetsen aangeboden. Het hoofddoel ervan is de leraar inzicht te geven in de ontwikkeling en in de leerwinst van de leerlingen inzake enkele essentieel geachte vaardigheden. Hiertoe moeten alle leerlingen minimaal één maal per jaar een toets per vaardigheid afleggen. De resultaten van deze toetsen zijn in geen enkel geval bindend voor attestering of diplomering, maar kunnen gebruikt worden als informatiebron bij bijvoorbeeld differentiatie of studieloopbaanbegeleiding.

Het derde scenario wordt een 'diagnostisch meetinstrumentarium en toetsenbank' genoemd. In dit scenario in het kwadrant 'ontwikkelingsgericht' en 'facultatief' wordt een leer(winst)monitoringssysteem als een toetsenbank met gestandaardiseerde toetsen ontwikkeld. Scholen en leraren kunnen uit deze toetsenbank gestandaardiseerde toetsen gebruiken om na te gaan welke aspecten van een bepaald leerdomein door de leerling reeds beheerst worden en waar die leerling nog remediëring kan gebruiken. De toetsen zijn gericht op zeer specifieke leerstofonderdelen.

Het vierde en laatste scenario op microniveau ('high-stakes eindtoetsen voor leerlingen en scholen') situeert zich net als het eerste scenario in het verantwoordingsgericht en verplicht kwadrant. Het is het enige scenario dat zich over twee niveaus verspreidt, met name het micro- en mesoniveau. Binnen dit scenario worden gestandaardiseerde toetsen ingezet met een tweeledig hoofddoel: enerzijds het in kaart brengen van de geaggregeerde leerprestaties van leerlingen in scholen; anderzijds het objectiever maken van de beslissing over het uitreiken van getuigschriften en diploma's aan de leerlingen. Het eerste doel is verantwoordingsgericht op mesoniveau, het tweede doel is eveneens verantwoordingsgericht maar dan op microniveau.

Ook op mesoniveau kunnen vier scenario's onderscheiden worden (zie figuur C).



Figuur C. Overzicht van de verschillende scenario's op mesoniveau

Het vijfde scenario 'leerwinstmeting op schoolniveau met openbaar karakter' situeert zich tussen twee kwadranten. In dit scenario dragen de resultaten van de ontwikkelde gestandaardiseerde toetsen in het leerwinstmonitoringssysteem bij tot de kwaliteitszorg binnen scholen, waardoor dit scenario haar plaats binnen het ontwikkelingsgericht kwadrant verdient. Hoewel deze verplichte en schriftelijke toetsen een formatief karakter hebben (en als low-stakes gezien worden voor leerlingen), worden ze door de openbaarheid van gegevens inzake toegevoegde waarde op schoolniveau als high-stakes beschouwd, waardoor dit scenario ook thuishoort in het verantwoordingsgericht kwadrant.

In het zesde scenario 'leerwinstmeting op schoolniveau met afgeschermd karakter' dragen de resultaten van het leer(winst)monitoringssysteem bij tot de kwaliteitszorg binnen scholen. De digitale toetsen in dit monitoringssysteem zijn verplicht voor alle scholen, maar hebben een formatief karakter en de resultaten worden niet openbaar gemaakt.

Het zevende scenario bevindt zich in het kwadrant 'facultatief' en 'ontwikkelingsgericht'. In dit scenario wordt leer(winst)monitoringssysteem als een toetsenbank ontwikkeld waaraan scholen toetsen kunnen ontleen die hen helpen bij het in kaart brengen van de mate waarin het onderwijs in de school de leerlingen leidt tot het behalen van de eindtermen, aangevuld met andere vragenlijsten. Het gebruik van deze toetsen past binnen de interne kwaliteitszorg op schoolniveau en is vrijblijvend. De resultaten worden enkel aan de school ter beschikking gesteld. Het verschilt van scenario 3 door de gerichtheid van de toetsen op het mesoniveau: ze zijn in de eerste plaats gericht op het genereren van informatie op schoolniveau.

Uit de eerste en tweede ronde van de Delphi-studie bleek bij de bevroagde respondenten weinig expliciete vraag naar de ontwikkeling van een leer(winst)monitoringssysteem dat gericht is op het genereren van informatie ten behoeve van het macroniveau. Aangezien in de huidige situatie de initiatieven van de overheid inzake ontwikkeling en afname van gestandaardiseerde toetsen voornamelijk zijn gericht op het macroniveau (met name door de peilingsproeven), werd de vraag van respondenten om zoveel mogelijk binnen het huidig kader te blijven gecatalogiseerd als een scenario op het macroniveau. Dit scenario 8 met de naam 'de huidige trend wordt verdergezet' wil evenwel niet zeggen dat niets hoeft te veranderen. In dit scenario wordt immers sterker ingezet op het optimaliseren van de randvoorwaarden van het bestaande monitoringssysteem, zoals het valideren van bepaalde toetsen, om zo de effectiviteit van deze toetsen in het licht van onderwijskwaliteit te versterken. Daarnaast besteedt het scenario aandacht aan de ondersteuning en verdere professionalisering van scholen en lerarenteams in het kader van datageletterdheid.

BEOORDELING VAN DE SCENARIO'S NAAR WENSELIJKHEID EN HAALBAARHEID

In de derde ronde van de Delphi-studie werd bij elk van de 24 respondenten gepeild naar de standpunten inzake de wenselijkheid en haalbaarheid van elk van de acht scenario's. Deze standpunten werden geregistreerd op twee manieren. Allereerst kregen respondenten de vraag om voor elk scenario op een Likert-schaal aan te geven in welke mate zij dit scenario wenselijk en haalbaar achtten. Vervolgens kregen de respondenten de mogelijkheid om hun standpunt verder te verduidelijken en te motiveren in een interview.

