

Kansen en bekommernissen bij centrale toetsen: echo's uit het Vlaamse onderwijsveld

*Jan Vanhoof, Evelyn Goffin, Shana De Maeyer, Cin Eysermans en Sofie Snyers**



Vanaf schooljaar 2023-2024 zullen alle Vlaamse leerlingen tijdens hun schoolloopbaan vier keer een centrale toets afleggen. Het doel van deze studie was om de percepties en de verwachtingen van leraren, schoolleiders en pedagogisch begeleiders bij deze beleidsbeslissing in kaart te brengen. Aan de hand van 34 explorerend kwalitatieve semigestructureerde interviews werd daartoe empirische evidentie verzameld. Verwachte wenselijke en onwenselijke effecten van het voorziene systeem worden belicht. Daarnaast wordt ingegaan op een set van sleutelkenmerken die leerkrachten, schoolleiders en pedagogisch begeleiders vooropstellen zodat centrale toetsen een optimale kans op slagen zouden hebben bij verdere implementatie. De resultaten tonen aan dat vooral een ontwikkelingsgericht perspectief aanbevolen wordt. Verder wordt onder meer gepleit voor een brede toetsing, secure omgang inzake de openbaarheid van resultaten, en inzetten op de ontwikkeling van informatiegeletterdheid van toekomstige gebruikers.

I. Situering

De discussie over hoe onderwijskwaliteit te meten, te bewaken en te verbeteren, vormt het onderwerp van vele beleidsacties en initiatieven binnen schoolteams. Om kwaliteit binnen het onderwijs in kaart te brengen, opteren verschillende landen voor controlesystemen gebaseerd op gestandaardiseerde toetsen, die op regelmatige tijdstippen tijdens een volledige schoolcarrière de leerprestaties en de leerwinst van leerlingen meten (OECD, 2013). Een gestandaardiseerde toets houdt in dat de condities waaronder de toets wordt afgelegd en de inhoud voor elke leerling identiek zijn. Deze toetsen komen in diverse vormen voor, onder andere inzake verplichting of vrijwilligheid van afname, het doel van de toets, de frequentie van afname, de getoetste leerdomeinen en de toetsvorm. In het geval van een centrale toets zijn de toetsen verplicht en bepaalt

de overheid het moment waarop leerlingen de toetsen afleggen. Ook staat de overheid in voor het bepalen van de manier waarop gescoord en geëvalueerd wordt. Een monitoringsysteem op basis van gestandaardiseerde toetsen is een aangewezen middel om onderwijskwaliteit te bewaken en te verbeteren (BAIRD, AHMED, HOPFENBECK, BROWN en ELLIOT, 2012; VAN GASSE, VANHOOF, MAHIEU en VAN PETEGEM, 2015).

Er bestaat ook in Vlaanderen al langer een behoorlijk breed spectrum aan gestandaardiseerde toetsen, allemaal met een ander doel of opzet, die in Vlaamse scholen (op vrijwillige basis) worden afgenomen. Denk daarbij aan de periodieke peilingen en paralleltoetsen, koepelgebonden toetsen en oriënteringstoetsen. Deze gestandaardiseerde toetsen hebben echter een wisselende focus waardoor de resultaten van deze metingen een weinig systematisch beeld geven van de schoolwerking. Daarnaast nemen Vlaamse scholen (vrijwillig) deel aan

* Alle auteurs zijn verbonden aan de Universiteit Antwerpen.

verschillende internationaal vergelijkende studies die gericht zijn op de meting van leerlingenprestaties. Ook via deze internationaal vergelijkende prestatietingen (zoals TIMSS, PIRLS, PISA) verkrijgt men informatie over leeruitkomsten op systeem- en schoolniveau. De lacune aan objectieve data in Vlaanderen is er evenwel vooral op het vlak van het meten van leerwinst bij de leerlingen, waardoor SHEWBRIDGE *et al.* (2011) stellen dat het Vlaamse onderwijs gebaat is met het ontwerpen van centrale toetsen die vergelijkbare gegevens verstrekken over een aantal kerngebieden.

Tot voor kort was Vlaanderen een van de weinige Europese onderwijssystemen waarin geen vorm van centrale toetsing op grote schaal werd geïmplementeerd. Een van de redenen hiervoor is dat de Vlaamse onderwijscontext wordt gekenmerkt door een hoge mate van autonomie van scholen (OECD, 2013). Deze autonomie wordt bepaald door het grondwettelijk verankerde principe van 'vrijheid van onderwijs'. Dit principe geeft scholen de autonomie om volgens een eigen pedagogisch project naar de decretaal vastgelegde minimumdoelen toe te werken. De visie dat verschillen tussen scholen mogelijk en wenselijk zijn en dat deze net zorgen voor een rijk onderwijsaanbod, is in Vlaanderen sterk aanwezig. Daar staat tegenover dat de Vlaamse Regering recent beslist heeft dat vanaf 2023 alle leerlingen in Vlaanderen in de loop van hun schoolcarrière vier keer een centrale toets zullen afleggen. De toetsen zullen worden afgenomen aan het einde van het 4de en 6de leerjaar basisonderwijs en aan het 2de en 6de jaar secundair onderwijs en zullen in eerste instantie focussen op Nederlands (begrijpend lezen, schrijven en grammatica) en op wiskunde. Het doel van de toetsen is nagaan in hoeverre leerlingen eindtermen beheersen en scholen erin slagen leerwinst te boeken (WEYTS, 2019).

Het algemene principe van het ontwikkelingsgerichte gebruik van gestandaardiseerde toetsen maakt opgang in Vlaanderen (PENNINCKX *et al.*, 2017; DE MAN en DENYS, 2021). Er weerklinken het voorbijende decennium steeds meer stemmen die het wenselijk achten om op een systematische en gestandaardiseerde manier de leerresultaten van leerlingen in kaart te brengen. De meningen van uiteenlopende stakeholders in het Vlaamse onderwijsveld over de meest geschikte manier van vormgeven en invoeren van centrale toetsen zijn echter verdeeld. De discussie over waarom, wat en hoe te toetsen, is volop gaande. Dit gaat dan niet zozeer over het principe van gestandaardiseerd toetsen, wel over hoe de resultaten net wel of net niet te gebruiken. Bovendien zijn de details van hoe deze uitrol vorm zal krijgen vooralsnog nog niet ingevuld. Dit maakt de implementatie van centrale toetsen om leeruitkomsten bij leerlingen te meten nog steeds erg omstreken. Dat blijkt onder meer ook uit de adviezen die de Vlaamse Onderwijsraad hieromtrent uitbracht (VLOR, 2021).

De implementatie van een systeem van centrale toetsen heeft vooral kans op slagen als ze breed wordt gedragen door de verschillende stakeholders (SCHILDKAMP en KUIPER, 2010; HEMKER *et al.*, 2021; VLOR, 2021). Onderzoek naar de houding van het Vlaamse onderwijsveld is echter erg schaars, aangezien het nog over een zeer recente en deels 'hypothetische' situatie gaat. Nu het invoeren van centrale toetsen in Vlaanderen evenwel een feit is, zijn het de scholen en hun pedagogisch begeleiders die in eerste instantie betrokken partij zijn. Zowel in de aanloop naar, als in de nasleep van de beslissing tot invoering van centrale toetsen werden de opinies van deze actoren uit het Vlaamse onderwijsveld nog niet op systematische manier onderzocht. Hun houding ten opzichte van de bruikbaarheid van centrale toetsen zal, zo leert onderzoek (bv. PENNINCKX *et al.* (2017)), nochtans meebepalen in welke mate beoogde effecten op het vlak van kwaliteitsverbetering in de klas en op het niveau van het onderwijssysteem worden bereikt. Hun opvattingen omtrent het debat over centrale toetsen zijn onontbeerlijk om de conceptualisering van de centrale toetsen en de implementatie ervan de komende jaren op een positieve en gedragen manier verder uit te werken. Er zijn weliswaar de adviezen van de Vlaamse Onderwijsraad over dit thema (VLOR, 2021), maar hoewel verschillende actoren in deze raad vertegenwoordigd zijn, hebben deze adviezen niet de ambitie om empirische evidentie te verzamelen. Om die reden draagt deze exploratieve studie bij aan een breder begrip van de houding van diverse stakeholders in het Vlaamse onderwijsveld ten aanzien van deze centrale toetsen. Een verkenning en een uitdieping van bestaande visies, gepercipieerde kansen en bestaande bekommernissen is het schragende projectdoel.