Vier scenario's kregen een gemiddelde score boven de neutrale score '3'. Het betreft de vier scenario's waarin een leer(winst)monitoringssysteem wordt ontwikkeld in een ontwikkelingsgericht perspectief. Drie scenario's kregen een gemiddelde score die duidelijk onder de neutrale score '3' ligt. Het betreft de drie scenario's waarbij een leer(winst)monitoringssysteem expliciet in een verantwoordingsgericht perspectief wordt geplaatst. We kunnen hieruit concluderen dat de implementatie van een leer(winst)monitoringssysteem in de Vlaamse onderwijscontext het meeste kans maakt op een breed draagvlak indien het gericht is om een ontwikkelingsgericht doel op micro- of mesoniveau. Het andere spanningsveld dat hierboven als 'primaire bouwsteen' werd onderscheiden (verplichte versus facultatieve deelname van scholen) blijkt geen scherprechter te zijn in de beoordeling van de wenselijkheid door de 24 respondenten. In verband met scenario 8 is het moeilijk om een eenduidige

conclusie te trekken inzake wenselijkheid: negen respondenten beoordelen dit scenario als (eerder) wenselijk, maar elf respondenten menen dat dit scenario (eerder) onwenselijk is.

Verder valt op dat de vier scenario's die eerder positief worden beoordeeld inzake wenselijkheid, de twee scenario's omvat waarin een facultatieve toetsenbank (op microniveau ten behoeve van leerlingen, en op mesoniveau ten behoeve van scholen) wordt ontwikkeld, en de twee scenario's waarin leerwinstmonitoring wordt voorgesteld in een ontwikkelingsgericht perspectief. Belangrijke criteria die door vele respondenten worden gehanteerd bij de beoordeling van de wenselijkheid, zijn de mate waarin het leer(winst)monitoringssysteem geacht wordt een bijdrage te kunnen leveren tot het leren van leerlingen (met onderwijsprocessen op maat van elke leerling) en tot de interne kwaliteitszorg van de scholen.

In de inschatting van haalbaarheid spelen verschillende overwegingen mee. Meest prominent komen evenwel de volgende drie criteria naar boven in de inschatting van de 24 respondenten: de mate waarin men meent dat voor het betreffende scenario een draagvlak bestaat in de Vlaamse onderwijscontext; de inschatting van de haalbaarheid van de vereiste inspanningen om het betreffende leer(winst)monitoringssysteem te ontwikkelen; en de inschatting van de vereiste financiële inschatting. Voornamelijk het eerste criterium maakt dat de scores inzake haalbaarheid een sterke samenhang vertonen met de scores inzake wenselijkheid. De drie scenario's die gemiddeld als onwenselijk worden bestempeld, worden ook eerder onhaalbaar geacht, onder meer omwille van het veronderstelde beperkte draagvlak. Enkel scenario 8 vormt een uitzondering: hoewel de wenselijkheid gemiddeld als neutraal werd bestempeld, wordt de realisatie ervan als zeer haalbaar geacht. Aangezien dit scenario geen ontwikkeling van een nieuw leer(winst)monitoringssysteem voorziet, is de vereiste inspanning immers veel kleiner dan in de andere scenario's.

Verder zien we enkele opvallende verschillen tussen de categorieën van respondenten. Deze komen het meest tot uiting in de beoordeling van de wenselijkheid van scenario 1 en scenario 8. Scenario 1 wordt door elke respondent in de categorie 'onderwijsverstrekkers' als helemaal onwenselijk beschreven, terwijl vijf respondenten uit scholen dit scenario wel wenselijk tot zeer wenselijk achten. Dit reflecteert het verschil in opinie dat reeds in Deel 4 werd beschreven, met name dat respondenten uit de categorie 'scholen' meer dan onderwijsverstrekkers zich kunnen scharen achter het idee dat een leer(winst)monitoringssysteem deels in een verantwoordingsgericht perspectief op microniveau wordt ingezet. Scenario 8 - de huidige trend wordt verdergezet - wordt dan weer sterk gesteund door de onderwijsverstrekkers. Ook dit reflecteert de verschillen in opinies uit de eerdere ronden van de Delhistudie, waar de drie respondenten die meenden dat Vlaanderen niet bijkomend moet investeren in een leer(winst)monitoringssysteem omwille van de verwachte neveneffecten, alle drie tot de categorie 'onderwijsverstrekkers' behoorden. In de categorie 'beleidsmakers' en 'scholen' wordt dit scenario evenwel veel minder gedragen.

CONCLUSIES

Uit bovenstaande studies worden zes globale conclusies getrokken, gerelateerd aan (1) de doelen van het leer(winst)monitoringssysteem, aan (2) leerwinst en toegevoegde waarde, aan (3) openbaarheid

van resultaten, aan (4) eigenschappen van de toetsen, aan (5) ondersteuning voor scholen, en aan (6) het draagvlak dat hiervoor vereist is.

Doelen van het leer(winst)monitoringssysteem

Indien de Vlaamse overheid wenst te investeren in een leer(winst)monitoringssysteem, dient men in de eerste plaats een concrete set aan doelen te bepalen voor dit monitoringssysteem. Uit de internationale studie blijkt immers dat de impact van een leer(winst)monitoringssysteem het grootste is, wanneer aan elk van de toetsen een éénduidig doel wordt gesteld, ingebed in een transparante rationale. De rationale maakt duidelijk hoe elke gestandaardiseerde toets in het leer(winst)monitoringssysteem bijdraagt tot het vooropgestelde overkoepelende doel.

De literatuurstudie toont aan dat zowel verantwoordingsgerichte systemen, als ontwikkelingsgerichte systemen bepaalde positieve impulsen kunnen geven aan het onderwijs. Sommige studies wijzen uit dat deze positieve effecten versterkt worden indien de leer(winst)monitoring een high-stakes karakter heeft, maar deze bevinding wordt niet door alle studies gedeeld. Verder toont de literatuurstudie aan dat een high-stakes leer(winst)monitoringssysteem vaker dan een low-stakes systeem aanleiding geeft tot verschillende onwenselijke neveneffecten.

Uit de resultaten van de Delphi-studie blijkt evenwel dat low-stakes ontwikkelingsgerichte doelstellingen op micro- of mesoniveau in de Vlaamse onderwijscontext de voorkeur verdienen bij de ontwikkeling van een leer(winst)monitoringssysteem. Scenario's die uitgaan van dit ontwikkelingsgerichte perspectief worden hoger ingeschat inzake zowel wenselijkheid als haalbaarheid. Meer concreet dient een leer(winst)monitoringssysteem in Vlaanderen bijkomende betrouwbare en valide informatie te genereren op microniveau waardoor leraren in staat worden gesteld om het onderwijsaanbod maximaal af te stemmen op de noden van de leerlingen alsook om leerlingen beter inzicht te geven in hun eigen voortgang. Op mesoniveau is de verwachting dat het leer(winst)monitoringssysteem een bijdrage levert aan een informatierijke omgeving voor de school door het systeem van interne kwaliteitszorg van betrouwbare en valide informatie over de leerprestaties of de leerwinst van leerlingen te voorzien.