Gegeven het bovenstaande, rapporteren we in deze bijdrage over de houding van ruim dertig leraren, schoolleiders en pedagogisch begeleiders ten opzichte van de invoering van centrale toetsen in het Vlaamse onderwijs. We beschrijven welke wenselijke en onwenselijke vormen van gebruik zij associëren met de centrale toetsen zoals die op dit moment door de overheid uitgetekend worden. Verder bouwend op gewenste effecten, brengen we in kaart hoe een ideaal systeem van centraal toetsen er volgens deze groepen van respondenten uitziet.

II. Thematische afbakening

Deze bijdrage brengt percepties van kernstakeholders uit het Vlaamse onderwijsveld in kaart, zonder deze percepties te willen veralgemenen. Deze percepties worden neutraal beschreven en niet-normatief geïnterpreteerd, zonder ze dus systematisch te willen toetsen aan de bestaande ruimere kennisbasis van werken met centrale toetsen. Dat verklaart mede waarom we deze bijdrage niet starten met een uitvoerig overzicht van eerder empirisch onderzoek. Voor een inventarisatie van elders beschreven (neven)

effecten en randvoorwaarden verwijzen we naar het overzicht dat PENNINGX *et al.* (2017) compileerden. Toch willen we hier kort zicht geven op een aantal oriënterende concepten en theoretische dimensies die het gesprek met de respondenten (via de interviewleidraad) en de analyse (via de gebruikte codeboom) gestructureerd hebben.

II.1. Wenselijk en onwenselijk gebruik van centrale toetsen in de klas en de school

Het vertrekpunt zijn mogelijke (effecten van) vormen van gebruik van centrale toetsen. Dit gaat in wezen terug op de vraag waarom te investeren in vormen van centrale toetsen. Daarbij is voor deze studie vooral het onderscheid tussen effecten op klas- en schoolniveau zinvol. Het klasniveau heeft daarbij betrekking op resultaten van individuele leerlingen en op hoe leraren gebruikmaken van de resultaten van de leerlingen in hun klas. Het schoolniveau omvat datgene wat de leerling en de klas binnen de school overstijgt.

Voorstanders menen bijvoorbeeld dat gestandaardiseerde toetsen op klasniveau kunnen zorgen voor een nauwkeurige, objectieve beoordeling van leerlingprestaties die niet gekleurd wordt door vooringenomenheid op basis van bijvoorbeeld gender, socio-emotionele achtergrond of inzet in de klas (BAIRD *et al.*, 2012; MARLOW *et al.*, 2014). Daarnaast is het hanteren van een 'gelijke lat' voor alle leerlingen bij een verantwoordingsperspectief (summatieve eindtoetsen) volgens sommigen niet alleen belangrijk voor de leerlingen zelf maar ook voor het mogelijk vervolgonderwijs en potentiële werkgevers (BISHOP en WÖSSMANN, 2004). Wenselijke effecten in de klas hebben ook een keerzijde (STOBART, 2008; CHAKRABARTI en SCHWARTZ, 2013). Een vaak vermeld neveneffect is dat leerkrachten hun onderwijs- of evaluatiepraktijk te eenzijdig aanpassen aan de toets om zo leerlingen een beter toetsresultaat te laten bekomen (*cf. teaching to the test*). Zo kan men bepaalde leerinhouden die in de toets voorkomen, overbeklemtone ten koste van andere leerinhouden (BAIRD *et al.*, 2012; JANSSENS, REKEN-MOMBARG en LACOR, 2014).

Wat betreft het schoolniveau geven MARSH, PANE en HAMILTON (2006) aan dat centrale toetsen een indicatie kunnen geven van sterktes en pijnpunten binnen de school en zo de aanzet tot verbetertrajecten. Net in de koppeling tussen geïnformeerde interne kwaliteitszorg met professionele ontwikkeling van schoolteams schuilt in die logica een belangrijke drijfveer (KLEIN en VAN ACKEREN, 2012). Daar staat tegenover dat sterk verantwoordingsgericht gebruikmaken van resultaten van centrale toetsen (zie verder) tot onwenselijk strategisch gedrag kan leiden in scholen. De meest onwenselijke vorm is bewust en gepland vervalsen. De meest voorkomende vorm (AMREIN-BEARDSLEY *et al.*, 2010) is echter het meer

subtiele, toevallige en niet op voorhand geplande helpen van leerlingen bij toetsafname waardoor validiteit van het toetsresultaat in gedrang komt. Denk daarbij bijvoorbeeld aan toetsvragen voorlezen, moeilijke woorden verklaren, didactische posters laten hangen of meer tijd geven dan voorzien. Eerder dan via goede onderwijspraktijken kan men dus mogelijks ook op zulke niet-duurzame manieren tot verbeterde toetsresultaten willen komen.

II.2. Sleutelkenmerken van systemen van centrale toetsing

Naast het inventariseren van mogelijke effecten beoogt deze studie ook in kaart te brengen hoe de bevraagde Vlaamse stakeholders invulling geven aan een aantal concrete maar impactvolle beslissingen bij het vormgeven van systemen van centrale toetsing. De volgende zeven oriënterende vragen golden als vertrekpunt om naar de overtuigingen van de bevraagde respondenten te peilen. Naast deze structurerende thema's legden we in de gesprekken een grote openheid aan de dag voor andere elementen die respondenten naar voren wilden schuiven.

Ontwikkelings- en/of verantwoordingsgericht?

Centrale toetsen kunnen op schoolniveau zowel een ontwikkelings- als verantwoordingsdoel nastreven (WANG *et al.*, 2006) en er wordt verondersteld dat een verantwoordingsdoel indirect kan leiden tot een ontwikkelingsdoel (LOONEY, 2014). Daartegenover staat dat bij het verantwoordingsperspectief het gevaar bestaat dat scholen de resultaten op oneerlijke wijze proberen te verbeteren, waardoor ze meteen ook niet meer of minder bruikbaar zijn vanuit ontwikkelingsperspectief (SMITH, 2014; SAUNDERS, 1999). Ook op klas- en leerlingniveau kunnen centrale toetsen formatief-ontwikkelingsgericht (met het oog op bijsturen en differentiëren) of summatief-verantwoordingsgericht worden ingezet (met het oog op beslissen over studieloopbaan van leerlingen of toegang vervolgonderwijs). Hier gaat het om de vraag wie op welke manier gebruik kan maken van de resultaten. Blijven deze vertrouwelijk binnen de school of kan bijvoorbeeld ook de onderwijsinspectie de toetsresultaten een plaats geven in haar doorlichtingsaanpak?

Verplicht of facultatief? Hoewel de Vlaamse centrale toetsen in principe verplicht zullen zijn, zijn ook hier in theorie variaties mogelijk. De keuzevrijheid om centrale toetsen bij alle, dan wel slechts bij specifieke leerlingen (groepen) af te nemen zou bijvoorbeeld ook op schoolniveau kunnen liggen. Daar staat tegenover dat het willen vergelijken van scholen wellicht betekent dat deelname van klassen en leerlingen voldoende centraal aangestuurd moet worden.

Brede of kerngerichte focus? Centrale toetsen kunnen al dan niet een brede focus hebben. Zo is een vraag bijvoorbeeld hoe de verdeling over kennis, vaardigheden en attitudes

ingevuld dient te worden. Daarnaast is het de vraag of klassieke leerdomeinen zoals Nederlands en wiskunde volstaan, dan wel of ook andere leerdomeinen in de focus zouden moeten zitten. In onderzoek naar schooleffectiviteit wordt bijvoorbeeld aangeraden kennis en vaardigheden binnen verschillende leerdomeinen te toetsen omdat cognitieve resultaten op rekenen en taal niet zonder meer veralgemeend kunnen worden naar andere domeinen (OPDENAKKER en VAN DAMME, 2000).

Afstemming op eindtermen of leerplannen? De afstemming van centrale toetsen op het curriculum genereert veel bruikbare informatie voor het micro- en mesoniveau, aangezien leerkrachten hun lespraktijk op dit curriculum afstemmen (GOODMAN en HAMBLETON, 2005; WANG *et al.*, 2006). Inzetten op eindtermgebaseerde toetsen kan in Vlaanderen echter ook tot een aantal moeilijkheden leiden. Scholen (uit verschillende onderwijsnetten) bieden niet altijd op hetzelfde tempo of volgens dezelfde planning leerstofonderdelen aan doorheen de schoolloopbaan van leerlingen. Daarnaast zullen de toetsitems bij elke curriculumwijziging moeten worden nagekeken en aangepast. Dat laatste geldt uiteraard zowel voor afstemming op eindtermen als op leerplannen.

Leerwinst en/of toegevoegde waarde? Leerwinst is een maat voor de groei van individuele leerlingen, in het algemeen of binnen een specifiek leerdomein (JANSSENS *et al.*, 2014). Leerwinst wordt bijvoorbeeld berekend op basis van twee metingen, het verschil tussen beide is de leerwinst. Een systeem van centrale toetsing kan daar stoppen of ook de ambitie koesteren om de bijdrage van de school aan de leerwinst van haar leerlingen zichtbaar te maken. In dat geval spreekt men over de toegevoegde waarde van scholen. Door leerwinst zo goed mogelijk te corrigeren voor factoren die buiten de directe invloed van scholen liggen (zoals socio-economische status van leerlingen), wordt het effect van de school zichtbaar (JANSSENS *et al.*, 2014; AMREIN-BEARDSLEY, 2008).

Al dan niet openbaar maken van toetsresultaten? Het openbaar maken van kwaliteitsgegevens van scholen is een tendens die in de meeste landen toeneemt (OECD, 2013). Een vrijwel onvermijdelijk gevolg hiervan zijn de zogenaamde rankings van scholen die kaderen in een cultuur van ‘*naming and blaming*’ waarin de media scholen onderaan de klassering viseert, met demotivaties van scholen en leraren als gevolg (ELSTAD, 2009; SLOANE *et al.*, 2013). Er zijn echter ook argumenten die stellen dat openbaarheid en zelfs competitie tussen scholen niet negatief hoeft te zijn. Vanuit economisch standpunt zullen scholen door deze competitie streven naar de beste onderwijskwaliteit (BLACK en WILLIAM, 2005). Bovendien biedt informatie over de prestaties van diverse scholen ouders de kans om op een geïnformeerde basis een schoolkeuze te maken.

Hoe inzetten op gebruikersvaardigheden? De mate waarin centrale toetssystemen impact hebben op de interne kwaliteitszorg van de scholen, alsook de mate waarin scholen zich laten verleiden tot ongewenst strategisch gedrag, is sterk gelinkt aan het beleidsvoerend vermogen van scholen (PENNINCKX *et al.*, 2017). Ondertussen rapporteren verschillende onderzoekers dat behoorlijk wat scholen slechts in beperkte mate in staat zijn om de resultaten van dergelijke toetsen op een doeltreffende manier aan te wenden (HOPSTER-DEN OTTER *et al.*, 2017; VAN GASSE *et al.*, 2015). Dat roept de vraag op waarop ondersteuning en begeleiding zich precies dient te richten en hoe effectieve professionalisering er in de context van gebruik van data uit centrale toetsen dient uit te zien.

III. Gegevensverzameling

Opvattingen van een toevallige selectie van Vlaamse leraren, schoolleiders én pedagogisch begeleiders werden in kaart gebracht. Aan de hand van kwalitatief explorerend onderzoek en meer specifiek het afnemen van (semigestructureerde) interviews wilden we standpunten begrijpen en doordringen in de leefwereld van de respondenten (MORTELMANS, 2007). De opzet was vooral om informatie te verkrijgen over de kansen en de bekommernissen bij centrale examens die leven bij de bevroegde groepen respondenten, zonder expliciet te willen generaliseren naar de populatie.

Op basis van de elementen in de thematische afbakening werd een interviewleidraad opgesteld. De leidraad bevatte zoveel mogelijk open vragen, om te vermijden dat de respondenten in een bepaalde richting zouden worden geduwd. Voorbeeldvragen zijn: Welke meerwaarde verwacht u van het gebruik van centrale toetsen? Welke problemen in ons onderwijssysteem zou zulk een systeem kunnen helpen oplossen? Welke negatieve effecten vreest u? Ziet u kansen of eerder bedreigingen in het integreren van centrale toetsen binnen het huidige evaluatiebeleid van uw school? Er werd voor en tijdens het interview uitdrukkelijk benadrukt dat er geen juiste of foute antwoorden zijn, om sociaalwenselijke antwoorden te vermijden. De interviews werden opgenomen na toestemming van de respondent. Anonimiteit werd gegarandeerd via een consentformulier. De lengte van de interviews varieerde van 23 tot 100 minuten.

In totaal werden 34 interviews afgenomen. Daarvan vonden 12 interviews plaats bij leerkrachten uit verschillende vakgebieden in de eerste en tweede graad secundair onderwijs, met aandacht voor variatie in het aantal jaren ervaring en de finaliteit van het onderwijsaanbod in de school. Eenzelfde variatie werd nagestreefd in de 12 interviews die bij schoolleiders secundair onderwijs afgenomen werden. De respondentengroep bestond ten slotte ook uit 10 medewerkers van pedagogische begeleidingsdiensten.

Alle interviews werden eerst woordelijk getranscribeerd. Nadien werden de uitgeschreven gegevens met behulp van het softwarepakket NVivo 12 gecodeerd, met als vertrekpunt de codeboom die deductief op basis van het theoretisch kader opgesteld was. Voorbeeldcodes zijn: effecten (macro, meso, micro), doelen (ontwikkelingsgericht, verantwoordingsgericht), verplichte/facultatieve deelname, randvoorwaarden, draagvlak, openbaarmaking, toetsinhouden, toetsafname, toetsfrequentie. De codeboom werd vervolgens ook inductief aangevuld met een aantal nieuwe codes die uit de interviews naar voren kwamen. De interviews en de analyses werden uitgevoerd door drie onderzoekers die onderling afstemden over de interviewleidraad, het codeerproces en de analyseaanpak.

IV. Bevindingen

We organiseren deze resultatensectie rond de kernbevindingen zoals deze uit de interviews overheen de doelgroepen naar voren gekomen zijn. Er is eerst de vaststelling dat de door ons bevraagde respondenten het potentieel erkennen van het gebruik van centrale toetsen (zie IV.1.). Dat potentieel realiseren blijkt echter verre van eenvoudig te zijn en mogelijk fragiel. Het vergt het wegwerken van verscheidene bezorgdheden en het creëren van passende randvoorwaarden (zie IV.2.). Beide elementen worden hieronder uitgewerkt. Citaten uit de interviews illustreren de resultaten en expliciteren de link met het tijdens de interviews verzamelde empirische materiaal.

IV.1. De bevraagde respondenten zien kansen in centrale toetsen

Het potentieel van centrale toetsen zit, samengevat, in het beter informeren van beslissingen die in het onderwijsveld gemaakt worden.

IV.1.a. Centrale toetsen kunnen het beeld van de leraar over diens leerlingen aanvullen, maar niet vervangen

Centrale toetsen hebben hun waarde volgens respondenten in het ondersteunen van beslissingen van leraren over hun leerlingen en over hun onderwijspraktijk. Respondenten willen echter geen van allen dat resultaten van centrale toetsen direct bepalend zijn voor slaagkansen of studiekeuze van individuele leerlingen. Respondenten waarschuwen voor het gevaar van verkeerde oriënteringen op basis van de centrale toetsen. Een argument hier is dat centrale toetsen een momentopname zijn, bijgevolg niet breed genoeg om de volledige capaciteiten van leerlingen in kaart te brengen. Volgens de respondenten dient er bij het beoordelen van leerlingen ook gekeken te worden naar

het totaalplaatje en zijn bijvoorbeeld ook leermotivatatie en draagkracht cruciaal in de overweging. Respondenten verwachten ook dat bepaalde leerlingen bij te grote druk door stress en angst kunnen onderpresteren op de centrale toets. Zulk onderpresteren kan dan weer negatieve gevoelens uitlokken volgens hen, zoals een daling van het zelfbeeld en de leermotivatatie. Ook het idee dat leerlingen die niet gemotiveerd zijn zomaar wat invullen of helemaal niets invullen in de toets, leeft.