Het internationale onderzoeksluik bevestigt doorheen de vier cases het inzicht uit de literatuurstudie dat het onwenselijk is om zowel ontwikkelingsgerichte als verantwoordingsgerichte doelen te stellen aan één en dezelfde gestandaardiseerde toets. Door het verantwoordingsgerichte karakter wordt de inzet van de toets immers verhoogd, wat in elke case leidt tot indicaties van onwenselijk strategisch gedrag van scholen en waardoor de resultaten aan betrouwbaarheid en daarom aan nut verliezen voor gebruik in een ontwikkelingsgericht perspectief.

Leerwinst en toegevoegde waarde

In het kader van de hierboven besproken voorkeur voor ontwikkelingsgerichte doelen met focus op micro- of mesoniveau, is het belangrijk dat dit monitoringssysteem voldoende nieuwe informatie kan genereren voor zowel scholen als voor individuele leraren. Hoewel in de internationale cases vaak bleek dat men de bevestiging van bestaande inschattingen ook als waardevol beschouwt, geven Vlaamse

respondenten in de Delphi-studie aan dat men van dit monitoringssysteem verwacht dat het nieuwe inzichten aanreikt. Deze doelstelling kan het best worden bereikt indien het monitoringssysteem zich niet beperkt tot het aanleveren van informatie over de leerprestaties van leerlingen. Twee van de vier werkscenario's die in de derde ronde van de Delphi-studie een positieve inschatting krijgen inzake wenselijkheid, betreffen de werkscenario's waarin de leerwinst van leerlingen in kaart wordt gebracht binnen een ontwikkelingsgericht perspectief.

De literatuurstudie onderscheidt uiteenlopende manieren waarop de leerwinst van leerlingen in kaart kan worden gebracht. Uit de internationale studie blijkt dat leerwinstmonitoring op initiatief van de overheid eerder zeldzaam is, maar eerder ligt bij private aanbieders. Bij deze leer(winst)monitoringssystemen bleek in de praktijk een gain-based model (en met name het gain score model) het meest gebruikelijke te zijn. Bovendien wordt dit model positief gewaardeerd door de meeste gebruikers ervan. Bij dit model dienen alle items op een geijkte vaardigheidsschaal te worden ondergebracht, bijvoorbeeld via Item Response Theory (IRT).

Inzake toegevoegde waarde, houdt de focus op een ontwikkelingsgerichte impact op mesoniveau voor de meeste respondenten in de Delphi-studie in dat scholen de leerwinst van hun leerlingen kunnen vergelijken met de leerwinst van vergelijkbare scholen (scholen met een soortgelijke input en context). Er wordt dus duidelijk de voorkeur gegeven aan vergelijking van leerwinst boven vergelijking van leerprestaties. Bij de leerwinstberekening ten behoeve van deze vergelijking tussen scholen wordt idealiter verder gebouwd op relatief eenvoudige modellen van leerwinst. Zeker wanneer de focus op een ontwikkelingsgericht perspectief behouden blijft, kan er uit de Delphi-studie en het internationale onderzoeksluik geconcludeerd worden dat een relatief eenvoudig (en voor scholen makkelijk te begrijpen en te hanteren) model van toegevoegde waarde de voorkeur verdient.

Openbaarheid van de resultaten

In elk van de case studies in het internationale onderzoeksluik werden resultaten van bepaalde gestandaardiseerde toetsen openbaar gemaakt. Hoewel de openbaarheid van resultaten in verschillende cases een relatief groot draagvlak heeft, worden in elke case aspecten ervan door verschillende stakeholders gecontesteerd. Ook in de literatuur blijkt dat de openbaarheid van resultaten van een leer(winst)monitoringssysteem op schoolniveau kan leiden tot een verhoogde transparantie die door bepaalde stakeholders gewaardeerd wordt, maar dat ze bijna onvermijdelijk tot gevolg heeft dat rankings van scholen worden opgemaakt, waarbij een competitie tussen scholen kan ontstaan en waarbij scholen beoordeeld worden op een slechts beperkt deel van hun ruime opdracht. Dit kan onwenselijk strategisch gedrag en zelfs fraude in de hand werken.

De resultaten van de Delphi-studie wijzen uit dat het in de Vlaamse context niet als wenselijk wordt gezien dat resultaten van gestandaardiseerde toetsen in een leer(winst)monitoringssysteem openbaar zouden worden gemaakt. De twee werkscenario's waarbij openbaarheid van resultaten op schoolniveau werd vooropgesteld werden elk door iedere categorie van respondenten gemiddeld sterk negatief

beoordeeld. De openbaarheid van de gegevens is daarbij de belangrijkste reden om beide werkscenari's af te keuren.

Eigenschappen van de toetsen in een leer(winst)monitoringssysteem

Uit de resultaten van de derde ronde van de Delphi-studie kan geconcludeerd worden dat het voor Vlaanderen wenselijk is om niet enkel te investeren in één enkele gestandaardiseerde toets, maar eerder in een omvattend leer(winst)monitoringssysteem waaruit scholen verschillende toetsen kunnen ontlelen. De werkscenari's die in deze richting werden ontwikkeld werden beiden inzake wenselijkheid en haalbaarheid gemiddeld positief beoordeeld door de respondenten in de Delphistudie. Een toetsenbank is dan een uitgebreide database waarin betrouwbare en gevalideerde gestandaardiseerde toetsen zijn opgenomen in uiteenlopende leerdomeinen (al dan niet uitgebreid met niet-cognitieve aspecten).

In de literatuurstudie blijkt dat heel veel is geschreven omtrent het belang van een sterke psychometrische onderbouwing van gestandaardiseerde toetsen die deel uitmaken van een leer(winst)monitoringssysteem. Betrouwbaarheid en validiteit zijn essentiële vereisten van de gestandaardiseerde toetsen in zulk systeem. In het internationale onderzoeksluik stellen we vast dat in elk van de bezochte onderwijssystemen met heel veel zorg wordt omgegaan met de betrouwbaarheid van de toetsen. Ondanks de grote zorg voor kwaliteit en psychometrische onderbouwing, staat soms vooral de validiteit van gestandaardiseerde toetsen ter discussie: de vele voorbeelden uit de internationale studie tonen inderdaad aan dat tot nog toe vaak een beperkt aspect van het hele gamma aan leerdoelen door middel van gestandaardiseerde toetsen wordt geëvalueerd. Om dit tegen te gaan, moeten inspanningen worden geleverd om via het leer(winst)monitoringssysteem ook zogenaamde 'hogere orde-vaardigheden' te toetsen.

In de literatuurstudie worden enkele auteurs vernoemd die ervoor pleiten om bij de ontwikkeling van een leer(winst)monitoringssysteem te vertrekken van het curriculum - in het geval van Vlaanderen: de eindtermen en ontwikkelingsdoelen van de verschillende leerdomeinen. Hierdoor krijgen leraren en scholen een zicht op de mate waarin leerlingen bereikt hebben wat het onderwijsaanbod heeft beoogd. In het internationale onderzoeksluik werd evenwel geconcludeerd dat ook niet-curriculum gebonden gestandaardiseerde toetsen in gebruik zijn en gretig worden afgenomen door de scholen. Voor het Vlaamse onderwijssysteem kunnen we voorzichtig concluderen dat een leer(winst)-monitoringssysteem gebaseerd op de eindtermen en ontwikkelingsdoelen (eventueel leerplannen) de voorkeur geniet boven de systeem dat niet van deze doelen vertrekt. In zes van de acht werkscenari's die werden opgebouwd ter voorbereiding van de derde ronde van de Delphi-studie baseert het leer(winst)monitoringssysteem zich op de eindtermen en ontwikkelingsdoelen.

In de eerste en tweede ronde van de Delphi-studie werd door verschillende respondenten uitdrukkelijk gewezen op het belang van 'brede toetsing', waarbij zowel cognitieve als niet-cognitieve leerprestaties en zowel de leerdomeinen rekenen en taal als andere minder evident te toetsen leerdomeinen aan bod zouden komen. Hierdoor wordt de impact van het onwenselijke neveneffect van teaching to the

test geminimaliseerd. De respondenten vinden het daarom belangrijk om te streven naar een brede evaluatie van de verschillende leerdomeinen, gericht op het in kaart brengen van leerprestaties of leerwinst inzake onder meer (beroeps)competenties, attitudes en talenten van leerlingen. Bovendien werd in de literatuurstudie beschreven dat de resultaten op kennistoetsen (of toetsen van relatief eenvoudige vaardigheden) binnen de meest getoetste leerdomeinen (rekenen en taal) niet zonder meer kunnen veralgemeend worden naar andere vormen van output ('hogere vaardigheden', attitudes, competenties) of naar andere leerdomeinen. De ambitie om brede toetsing te realiseren, is evenwel niet eenvoudig. Er is dan ook verder onderzoek nodig naar de ontwikkeling van betrouwbare en valide toetsing van niet-cognitieve aspecten.

Daarnaast blijkt uit elk van de drie onderzoeksluiken de meerwaarde van adaptieve toetsen voor toetsen die gebruikt worden in een ontwikkelingsgericht perspectief groot te zijn. Bij een adaptieve toets wordt het niveau van de aangeboden toetsitems maximaal aangepast aan het kennis- of vaardigheidsniveau van elk van de individuele leerlingen. Voordelen zijn dat de toetsen een grotere uitdaging bieden voor elke leerling en dat het leer(winst)monitoringssysteem meer verfijnde informatie kan opleveren over de leerprestatie of leerwinst van elke leerling. Een voorwaarde hiervoor is dat de toets digitaal wordt afgenomen, wat door het doorgaans ontwikkelingsgericht karakter van deze toetsen evenwel geen probleem kan zijn (bij sterk verantwoordingsgerichte toetsen bleken klassieke pen-en-papier toetsen meer aangewezen om de (perceptie van) gelijkheid tussen leerlingen niet te bedreigen).

Ondersteuning

Zowel in de literatuurstudie als in elk van de cases van het internationale onderzoeksluik werd benadrukt dat de ontwikkeling van een leer(winst)monitoringssysteem op zich niet voldoende is. De vooropgestelde impact kan maar bereikt worden indien bij de gebruikers (leraren en scholen) aan bepaalde voorwaarden is voldaan. Het ruime aanbod aan Vlaamse en internationale onderzoeksliteratuur in het schooleffectiviteitsonderzoek bevestigt dat er grote verschillen zijn tussen scholen inzake datagebruik. Een belangrijke verklaring voor deze verschillen ligt in het beleidsvoerend vermogen en de datageletterdheid van scholen. Om deze voorwaarden in te vullen, bleek zelfs gerichte ondersteuning van scholen noodzakelijk in onderwijssystemen waar scholen reeds een langere traditie hebben in het omgaan met outputgegevens inzake de leerprestaties van leerlingen.

Draagvlak creëren

Ten slotte kan uit de literatuurstudie en uit het internationale onderzoeksluik worden geconcludeerd dat er inspanningen nodig zijn om bij de uiteenlopende stakeholders een draagvlak te creëren voor de ontwikkeling en het gebruik van een leer(winst)monitoringssysteem. Om dit draagvlak te creëren is het raadzaam rekening te houden met de bovenstaande conclusie dat Vlaanderen zulk systeem in een verantwoordingsgericht perspectief momenteel wenselijk noch haalbaar acht. De aanvaarding van de toetsen zelf hangt volgens de resultaten van de literatuurstudie onder meer af van de betrokkenheid van leraren in de ontwikkeling van de toetsen en van de gepercipieerde betrouwbaarheid en validiteit.

Uit het internationale onderzoeksluik kan worden afgeleid dat een draagvlak voor een leer(winst)monitoringssysteem kan groeien naarmate gebruikers inzien in welke mate de resultaten ervan kunnen bijdragen tot de onderwijspraktijk.