De schoolleiders en de leerkrachten geven aan dat een klassenraad het best de capaciteiten van leerlingen kan inschatten en kan aangeven welke evolutie die leerling gemaakt heeft. Wel zijn zij het erover eens dat de resultaten van centrale toetsen adviserend kunnen zijn bij studiekeuze. De resultaten van de centrale toetsen kunnen bijvoorbeeld worden meegenomen in oudercontacten, op voorwaarde dat ze met bijkomende informatie aangevuld worden. Tevens halen de respondenten aan dat de leerlingenbegeleiding door middel van de leerwinstmonitoring sterker afgestemd kan worden op de vastgestelde achterstand bij bepaalde leerlingen, zowel preventief als remediërend. Het doel van het systeem moet de groei van leerlingen zijn. Evaluatie en toetsing zijn het middel om daartoe bij te dragen.

“Om op die manier leerlingen te oriënteren naar een finaliteit of een hogere studie. Op die manier zal het wel opbrengen en een meerwaarde zijn. (...) Een extra middel om te adviseren.” (Schoolleider, respondent 7)

“Dan lijkt mij dat ook wel een sterk document om aan ouders laten zien, omdat dat niet gebonden is aan de scholen, niet gebonden is aan leerkrachten. Dus daar lijkt het me ook wel een krachtig middel voor.” (Leerkracht, respondent 4)

Respondenten erkennen dat centrale toetsen kunnen bijdragen aan een meer objectieve beoordeling van leerlingen. Ze geven aan dat tal van (niet-schoolgebonden) factoren onbedoeld de inschatting van de leraar over individuele leerlingprestaties kunnen kleuren. In lijn daarmee geven ook de geïnterviewde leraren (zelf) aan dat leerlingen nu vaak, afhankelijk van welke leerkracht ze voor zich hebben staan, andere resultaten behalen. Om zulke vertekeningen te vermijden, klinkt het dat het wenselijk is dat leraren hun inschatting vanuit de eigen professionaliteit aftoetsen aan de resultaten van centrale toetsen. Dat geeft hun de kans om na te gaan of deze resultaten hen al dan niet verrassen en om hun beeld van de leerlingen, indien nodig, bij te sturen. Zo fungeren de resultaten van centrale toetsen ook als een reflectie voor de leraar over het eigen inschattingsvermogen.

Meerdere schoolleiders zeggen de resultaten van de centrale toetsen te willen gebruiken om in gesprek te gaan met leerkrachten over hun sterktes, hun lespraktijk en over mogelijke werkpunten en bijhorende ondersteuning

of professionalisering. Enkele respondenten benadrukken daarbij dat een systeem van centrale toetsen evenwel moet bewaken dat leraren zich niet miskend weten in hun rol als expert in leren en evalueren. Anderzijds verwachten respondenten eveneens dat leraren net door de resultaten van de toetsen gevoelens van persoonlijke eer en bevestiging van hun dagelijks functioneren kunnen ervaren.

“Er zijn nogal wat leraren, daar moeten we niet flauw over doen, die natuurlijk een bepaalde vorm van verwachtingspatronen in leerlingen hebben en je kunt je afvragen in hoeverre dat evaluatie door leraren objectief is. (...) Dan zou je kunnen zeggen dat een gestandaardiseerd en gevalideerd systeem eigenlijk een objectieve maatstaf is die voor iedereen dezelfde is en waar iedereen op dezelfde manier wordt gemeten, dat je eigenlijk die subjectiviteit eruit haalt.” (PBD, respondent 5)

IV.1.b. Centrale toetsen kunnen een bijkomend perspectief bieden in de evaluatie van het schoolbeleid

Respondenten zien de centrale toetsen ook als een bijkomende bril voor scholen om meer doelgericht naar hun onderwijskwaliteit te kijken. Door de systematische toetsing over de jaren heen kan een centraal toetsingssysteem volgens respondenten zowel een bevestiging van de huidige aanpak bieden als een wake-upcall zijn om zo nodig bepaalde zaken anders aan te pakken. Een aantal respondenten beschouwen de centrale toetsen door de wetenschappelijke onderbouwing als een sterker evaluatie-instrument in vergelijking met zelfontwikkelde toetsen.

Al de schoolleiders en leerkrachten geven aan dat de resultaten van centrale toetsen een aanzet kunnen bieden tot verbetertrajecten binnen de interne kwaliteitszorg. Schoolleiders willen deze verbetertrajecten initiëren door vanuit het leerkrachtenteam zelf te onderzoeken op welke domeinen ondersteuning wenselijk is. Respondenten omschrijven het systeem van centrale toetsen als een positieve hefboom om met de interne kwaliteit aan de slag te gaan.

“Wat ga ik met die cijfers doen? Ik denk vooral kijken naar hoe goed geven wij onderwijs. Kunnen wij onze manier van lesgeven en onze manier van omgaan met kinderen verbeteren, zodanig dat leerlingen maximale leerwinst hebben? Ik denk dat dat nog altijd het doel is: hoe kunnen we ons verbeteren.” (Schoolleider, respondent 3)

Tevens wijzen respondenten op schoolniveau op het gegeven dat de centrale toets een momentopname is en als evaluatie-instrument bijgevolg enkel een deelelement kan zijn van een ruimer geheel van gegevens om kwaliteit te onderzoeken. Meerdere respondenten beamen dit en stellen dat enkel de centrale toetsen een te enge

focus zijn voor scholen om voldoende breed naar de onderwijskwaliteit te kijken.

IV.1.c. Centrale toetsen kunnen bovenschoolse beleidsbeslissingen informeren

Inzake kwaliteitsbewaking stellen respondenten dat ook de overheid en intermediaire organisaties met centrale toetsen op een meer objectieve en systematische manier aspecten van onderwijskwaliteit kunnen monitoren. Enkele respondenten vermelden bijvoorbeeld dat het systeem onderliggende problemen zoals de taaldiversiteit in het Vlaamse onderwijs kan blootleggen. Ze hopen dat de beleidsinformatie uit de centrale toetsen een aanzet kan zijn voor de overheid om deze achterliggende problemen de nodige aandacht te geven.

“Ik vind het goed dat er nagedacht wordt over hoe we kwaliteit kunnen controleren. Kunnen we er ergens een lijn in trekken of dat het al dan niet goed gaat met ons onderwijs. Bijvoorbeeld lezen hadden we misschien vroeger kunnen detecteren op die manier.” (PBD, respondent 9)

Ondanks het potentieel dat men ziet, zijn zowel voor- als tegenstanders van de centrale toetsen kritisch over druk van deze bijkomende investering op het Vlaamse onderwijsbudget. Zo vragen respondenten zich luidop af of deze toetsen wel voldoende meerwaarde gaan creëren in vergelijking met de output van de huidige gevalideerde toetsen en internationale onderzoeken. Daarnaast zijn een aantal pedagogische begeleiders van mening dat de reeds jarenlange deelname aan internationale onderzoeken tot op heden de pijnpunten aanduiden, maar op zich onvoldoende bijdragen tot een aanzienlijke verbetering van de onderwijskwaliteit. Vanuit die visie stellen enkele respondenten zich de vraag of de beschikbare middelen niet beter ingezet worden om die problemen direct aan te pakken. Verschillende respondenten zijn ook stellig over het feit dat onderwijskwaliteit vele dimensies kent en dat men binnen onderwijs moet durven kijken naar de huidige complexe onderwijsrealiteit in plaats van te focussen op een dalende onderwijskwaliteit.

“Er zit veel meer achter de onderwijskwaliteit dan het niveau van de lessen. Dat is gekoppeld aan maatschappelijke problemen. Het feit dat leerlingen niet meer dat niveau halen, heeft niet alleen te maken met wat er in de les gegeven wordt. Dat is veel ruimer dan dat.” (Schoolleider, respondent 6)

Afgezien van deze bedenkingen zien de respondenten globaal de meerwaarde van centrale toetsen in het school- en koepeloverschrijdend karakter en de grotere wetenschappelijke onderbouwing in vergelijking met de huidige gevalideerde proeven. Mede door het Vlaanderenbreed bereik, en bijgevolg de omvangrijke

vergelijkingsbasis, zien ze kansen om op een objectieve manier indicaties van sterktes en pijnpunten inzake onderwijskwaliteit op alle niveaus in beeld te brengen.

IV.2. Bekommernissen en randvoorwaarden bij het opzetten van centrale toetsen

Respondenten zijn een systeem van centrale toetsing echter niet zonder meer genegen. In de gesprekken worden mogelijkheden vaak in één adem met bestaande bekommernissen vernoemd. Het al dan niet kunnen realiseren van de mogelijkheden van centrale toetsen is volgens de respondenten afhankelijk van een veelheid aan randvoorwaarden.

IV.2.a. Pleidooi voor een ontwikkelingsgericht perspectief, met enkel een indirecte verantwoordingscomponent

De respondenten schuiven in hun reacties duidelijk een ontwikkelingsperspectief naar voor als voorwaarde voor een vruchtbare implementatie van de centrale toetsen. Eerder bleek reeds dat de respondenten centrale toetsen omschrijven als een spiegel voor de eigen onderwijskwaliteit en de resultaten beschouwen als kritische bril en hefboom voor de interne kwaliteitszorg. De meeste respondenten willen de resultaten van de centrale toetsen dan ook in de eerste plaats gebruiken om informatie te genereren over werkpunten en sterktes binnen de school om zo de interne werking te ondersteunen. Zowel schoolleiders als leerkrachten geven aan zicht te willen krijgen op de ontwikkeling van leerlingen en om aan de hand van leergerichte feedback aan de slag te willen gaan om leerwinst bij leerlingen te vergroten. Leerkrachten willen aan de hand van de resultaten hun lespraktijk evalueren en beogen om op basis van de resultaten aangepaste trajecten te kunnen opstarten voor de leerlingen. De pedagogische begeleiders zien de output van de toetsen als een extra instrument om de begeleiding van scholen meer gericht in te zetten en bijgevolg effectievere trajecten op maat aan te bieden. Op die manier krijgen scholen de kans om hun eigen evaluatie in handen te nemen en hoeven ze bijvoorbeeld niet te wachten op een evaluatie door de onderwijsinspectie.

“Als school moeten we de kans krijgen om op zelfonderzoek te gaan en te onderzoeken waar we beleidsmatig moeten bijsturen. Waar ligt het aan? Aan onze didactische aanpak, aan ons publiek, aan ons beleid?” (Schoolleider, respondent 9)

Over de mate waarin ook verantwoording een rol kan spelen, lopen meningen uiteen. De meerderheid van de respondenten is voorstander om resoluut te kiezen

voor een van beide doeleinden en in dat geval geniet het ontwikkelingsgericht perspectief de voorkeur. Enkele respondenten sluiten verantwoordingsgericht gebruik echter niet uit. Wanneer er ook sprake is van verantwoording als doel van de centrale toetsen, pleiten zij er wel voor om in de verantwoording te focussen op het gebruik van de toetsen als instrument voor kwaliteitszorg, eerder dan voor een directe focus op de resultaten zelf.

“Het hoeven geen twee uitersten van een continuüm te zijn. Het uitgangspunt is voor mij ontwikkelingsgericht, maar er mag een soort van controle zijn en men mag perfect bekijken of een school daar goed mee aan de slag gaat.” (PBD, respondent 4)

Geen enkele respondent is voorstander van een verantwoordingssysteem dat zich eenzijdig focust op de outputgegevens. We haalden reeds eerder aan dat respondenten bijvoorbeeld niet willen dat de toetsen op zich bepalend zijn bij beslissingen voor de leerlingen over het al dan niet overgaan en studiekeuze. Het gaat wel om beter onderbouwde studiekeuzes te maken. Daarnaast staan leerkrachten zelf huiverachtig tegenover het idee dat het systeem wordt ingezet om bij hen te controleren of eindtermen voldoende worden nagestreefd en gerealiseerd. De respondenten vrezen dat een te grote verantwoordingsgerichte inzet van de toetsen ongewenst strategisch gedrag zal uitlokken dat de ontwikkelingsgerichte focus inzake de bewaking en de verbetering van onderwijskwaliteit zal belemmeren.

“Ik ben tegen als het gaat over centrale toetsen waarbij je ‘leren voor de toets’ krijgt en waar er sterke consequenties aan vasthangen, zoals in Frankrijk met de BAC. Dat is vrij belangrijk, dus dat liever niet. Maar als het gaat om de groei van leerlingen proberen in kaart te brengen, en ook om onderwijskwaliteit van een school proberen in kaart te brengen, om het te gebruiken als feedback over ‘wat loopt er goed, wat loopt er niet goed’, ja, dan wel.” (Leerkracht, respondent 2)

Tot op heden legde de beleidsnota van de minister van Onderwijs vast dat scholen, waarvan de leerlingen significant minder leerwinst genereren op de centrale toetsen, in een begeleidingstraject moeten stappen om de kwaliteit van hun onderwijs te versterken. Enkele pedagogisch begeleiders bestempelen deze verplichting tot deelname aan een begeleidingstraject als een sanctie die door de overheid wordt opgelegd. Ze vrezen dat de toetsen hierdoor als sterk verantwoordingsgericht geïnterpreteerd zullen worden, ten nadele van de gewenste ontwikkelingsgerichte focus. In dat opzicht verkiezen ze unaniem een ondersteunende rol boven een controlerende.

IV.2.b. Centrale toetsen richten zich best op de brede kern van het curriculum

Leerkrachten rekenen erop in de toetsen een goede afspiegeling te vinden van wat er verwacht wordt van leerlingen, in lijn met de eindtermen. Hier heerst tegelijkertijd ongerustheid over. De meeste respondenten verwachten dat de resultaten van centrale toetsen (gebaseerd op twee vakdomeinen) slechts een onvolledig beeld zullen geven van onderwijskwaliteit en de leerwinst die binnen een school wordt gerealiseerd. Hierdoor dreigt volgens respondenten een verschuiving van de focus in het onderwijs naar de kerngebieden wiskunde en taal waardoor overige vakken en leercompetenties naar de achtergrond verdwijnen. Men vreest een vernauwing van het curriculum tot de leerdomeinen die zullen worden getoetst.

“Het kan gebeuren dat men strategisch gaat kiezen om in de onderwijstijd meer in te zetten op de vakken of de domeinen die zullen worden getest dan op andere domeinen. Er bestaat een zekere vrijheid om uren in te zetten in vrije ruimte en dergelijke. Als dan de reden om die uren te kiezen of toe te wijzen enkel het scoren op die tests is, ja dan zijn we verkeerd bezig.” (PBD, respondent 10)

“Het begrip centrale toetsen, het klinkt als iets dat heel betekenisvol is en dat gaat dan eigenlijk, in mijn ogen, toch maar over een heel klein aspect van het leren.” (Leerkracht, respondent 4)

Voor verschillende respondenten heeft het momenteel uitgetekende systeem inhoudelijk een te beperkte reikwijdte om de gehele onderwijskwaliteit in beeld te brengen. Naast het kennisluik vragen zij ook voldoende aandacht voor vaardigheden en attitudes. Tegelijkertijd erkennen meerdere respondenten, mede uit eigen ervaring met de huidige gevalideerde toetsen, dat deze vaardigheden en attitudes moeilijk(er) te evalueren zijn binnen een gestandaardiseerde toetsvorm. Desondanks ijvert de meerderheid van de respondenten voor de integratie van bijvoorbeeld een praktische proef, met objectieve en transparante beoordelingscriteria. Respondenten koppelen hun inschatting van de validiteit van de toetsen ook aan de gebruikte evaluatievorm. Een digitale toets wordt aangehaald als het meest geschikt instrument, aangezien die volgens de respondenten leidt tot een meer objectieve en snellere verwerking van de resultaten. Als kanttekening verwijzen enkele respondenten evenwel naar de mogelijk te enge focus op de soorten vragen, inhoud en competenties die gemakkelijker te toetsen zijn binnen een digitaal systeem.

“Ik denk dat er met wat creativiteit van de mensen die die toetsen ontwikkelen wel een en ander mogelijk is. Tegelijkertijd zal dat wat mogelijk is om te toetsen, bepalen wat wel of niet getoetst wordt. We moeten ons daar wel degelijk van bewust zijn, dat dat ook maar een beeld is. Dat

we zeker niet alles wat we toetsbaar willen hebben, zullen kunnen toetsen.” (PBD, respondent 2)

Als basis voor centrale toetsen verkiezen de respondenten zo goed als unaniem de eindtermen. Dat is volgens hen nodig om tot een Vlaanderenbreed beeld van competenties van de Vlaamse leerlingen te komen.