Hierdoor wordt het belang van een focus op micro- en mesoniveau opnieuw bevestigd. Verschillende respondenten in de Delphi-studie pleiten er dan ook voor om relatief kleinschalig te beginnen met een groep van scholen die hier aan willen deelnemen, zodat de meerwaarde kan worden aangetoond en de impact kan worden geëvalueerd. Op basis hiervan kan een draagvlak worden gecreëerd. Ten slotte leren we ook uit het internationale onderzoeksluik dat het wellicht onmogelijk is om alle critici ervan te overtuigen dat de positieve effecten van een leer(winst)monitoringssysteem opwegen tegen de mogelijke onwenselijke neveneffecten.

BELEIDSAANBEVELINGEN

Onderwijsonderzoek kan niet in de plaats van het politieke beleid treden. Het beleid neemt beslissingen en treft maatregelen. Daarbij laat het zich inspireren en leiden door een geheel van kennis, waarden en opvattingen inzake onderwijs, individu en samenleving. Wat beleidsondersteunend onderzoek doet is op een wetenschappelijk verantwoorde wijze gegevens verzamelen, ordenen en in samenhang presenteren. In deze studie betreft het gegevens die beleidsmakers kunnen inbrengen in hun beslissing of, en zo ja volgens welke modaliteiten, in het Vlaamse onderwijs een (verder) uitgebouwd systeem van leer(winst)monitoring opzetten wenselijk en haalbaar is.

Vertrekkende vanuit bovenstaande positionering willen we desalniettemin de kans nemen om ook als onderzoeksteam standpunt in te nemen. Dat standpunt hebben we vertaald in principes die in het uitdenken van een beleid inzake leer(winst)monitoring van wezenlijk belang zijn. In wat volgt beschrijven we vanuit het voorliggende onderzoek de uitgangspunten op basis waarvan leer(winst)monitoring(systemen) ons inziens van betekenis kan zijn in het Vlaamse onderwijs. Het realiseren van deze uitgangspunten is wat wij aanbevelen. Daarom hebben we ze als imperatief geformuleerd.

Gegeven de aard van de onderzoeksopdracht richten we ons in onze aanbevelingen in eerste instantie tot beleidsmakers in het Vlaamse onderwijs. Dat neemt niet weg dat wij schoolleiders en leerkrachten als de eerste verantwoordelijken beschouwen voor het behouden en verbeteren van onderwijskwaliteit. Voorbij gaan aan het schoolniveau zou er toe leiden dat we aanbevelingen in een vacuüm zouden presenteren. Hoewel we ons richten tot beleidsmakers zal de link met de school- en klaspraktijk dan ook een rode draad doorheen de aanbevelingen blijken te zijn.

De principes zijn niet zonder meer nieuw, en bouwen verder op empirische evidentie met betrekking tot leer(winst)monitoring(systemen) in verschillende onderwijscontexten, deze studie inclusief. Daarnaast spelen de aanbevelingen in op de inschatting van de wenselijkheid en haalbaarheid van mogelijke werkszenario's voor leer(winst)monitoring(systemen) voor Vlaanderen door een breed veld van stakeholders (cf. de Delphi-studie). Doelbewust richten we ons in de aanbevelingen nog op de eerder grote vraagstukken, wetende dat verdere uitwerking nodig is, dat meer verfijnde lessen uit deze studie

mogelijk (en beschikbaar) zijn en dat antwoorden op de grote vraagstukken nieuwe vragen zullen oproepen. Het zijn echter precies de fundamentele aspiraties en mogelijkheden die in het beleidsdebat eerst dienen uitgeklaard te worden.

Aanbeveling 1: Kijk en beslis voorbij de karikatuur

Het voorliggende onderzoek heeft aangetoond dat de problematiek van leer(winst)monitoring niet eenduidig is. In de bestudeerde landen is het politieke en academische debat nog steeds actueel, en laten bestaande systemen nog steeds disfuncties zien. Het gevaar is dan ook groot dat het debat in Vlaanderen verengd zou blijven tot een discussie over de wellicht meest extreme vorm van externe publicatie, namelijk in de vorm van rankings zoals die o.a. in de Britse context gangbaar is. Tegenstanders van vormen van centrale toetsing in Vlaanderen hebben het waardeoordeel over leer(winst)monitoringsystemen te vaak verengd tot zulke onwenselijke uitwassen in andere onderwijssystemen. Op die manier hebben ze een toon gezet die een sereen debat over voor- en nadelen bemoeilijkt heeft. Vanuit wetenschappelijk oogpunt pleiten wij voor het duidelijk kaderen van de problematiek in een bredere onderwijscontext en voor een analyse zonder a priori's. Een eerste evidente aanbeveling aan het Vlaamse onderwijsveld is dan ook goed kennis te nemen van de inhoud van dit onderzoeksrapport, om van daaruit tot geïnformeerde oordeelsvorming te komen. Dat vergt een bereidheid om taboedoorbrekend bestaande denkbeelden en overtuigingen te expliciteren, te confronteren met bestaande evidentie, in vraag te stellen en zo nodig te herformuleren. De variatie in verschijningsvormen en werkscenario's inzake leer(winst)monitoring maakt duidelijk dat het debat veel rijker en genuanceerder is dan het tot dusver in Vlaanderen gevoerd werd. Kortom: als het Vlaamse onderwijssysteem niet oplet, dan loopt ze het gevaar het goede en bruikbare te verwerpen omdat het niet onderscheiden wordt van het misleidende en onwenselijke.

Vlaanderen is vooralsnog terughoudend geweest in het uitwerken van systemen voor leer(winst)monitoring en van het openbaar maken van outputinformatie over scholen. Wij hanteren graag de invalshoek dat deze omzichtige houding Vlaanderen (internationaal) in een unieke context gebracht heeft. Het onderwijssysteem heeft de kans - en zonder meer ook de plicht - om te onderzoeken hoe de voordelen van een leer(winst)monitoringsysteem kunnen benut worden zonder de nadelen ervan te hoeven ondergaan. Vlaanderen dient de ambitie te hebben om de (zogenaamde) voorsprong van andere onderwijssystemen om te buigen tot een situatie waarin 'de wet van de remmende voorsprong' van toepassing is. Vanuit de in deze studie gedocumenteerde systemen, hun randvoorwaarden en (neven)effecten kan een systeem uitgedacht worden dat vanuit een Vlaams perspectief efficiënt en effectief is. Door middel van een sterk bewustzijn van de gedocumenteerde valkuilen kan een leer(winst)monitoringsysteem gecreëerd worden waarvan de meerwaarde en performantie algemeen erkend wordt. Dat vergt bereidheid om incrementeel te werken, bereidheid tot investering en de noodzaak tot het creëren van een juridisch kader dat mogelijke nadelen (met name inzake openbaarheid) onder controle houdt.

Aanbeveling 2: Bouw op vanuit de beoogde doelstellingen

Eén van de conclusies van deze studie is dat de impact van een leer(winst)monitoringssysteem het grootst is wanneer aan de verschillende onderdelen een éénduidig doel wordt gekoppeld. Het expliciteren van de rationale en assumpties achter mogelijke systemen moet dan ook het vertrekpunt zijn van het beleidsdebat. Eerst dient op een transparante wijze geduid te worden wat men wil bereiken om pas vervolgens de vraag te stellen hoe dat kan bereikt worden. We toonden aan dat een breed pallet aan doelstellingen ten gronde kan liggen aan leer(winst)monitoringsystemen. Het spreekt voor zich dat de uitrol anders is wanneer men bijvoorbeeld gelijkwaardigheid van diploma's overheen scholen wil garanderen dan wanneer men concrete onderwijs- en leerprocessen in de klas wil informeren. We toonden aan dat zowel verantwoordingsgerichte als ontwikkelingsgerichte systemen bepaalde positieve impulsen kunnen geven aan onderwijskwaliteit. Over de mogelijkheden om beide doelstellingen te combineren in één systeem zijn buitenlandse lessen echter eerder pessimistisch. We willen aanbevelen om met verschillende partners in kaart te brengen over welke doelstellingen overeenstemming kan bereikt worden. Vervolgens dient te worden nagegaan of bestaande systemen deze doelen realiseren en wat in termen van leer(winst)monitoring verder dient uitgebouwd te worden. Wij zien in het Vlaamse onderwijs te vaak beargumenteerd worden dat bestaande systemen volstaan zonder grondige analyse van de doelstellingen die deze systemen al dan niet realiseren. Zo is het bijvoorbeeld de vraag in welke mate koepelgebonden initiatieven scholen en leraren momenteel in staat stellen om de leerwinst van hun leerlingen in kaart te brengen en over jaren heen op te volgen. Het Vlaamse onderwijssysteem dient zich ervan bewust te zijn dat het behoud van de status quo ook nadelen heeft. In dat geval kiest men er immers voor om waardevolle doelstellingen niet te realiseren.

De vraag naar de wenselijke doelstellingen van en leer(winst)monitoringsystemen is geenszins een geïsoleerd vraagstuk. De vraag is gekoppeld aan andere beleidsdossiers die op elk op hun eigen manier deze discussie kunnen beïnvloeden, maar ook op hun beurt onvermijdelijk beïnvloed worden door de beslissingen die inzake de implementatie van een leer(winst)monitoringssysteem zullen worden genomen. Daarbij denken we bijvoorbeeld aan het debat rond eindtermen, implementatie Mdecreet, loopbaandebat, en het debat over de professionalisering van schoolbesturen. Aanbeveling 3: Stel ontwikkeling centraal, niet een enge focus op verantwoording

De onderzoeksresultaten leren dat zowel schoolontwikkeling als verantwoording een uitgangspunt kunnen zijn voor leer(winst)monitoring. Onze aanbeveling is evenwel om in het uitwerken van systemen zonder meer de nadruk te leggen op leerprocessen en de ontwikkeling die daarmee gepaard gaat. De focus op individueel (leerlingen), professioneel (leraren) en organisationeel (scholen) leren is van cruciaal belang. Leer(winst)monitoring met een te sterke focus op controle kan er toe leiden dat het leerpotentieel op diverse niveaus niet ten volle benut wordt. Over het algemeen blijkt bij diverse stakeholders in de Vlaamse onderwijscontext een positieve houding ten aanzien van ontwikkelingsgerichte leer(winst)monitoring. Het is dan ook aangewezen deze positieve houding verder te koesteren en te bewaken dat schoolleiders en leerkrachten autonoom gemotiveerd blijven om informatie te gebruiken in het kader van hun beleidsvoering en onderwijspraktijk.

We willen benadrukken dat gestandaardiseerd toetsen en een zogenaamde hoge inzet (high-stakes) niet hoeven samen te gaan. Een directe eenzijdige koppeling van de uitkomsten van

gestandaardiseerde toetsen aan beslissingen over voortgang/zittenblijven van leerlingen, toegang tot vervolgstudies, salarissen of bonussen voor personeelsleden maar ook beloningen of (publieke) sancties voor scholen dient vermeden te worden. Ons advies is dus om leer(winst)informatie vooral formatief in plaats van (enkel) summatief te gebruiken. In de leergerichte context die we beogen, wordt informatie in eerste instantie verzameld om het leren van de leerling, leraar en de school te informeren in een proces van (school)ontwikkeling en niet (enkel) om tot een (eind)oordeel te komen. Dat betekent ook dat in het ter verantwoording roepen van scholen en leraren ons inziens best gefocust wordt op de wijze waarop gebruik van leer(winst)monitoring hen tot (school)ontwikkeling brengt, eerder dan verantwoording te vragen voor de (output)resultaten die in het kader van het monitoringsysteem in kaart worden gebracht.

Wanneer scholen aan informatie uit gestandaardiseerde toetsen op leerlingniveau een rol willen toekennen in attestering en advisering, dan kan dit ons inziens enkel in een context waarin tevens de inbreng van de professionele leraar en van de delibererende klassenraad gegarandeerd blijft. Dat neemt niet weg dat we scholen willen stimuleren om - complementair aan de bestaande aanpak - van gestandaardiseerde schooloverstijgende toetsen gebruik te maken als referentiekader om aansluiting te (blijven) vinden bij de van leerlingen verwachte minimumprestaties.