“Ik vind het logisch dat het is gebaseerd op de eindtermen. De leerplannen zijn een vertaalslag van de netten om die eindtermen in een stramien te gieten. Dus als je vergelijkbare toetsen wil over de netten heen dan moet je je baseren op de eindtermen die voor iedereen hetzelfde zijn.” (Schoolleider, respondent 1)

Een aantal respondenten uiten hierbij hun bezorgdheid over de uitgebreide nieuwe eindtermen en het gegeven dat eindtermen kunnen veranderen. Daarnaast benadrukt men gelijkwaardigheid van en evenwichtigheid tussen de verschillende sleutelcompetenties in de nieuwe eindtermen. Men is voorstander van een bredere toetsing of een aanbod van meer genormeerd toetsmateriaal voor scholen, vertrekkende vanuit de sleutelcompetenties. Centrale toetsen zouden ook een uitgelezen kans kunnen en moeten zijn om ook sociaalrelationele competenties, burgerschap, zelfbewustzijn, cultureel bewustzijn, STEM en digitale competenties op brede schaal te toetsen. Een aantal respondenten dromen van een toetsenbank met genormeerd en wetenschappelijk onderbouwd toetsmateriaal voor alle eindtermen.

Met de wens naar uitbreiding van de leergebieden op langere termijn in het achterhoofd tonen meerdere respondenten wel begrip voor de keuze voor Nederlands en wiskunde in de opstart(fase) van de centrale toetsen. De respondenten bevestigen het feit dat een goede beheersing van Nederlands de basis vormt om andere vakken te begrijpen en te verwerken. Een aantal schoolleiders vullen dit nog aan met de hoop dat de inhoud van de centrale toetsen op het einde van het secundair onderwijs wordt afgestemd op wat voor een leerling belangrijke informatie is. De schoolleiders hopen dat centrale toetsen zoveel mogelijk domeinspecifieke vakken zullen omvatten en afgestemd zullen zijn op de richting en de finaliteit waarin de leerling zich bevindt.

“Voor een 6de jaar in de arbeidsmarktfinaliteit moet je eigenlijk een toets afnemen met de vraag ‘Is die leerling zelfredzaam in arbeidsmarktcontext?’ Kan hij bijvoorbeeld een verkoopgesprek voeren of een offerte maken?” (Schoolleider, respondent 12).

Zinvol toetsgebruik vergt metingen die door de toekomstige gebruikers als valide en betrouwbaar gepercipieerd worden. Wanneer gebruikers de relevantie en de degelijkheid van de toetsen of de resultaten ervan betwisten dan zal het systeem de beoogde doelen niet kunnen waarmaken.

Wat de validiteit van de toetsen betreft, hebben de respondenten globaal veel vertrouwen in de academische partners indien zij de opdracht krijgen voldoende ruime toetsen te maken. Leerkrachten en schoolleiders van scholen met arbeidsmarktfinaliteit benadrukken daarbij in één adem evenwel het belang van naast toetsexperts ook experts uit het werkveld (vakleerkrachten) te betrekken. Het idee dat één toets voor verschillende onderwijsfinaliteiten relevant kan zijn, nemen zij niet zonder meer voor waar aan.

IV.2.c. Ook overwegen om de getoetste scharniervormen uit te breiden

Naast een uitbreiding op het vlak van leergebieden en leerinhouden zien de meeste respondenten ook een meerwaarde in de uitbreiding van de toetsfrequentie, zeker in het secundair onderwijs. Meer dan de helft van de respondenten vindt de frequentie van het einde van de eerste graad en het einde van de derde graad in het middelbaar onderwijs een logische keuze omdat dit schakelmomenten zijn bij studieoriëntering. Leerkrachten benadrukken daarnaast echter ook de graadgebonden eindtermen en verkiezen, net als schoolleiders, een extra meetmoment op het einde van de tweede graad om het groeiproces van leerlingen beter op te volgen en bij te sturen. Een bezorgdheid die schoolleiders ook vermelden, betreft de neveninstromers. Respondenten stellen zich vragen bij het meenemen van hun resultaten over de scholen heen en vrezen dat de resultaten van neveninstromers hun schoolresultaat zullen beïnvloeden zonder dat ze daar, als school, enige impact op hebben gehad. Ook vanuit dat perspectief zou een extra meting op het einde van de tweede graad secundair onderwijs een meerwaarde betekenen.

De meerderheid van de respondenten begrijpt de huidige overheidskeuze van twee meetmomenten in het basisonderwijs op het vlak van de leerwinstmonitoring. Het vooropgestelde ontwikkelingsperspectief indachtig weerklinkt het evenwel ook dat een nulmeting om te kijken waar elke leerling start wenselijk zou zijn. Terwijl de meting van het 6de leerjaar basisonderwijs kan worden beschouwd als een nulmeting voor het secundair onderwijs, stelt men vast dat zulk een nulmeting er in het basisonderwijs niet is.

We geven ook nog mee dat respondenten zich vragen stellen bij de doelbepalingen op het einde van het vierde leerjaar, aangezien dat geen decretaal eindpunt met eindtermen is. Men verwacht dat de overheid hiervoor dan een cesuur of tussendoelen zal opstellen, wetende dat dit leerplanoverschrijdend niet evident zal zijn.

IV.2.d. Beducht blijven voor vormen van manipulatie, zoals 'teaching to the test'

Het bewust en gepland vervalsen van toetsresultaten, moest dat dan al mogelijk zijn, is niet meteen iets waar de

bevraagde respondenten voor vrezen. Anders ligt het voor kleinere vormen van manipulatie waarbij leerkrachten leerlingen bijvoorbeeld subtiel helpen door vragen te herformuleren of moeilijke woorden te verklaren tijdens de afname. Ook het gaan trachten uit te sluiten van potentieel slecht scorende leerlingen voor de test is volgens respondenten niet ondenkbaar. Verder wordt een specifieke vorm van *teaching to the test*, namelijk een training op de evaluatievormen die in de centrale toetsen aan bod zullen komen, door verschillende respondenten als een te vermijden neveneffect benoemd. Verwacht wordt dat deze training op toetsvaardigheden ten koste zal gaan van de onderwijstijd. Ter illustratie verwijst een pedagogische begeleider naar de eigen bevindingen hieromtrent in aanloop naar de koepelgebonden proeven, waarbij leraren zonder intentie tot vervalsing hun leerlingen voorbereiden op de gevalideerde proeven binnen de onderwijstijd.

"Heel wat van onze leraren in het zesde leerjaar sparen en verzamelen de testen van de voorgaande jaren en laten een aantal weken voor de nieuwe proef de kinderen ook oefenen met die testen. We weten dat dat gebeurt." (PBD, respondent 8)

Sommige respondenten formuleren ook bedenkingen bij het kunnen garanderen van een gelijkaardige en dus vergelijkbare manier van afname binnen de scholen. De afspraken, de communicatie en de opvolging over wie de toetsen afneemt (eigen leerkrachten en/of externen), op welke locatie (aula, klaslokaal, examencentrum), onder welke omstandigheden (tijdsindicaties, hulpmiddelen voor zorgleerlingen ...) moeten volgens hen zeer duidelijk worden uitgewerkt in concrete draaiboeken. Vanuit leerlingenperspectief verkiezen respondenten een afname in de eigen school met een combinatie van eigen en externe toezichthouders om zo een zo veilig mogelijke context voor hun leerlingen te creëren.

Om de betrouwbaarheid van de beoordelingen te waarborgen, wijzen respondenten ook op de vertekening die kan ontstaan wanneer 'verbetering' door eigen leerkracht dient te gebeuren (zelfs met duidelijke correctiesleutels). De beoordeling van de toetsen moet dus volgens hen of geautomatiseerd gebeuren of in de handen van externen liggen.

"Leraren verbeteren allemaal op een andere manier. Ik ben zelf heel, heel streng, ik vind wiskundige notatie heel belangrijk. Je hebt leerlingen die vergelijkingen opschrijven en tussen elke stap nog eens een gelijkheidsteken zetten. Dat is niet juist. Maar je weet wel wat die leerling bedoelt en het gevaar is, denk ik, dat je dat van je eigen leerlingen gaat verbeteren. Want je wil onvermijdelijk ook dat die goed scoren op die toets. Want ja, het is toch een beetje uw persoonlijke verwezenlijking dat daarin zit." (Leerkracht, respondent 4)

IV.2.e. Openbaarheid over resultaten vermijden, vooral omdat men eenzijdige vormen van schoolvergelijking vreest

Wanneer er sprake is van publieke openbaarheid van resultaten vrezen respondenten maatschappelijk een te eenzijdige kijk op onderwijskwaliteit. Angst voor rankings loopt als een rode draad doorheen de gesprekken en wordt automatisch aan de openbaarheid van de schoolresultaten gekoppeld. Op zich kan ongeveer de helft van de schoolleiders zich vinden in het publiceren van resultaten, als deze tenminste duidelijk worden gekaderd aan de hand van context- en achtergrondkenmerken van zowel leerlingen als van de school. De tegenstanders van openbaarheid, waaronder de meeste van de bevroegde leraren, leggen de focus op 'interne' kwaliteitszorg en menen dan ook dat de resultaten alleen intern gebruikt dienen te worden.