Aanbeveling 4: Bouw (verder) aan een informatierijke omgeving voor scholen (met inbegrip van leerwinstinformatie)

De vergelijking met cases in de landenstudie leert dat er in het Vlaamse onderwijs - ondanks recente initiatieven zoals 'Dataloop' - relatief weinig informatie beschikbaar is. Dat geldt bij uitstek voor informatie over leerwinst. Op basis van onze bevraging van Vlaamse stakeholders concludeerden we dat men rekent op een leer(winst)monitoringssysteem om op microniveau bijkomende betrouwbare en valide informatie te genereren waardoor leraren in staat worden gesteld om tot gelijkgerichte inschattingen van leerlingprestaties te komen, om het onderwijsaanbod maximaal af te stemmen op de noden van de leerlingen maar ook om leerlingen (en hun ouders) zelf beter inzicht te geven in hun eigen voortgang. Op mesoniveau is de verwachting dat het leer(winst)monitoringssysteem een bijdrage levert aan een informatierijke omgeving voor de school door het systeem van interne kwaliteitszorg van betrouwbare en valide informatie over de leerprestaties of de leerwinst van leerlingen te voorzien. Om dit doel op mesoniveau te verwezenlijken, dient de informatie over de school in perspectief te worden geplaatst met outputinformatie van scholen met een vergelijkbare input en context. We stellen een eerder grote consensus vast rond deze idee dat ijkpunten voor informatie meer betekenis geven aan informatie, en rond de vaststelling dat vergelijkende informatie momenteel (te) schaars is. Door alle stemmen aan het woord te brengen, toont deze studie aan dat de gedragenheid voor deze ideeën in het Vlaamse onderwijsveld globaal genomen aanzienlijk groter is dan beleidsdebatten nogal eens laten uitschijnen.

Indien we scholen willen stimuleren tot informatiegebruik, dan strekt het tot aanbeveling om voor scholen een meer informatierijke omgeving uit te bouwen. Het is zaak om op verschillende niveaus te investeren in relevante, valide en betrouwbare, en gebruiksvriendelijke leer(winst)informatie. De

ontwikkeling, verspreiding en het faciliteren van gebruik van gestandaardiseerd toetsmateriaal zou met andere woorden een collectief beleidsstreven moeten zijn.

Aanbeveling 5: Verplicht het principe van leer(winst)monitoring, niet de keuze voor een systeem

Met het 'Decreet betreffende de kwaliteit van onderwijs' werd benadrukt dat instellingen zelf de eerste verantwoordelijken zijn voor de eigen kwaliteit en daartoe een zelfstandig beleid moeten voeren. Het decreet verplicht scholen tot interne kwaliteitszorg maar blijft terughoudend in de invulling ervan. Het kwaliteitsdecreet beperkt zich tot de verplichting voor elke instelling om op systematische wijze haar eigen kwaliteit te onderzoeken en te bewaken en legt geen richtlijnen op over de organisatie ervan. Het luidt dat de instelling zelf kiest op welke wijze ze dit doet. Wij adviseren de overheid om een gelijkaardige terughoudendheid aan de dag te leggen in het uitvaardigen van inhoudelijke en organisatorische voorschriften inzake leer(winst)monitoring. Dat betekent dat scholen decretaal zonder meer de opdracht kunnen krijgen om leerprestaties en leerwinst van leerlingen in kaart te brengen, te onderzoeken en te bewaken, maar dat deze verplichting zich niet vertaalt in het opleggen van het gebruik van deze of gene systeem. Ondertussen beschouwen wij het als de opdracht van de overheid om in overleg met de onderwijsverstrekkers het instrumentarium ter beschikking te stellen dat scholen hierbij kunnen aanwenden en om vervolgens toezicht te houden op de mate waarin scholen hun decretale verwachting vorm geven.

Aanbeveling 6: Voorzie in een breed aanbod dat recht doet aan de veelzijdigheid van het leren van leerlingen en van de onderwijsopdracht van scholen

Uit de verschillende onderdelen van deze studie blijkt het belang van 'brede toetsing', waarbij zowel cognitieve als niet-cognitieve leerprestaties en zowel de leerdomeinen rekenen en taal als andere minder evident te toetsen leerdomeinen aan bod komen. Verder leest men in deze studie een pleidooi voor aandacht voor bijvoorbeeld 'leren leren' en sociale vaardigheden maar ook voor de uitbouw aan socio-emotionele vragenlijsten (welbevinden). Gestandaardiseerde toetsen hoeven niet beperkt te zijn (c.q. blijven) tot strikt cognitieve prestaties en hebben best ook betrekking op vaardigheden en attitudes. Dit vergt echter substantiële investeringen, omwille van de noodzaak aan verder onderzoek naar mogelijkheden om deze vaardigheden en attitudes op betrouwbare en valide manier in kaart te brengen, omwille van de kost ten behoeve van het ontwikkelen van vaardigheidstoetsen, alsook omdat de afname van zulke toetsen arbeidsintensiever is. Scenario's die een ruim geheel van toetsen over uiteenlopende leerdomeinen vooropstellen, worden inzake wenselijkheid positief beoordeeld. Een volwaardige inschatting van bereikte leerresultaten van leerlingen en van de gerealiseerde kwaliteit van scholen maakt ook ons inziens dat een beperking tot cognitieve leerprestaties een onwenselijke verenging van het streefbeeld zou zijn. Bovendien is een breed aanbod dé manier om tegemoet te komen aan één van de meest gehoorde kritieken op bestaande systemen, met name de verenging van het onderwezen curriculum. Wij adviseren om deze brede benadering van bij de start centraal te stellen en zijn ondertussen beducht voor een beleidsscenario waarbij in de drang om snel tot een opstap tot leer(winst)monitoring te komen, toch gewerkt zal worden met een beperkte set van cognitieve indicatoren.