Enkele respondenten vrezen dat rankings op hun beurt zullen leiden tot meer segregatie tussen sterke en zwakke scholen. Cognitief sterke leerlingen en bekwame leerkrachten zullen volgens hen hoog scorende scholen verkiezen en leerlingen met leerproblemen of taalachterstand zullen door deze scholen sneller geheroriënteerd worden naar andere scholen, die hierdoor een leerlingenpopulatie krijgen met een groot aantal leerlingen dat mogelijk zwakker scoort op centrale testen. Een schoolleider en enkele leerkrachten menen dat de rankings een competitieve positieve stimulans kunnen teweegbrengen en zelfs een jarenlange negatieve perceptie van een school kunnen ombuigen, als blijkt dat een school beter scoort dan publiekelijk verwacht.

"Op de lange duur krijg je in die scholen dat ze aan het selecteren gaan. Dan willen ze alleen nog met de fine fleur werken. Zoals je dat ziet bij onze buurscholen, die beginnen met 140 leerlingen en eindigen in het 6de jaar met 40 leerlingen. Waar zitten die andere 100? Zijn die allemaal niet competent?" (Schoolleider, respondent 4)

Het correct meenemen van de context- en achtergrondkenmerken is een randvoorwaarde die doorheen de interviews herhaaldelijk weerklinkt wanneer de respondenten hun percepties bespreken rond de openbaarheid van resultaten en het meten van leerwinst.

"Omdat interpretatie van iets altijd gepaard gaat met voorkennis. Als je resultaten zomaar ergens openbaar maakt en mensen die helemaal geen weet hebben van hoe onderwijs in elkaar zit of wat de randfactoren zijn ... Dan kan dat zeer smal worden geïnterpreteerd." (PBD, respondent 5)

IV.2.f. Krijtlijnen uitzetten voor de inbedding van de centrale toetsen in de bestaande leerlingevaluatie

De meningen van zowel leerkrachten als schoolleiders over welke plaats resultaten van centrale toetsen in

de bestaande evaluatieaanpak dienen te krijgen, zijn verdeeld. De kijk hierop is nog in ontwikkeling. Volgens de respondenten betekent dit dat zonder duidelijke richtlijnen bij de implementatie van het systeem erg grote verschillen tussen scholen kunnen ontstaan. Zo zijn er bijvoorbeeld argumenten om bestaande toetsen te vervangen door centrale toetsen. Respondenten vermelden hier het vermijden van dubbele evaluaties over dezelfde leerinhouden en een verhoging van de motivatie bij leerlingen om deel te nemen aan de centrale toetsen als de resultaten voldoende gewicht krijgen. Tweemaal hetzelfde toetsen, zorgt bij sommige leraren ook voor ongerustheid vanuit een verantwoordingsperspectief. Het gaat dan met name over de vraag wat te doen als er een discrepantie blijkt tussen de resultaten van de centrale toetsen en de eigen evaluatieresultaten, en leraren die discrepantie dienen te verklaren en te verantwoorden aan leerlingen, ouders en hun directies.

"Idealiter neem je de centrale toetsen mee op in je evaluatiebeleid, bijvoorbeeld einde van het 6de jaar Nederlands, en je weet wat erin getest wordt. Dan kan je je eigen examen daaraan aanpassen en die dingen niet dubbel gaan testen. Het lijkt me zwaar om dat nog boven op onze eigen examenreeks te gaan zetten. Naar leerlingen toe is het beter dat het geïntegreerd is" (Schoolleider, respondent 2)

"Ik zou ze integreren. Dat gaat ook het meest duidelijk zijn, want anders ga je boodschap geven op een oudercontact, ja even kort door de bocht hé, dit zijn de resultaten van Vlaanderen en dit zijn de resultaten van mij. Ja, die komen dan niet allemaal overeen. En dan moet de leerkracht zich al beginnen te verantwoorden. Dus als je die samen kan nemen, dan krijg je wel een volledig beeld. Dan kan de leerkracht ook meer nuance brengen, meer uitleg geven van, 'kijk, daar komt het resultaat van uw zoon of dochter vandaan.'" (Leerkracht, respondent 3)

Daarnaast weerklinken er argumenten om de centrale toetsen wel degelijk naast en niet in plaats van de eigen toetsen te plaatsen. Dat zou waardevoller zijn om het evaluatie- en groeiproces dat de leerkracht gedurende een schooljaar met leerlingen aflegt ten volle vorm te geven. In deze logica verkiest men om de resultaten van beide soorten toetsen mee te nemen in feedbackgesprekken, ter reflectie bij leerlingencontacten en als een soort aparte rubriek in de leerlingenvolgsystemen.

Sommige respondenten ervaren het meenemen van de resultaten van centrale toetsen in het eigen evaluatiesysteem dan weer als een beperking van de deliberatievrijheid van een klassenraad. Een schoolleider merkt op dat er ook hiervoor best een wettelijk kader gecreëerd wordt. Als de ene school de resultaten meeneemt in de leerlingevaluatie en een andere school niet, dan verwacht deze schoolleider mogelijks problemen in de vorm van het aanvechten van attesteringen.

“Ga dan maar eens bewijzen dat de centrale toetsen correct zijn afgenomen, dat de leerlingen daar voldoende waren op voorbereid ... Dat zijn juridische discussies die ik niet wil voeren.” (Schoolleider, respondent 11)

IV.2.g. Handvaten bieden om met de resultaten aan de slag te gaan

Respondenten vrezen dat de centrale toetsen, als de overheid zich onvoldoende bewust is van de lacune in capaciteit in scholen, enkel zullen leiden tot vaststellingen zonder dat er verbeteringen of veranderingen worden geïnitieerd. Op verschillende manieren wordt informatiegeletterdheid immers als een belangrijke pijler binnen het proces van de centrale toetsen naar voren geschoven. Dit vergt volgens velen in de eerste plaats nog een mentale bijsturing, om zich bewust te zijn van het belang van data ter verbetering van de onderwijsleerpraktijk en om als onderzoekende professionals op een onderbouwde manier de eigen onderwijskwaliteit te analyseren. De respondenten wijzen op een algemeen gebrek aan kennis en vaardigheden van het onderwijspersoneel op dit vlak. Zo geven de bevroegde schoolleiders aan dat de interpretatie en de vertaling van cijferresultaten naar concrete aanpassingen van het schoolbeleid niet evident is voor hen. In dat opzicht hopen schoolleiders ook woordelijke feedback en zelfs adviezen terug te vinden in het rapport van de centrale toetsen, onder meer om misinterpretatie te voorkomen.

In het verlengde van het correct lezen en interpreteren van de data wijzen respondenten ook op het belang van het beleidsvoerend vermogen van scholen om zinvol aan de slag te gaan met de verzamelde data uit de centrale toetsen. Ze verwachten dat de toetsen een extra bron aan data zullen vormen en benadrukken dat de school in staat dient te zijn om de verscheidene databronnen niet alleen te begrijpen en te verwerken, maar ook te vertalen in acties.

“Als de centrale toetsen ingevoerd worden, en scholen krijgen daar op een goede manier feedback over, én schoolleiders zijn bekwaam om met de feedbackrapporten aan de slag te gaan om dat in beleid om te zetten en op die manier beweging te creëren op leerkrachtniveau – ja, dan bereiken we een effect. Maar dat gaat een inspanning van jaren zijn.” (Schoolleider, respondent 9)

Op het vlak van opleidingen voor zowel informatiegeletterdheid als beleidsvoerend vermogen zien de pedagogisch begeleiders en schoolleiders voor zichzelf een belangrijke rol weggelegd. Respondenten beschouwen deze professionalisering als een belangrijkere prioriteit voor de komende jaren om scholen klaar te stomen voor de centrale toetsen. Zo wordt aangegeven dat het niet wenselijk is dat schoolteams pas na de invoering van deze toetsen starten met vormingen omtrent

informatiegeletterdheid, maar dat ze best nu reeds al intensiever met het gedachtegoed aan de slag gaan om een onderzoekende cultuur uit te bouwen.