Aanbeveling 7: Zet waar mogelijk in op leerwinst

In elke van deze aanbevelingen spreken we van een leer(winst)monitoringssysteem, om aan te duiden dat het zowel kan gaan om het documenteren van leerprestaties als van leerwinst. De vele gedocumenteerde voordelen inzake leerwinst, onder meer de verwachting dat hierdoor een veel grotere impact kan bereikt worden op het niveau van (de begeleiding van) het individuele leerproces van elke leerling, leiden ons evenwel tot de aanbeveling om, waar mogelijk, in te zetten op het in kaart brengen van de leerwinst van leerlingen. Uit de respons van de Vlaamse stakeholders blijkt dat het draagvlak van een monitoringssysteem door leerwinstinformatie sterk wordt vergroot. Het werkscenario waarin het meest expliciet wordt ingezet op de productie van leerwinstinformatie in een ontwikkelingsgericht perspectief werd door de Vlaamse respondenten sterk gewaardeerd. Een beperking van dit werkscenario was evenwel dat het niet strookt met de voorgaande aanbeveling, met name dat het aanbod geen brede toetsing inhield. Ervaringen uit de literatuur en het buitenland, maar ook van eerdere pogingen in de Vlaamse onderwijscontext, leren evenwel dat genereren van leerwinstinformatie in uiteenlopende leerdomeinen erg moeilijk kan zijn door het ontbreken van een graduele groeilijn in de kennis en vaardigheden van leerlingen in verschillende domeinen. Ten behoeve van de integratie van aanbeveling 6 en 7 dient daarom sterk geïnvesteerd te worden in debat omtrent welke vormen van leerwinstinformatie bij welke leerdomeinen beschikbaar zouden moeten zijn en in onderzoek en ontwikkeling om dit vervolgens mogelijk te maken.

Aanbeveling 8: Voorzie in een pallet aan initiatieven voor deskundigheidsontwikkeling

Het voorliggende onderzoek maakt duidelijk dat het noodzakelijk is te werken aan het menselijk en sociaal kapitaal van scholen met betrekking tot het gebruik van leer(winst)monitoringsystemen. Het internationale onderzoeksluik toont aan dat elders te vaak voorbij gegaan werd aan de nochtans onontbeerlijke expertise-ontwikkeling en expertisedeling bij gebruikers. Hoewel scholen hier zelf een wezenlijke rol in te vervullen hebben, willen we de overheid een aantal aanbevelingen meegeven waarbij we de beoogde professionalisering in een ruimer kader plaatsen. Het gebruik van leer(winst)monitoringsystemen past immers in een ruimer verhaal van gekende thema's als kwaliteitszorg, informatiegebruik en datageletterdheid, maar ook in een verhaal van ethisch datagebruik door alle betrokkenen binnen (en buiten) het onderwijsveld. Deskundigheidsontwikkeling wordt in eerste instantie best geïntegreerd in initiatieven die zich op deze thema's richten. Toch zijn ook meer gefocuste strategieën behulpzaam. Zo zou de overheid kunnen overwegen om (tijdelijk) middelen te reserveren in het werkpakket van scholen, specifiek voor interne kwaliteitszorg en gericht op gebruik van leer(winst)monitoringsystemen. Laat deze middelen dan na verloop van tijd deel uitmaken van de reguliere financiering. Verder lijkt ook het sterker integreren van informatiegebruik en werken met (resultaten van) gestandaardiseerde toetsen in het curriculum van opleidingen van (toekomstige) schoolleiders en leerkrachten verstandig. Prille ervaringen inzake schoolfeedbackgebruik toonden in Vlaanderen reeds de noodzaak en potentie aan om tezamen met een aanbod van gestandaardiseerde toetsen te voorzien in professionaliseringsactiviteiten voor verschillende doelgroepen. Hetzelfde geldt voor het aan scholen en schoolbesturen ter beschikking stellen van experts, waarop zij specifiek voor ondersteuning met betrekking tot gebruik van het leer(winst)monitoringsysteem een beroep kunnen doen. Dat kan bijvoorbeeld in de schoot van de pedagogische begeleidingsdiensten gebeuren. Met het oog op de transfer van inzichten met betrekking tot gebruik van leer(winst)monitoringsystemen is het

ook aangewezen om de uitbouw van voor scholen interne en externe thematische netwerken rond leer(winst)monitoring te ondersteunen.

Aanbeveling 9: Werk met pilootprojecten en monitor de werking van het systeem

Als de overheid overweegt om een leer(winst)monitoringsysteem op te zetten verdient het aanbeveling eerst een degelijk pilootproject uit te voeren. Het opzetten van een leer(winst)monitoringsysteem is een complexe onderneming die aan verscheidene facetten van de organisatie van scholen en leren raakt en waarbij zeer diverse stakeholders betrokken zijn. Kenmerken van het functioneren van het systeem en (beleefde) (neven)effecten laten zich niet (eenduidig) voorspellen. Bestaande (buitenlandse) systemen zijn onmogelijk zonder meer te transfereren naar een Vlaamse context. Wij adviseren de overheid daarom om van het opzetten van een leer(winst)monitoringsysteem een groeitraject te maken. Op die manier kan gezocht worden naar garanties om het systeem (in groeiende mate) aan de verwachte kenmerken en randvoorwaarden te laten voldoen. In deze groeifase kan vervolgens ook ingezet worden op informerende en draagvlakcreërende communicatie.

Een incrementele aanpak is ons inziens aangewezen. Het gebruik van een leer(winst)monitoringsysteem helpt namelijk zichzelf 'verkopen' indien het aanbod en de omkadering van bij het moment van globale uitrol kwaliteitsvol ingevuld wordt. Helaas kan het gebruik zichzelf bij wijze van spreken ook bemoeilijken tot onmogelijk maken indien dit niet het geval is. De regel 'baat het niet dan schaadt het niet' gaat niet op. Daarvoor is het prille draagvlak dat we in deze studie blootlegden te kwetsbaar en kostbaar.

Van bij de start dient verder een strategie voor monitoring naar de werking van het systeem ingebouwd te worden. Vanuit een overkoepelend kwaliteitszorgidee adviseren wij om vertrekkende vanuit de beoogde doelstellingen een set van prestatie- en perceptie-indicatoren te ontwikkelen en op te volgen. Op die manier kan het verbeterbare geïdentificeerd worden, kunnen mogelijke neveneffecten vroegtijdig getraceerd worden en kan verdere ontwikkeling van het leer(winst)monitoringsysteem geïnformeerd worden.