V. Afrondend

De bevroegde respondenten blijken er een behoorlijk positieve houding op na te houden ten aanzien van een ontwikkelingsgericht leerwinstmonitoringsysteem. Hoewel de methodologische aanpak van deze studie het niet toelaat het standpunt van de bevroegde respondenten te veralgemenen, is het volgens hen van belang deze positieve houding aan te wenden en vanuit deze basis de centrale toetsen verder vorm te geven. Respondenten verwachten dat het onderwijsveld zal openstaan voor centrale toetsen als de toetsen dienen om werkpunten en sterktes zichtbaar te maken, zowel in de interne schoolwerking als bij individuele leerlingen. De door ons bevroegde actoren vragen om duidelijke doelstellingen te formuleren, die vervolgens op een eenduidige manier naar het onderwijsveld gecommuniceerd worden. Het positief kaderen in interne kwaliteitszorg, inzetten op het leren van leerlingen én leraren, maar uiteraard ook het werken met psychometrisch onderbouwde en bovenal voor het curriculum in brede zin valide toetsen, zijn volgens de respondenten noodzakelijke voorwaarden om de gedragenheid verder te ontwikkelen. Precies hierin schuilt de beleidsrelevantie van deze bijdrage. De aanbevelingen die de geconsulteerde respondenten meegeven vanuit de invulling van een voor hen ideaal systeem, zijn wezenlijk om de gedragenheid van het systeem te waarborgen.

Respondenten vinden het noodzakelijk om bij de verdere communicatie over centrale toetsen ‘het waarom’ van de gemaakte keuzes goed te duiden. Het is in die zin ook belangrijk om bij de beschrijving van de bovenstaande resultaten mee te geven dat de meerderheid van de bevroegde respondenten tijdens deze studie nog botsten op het nog hypothetisch karakter van de centrale toetsen. Deze studie vond plaats in een periode waarin weliswaar een aantal cruciale bouwstenen van de centrale toetsen reeds vastgelegd waren, maar waarin de beleidsvoorbereiding nog onvoldoende ver gevorderd was om over een volledig uitgeschreven scenario van gedachten te wisselen. Zo liggen politieke discussies, het beleidsproces en de belangenbehartiging omtrent dit thema volgens respondenten nog in het verschiet. Respondenten zitten dan ook nog met veel vragen over het uiteindelijke doel, de impact en de vormgeving van het systeem. De bovenstaande resultaten tonen aan dat er nog veel vragen resten waarop volgens respondenten klaarblijkelijk nog niemand momenteel het antwoord kent. Onduidelijkheden wegwerken is dan ook een randvoorwaarde die talrijke respondenten uit de verschillende doelgroepen tijdens hun interview ter sprake brengen.

Centrale toetsen voor het onderwijsveld vergen gedragenheid door het onderwijsveld. Onze ambitie met deze studie was in kaart te brengen wat er in deze fase (voorjaar 2021) leeft bij een selectie van leraren, schoolleiders en pedagogisch begeleiders. Wellicht zullen de opinies van deze respondentengroepen echter nog wijzigen en zich verfijnen naarmate de invulling omtrent de centrale toetsen gedetailleerder wordt. Bovendien moeten we terughoudend zijn met het veralgemenen van de gepresenteerde bevindingen. In die zin loont het om aan de hand van gebruikersonderzoek de vinger aan de pols te houden en om meer representatief onderzoek op te zetten. Het is daarbij zinvol de opzet meteen ook uit te breiden naar leerlingen en ouders als gebruikers.

VI. Bronnenlijst

- AMREIN-BEARDSLEY, A. (2008), "Methodological Concerns About the Education Value-Added Assessment System", *Educational Researcher*, 37(2), 65-75.
- AMREIN-BEARDSLEY, A., BERLINER, D.C. en RIDEAU, S. (2010), "Cheating in the first, second, and third degree: Educators' responses to high-stakes testing", *Education Policy Analysis Archives*, 18(14), 1-36.
- BAIRD, J., AHMED, A., HOPFENBECK, T., BROWN, C. en ELLIOT, V. (2012), *Research evidence relating to proposals for reform of the GCSE*, Oxford, UK, Oxford University Centre for Educational Assessment.
- BISHOP, J.H. en WÖSSMANN, L. (2004), "Institutional Effects in a Simple Model of Educational Production", *Education Economics*, 12(1), 17-38.
- BLACK, P. en WILIAM, D. (2005), "Lessons from around the World: How Policies, Politics and Cultures Constrain and Afford Assessment Practices", *Curriculum Journal*, 16(2), 249-261.
- CHAKRABARTI, R. en SCHWARTZ, N. (2013), "Unintended consequences of school accountability policies: evidence from Florida and implications for New York", *Economic Policy Review*, (19), 44.
- DE MAN, L. en DENYS, K. (2021), *Vlaamse toetsen. Waar vandaan en waar naartoe?*, Brussel, Politeia.
- ELSTAD, E. (2009), "Schools which are named, shamed and blamed by the media: school accountability in Norway", *Educational Assessment Evaluation and Accountability*, (21), 173-189.
- GOODMAN, D. en HAMBLETON, R.K. (2005), "Some misconceptions about large-scale educational assessments" in PHELPS, R.P. (ed.), *Defending standardized testing*, Mahwah, NJ/Londen, Lawrence Erlbaum Associates Publishers, p. 91-110.
- HEMKER, B., FESKENS, R., ONNA, M.V., SMITS, I., BÉGUIN, A., STRAAT, H. en WOOLS, S. (2021), *Pedagogisch-psychometrische aspecten van centraal toetsen in Vlaanderen*, Arnhem, Stichting Cito.
- HOPSTER-DEN OTTER, D., WOOLS, S., EGGEN, T.J.H.M. en VELDKAMP, B.P. (2017), "Formative use of test results: A user's perspective", *Studies in Educational Evaluation*, (52), 12-23.
- JANSSENS, F.J.G., REKEN-MOMBARG, L. en LACOR, E. (2014), *Leerwinst en toegevoegde waarde in het primair onderwijs*, geraadpleegd via www.academia.edu/24276840/Leerwinst_en_toegevoegde_waarde_in_het_primair_onderwijs_eindrapportage_LTW_PO.
- KLEIN, E.D. en VAN ACKEREN, I. (2012), "Challenges and Problems for Research in the Field of Statewide Exams. A Stock Taking of Differing Procedures and Standardization Levels", *Studies in Educational Evaluation*, 37(4), 180-188.
- LOONEY, A. (2014), "Assessment and the Reform of Education Systems. From good news to policy technology" in WYATT-SMITH, C., KLENOWSKI, V. en COLBERT, P. (eds.), *Designing assessment for quality learning*, Londen, Springer, p. 233-247.
- MARLOW, R., NORWICH, B., UKOUMUNNE, O.C., HANSFORD, L., SHARKEY, S. en FORD, T. (2014), "A comparison of teacher assessment (APP) with standardised tests in primary literacy and numeracy (WIAT-II)", *Assessment in Education: Principles, Policy & Practice*, 21(4), 412-426, geraadpleegd via <https://doi.org/10.1080/0969594X.2014.936358>.
- MARSH, J.A., PANE, J.F. en HAMILTON, L.S. (2006), *Making Sense of Data-Driven Decision Making in Education*, Santa Monica CA, RAND Corporation.
- MORTELMANS, D. (2017), *Kwalitatieve analyse met Nvivo*, Leuven, Acco.
- OECD (2013), *Synergies for Better Learning. An international perspective on evaluation and assessment*, Parijs, OECD Publishing.
- OPDENAKKER, M. en VAN DAMME, J. (2000), "Effects of Schools, Teaching Staff and Classes on Achievement and Well-Being in Secondary Education: Similarities and Differences Between School Outcomes", *School Effectiveness and School Improvement*, 11(2), 165-196.
- PENNINCKX, M., VANHOOF, J., QUINTELIER, A., DE MAEYER, S. en VAN PETEGEM, P. (2017), *Zicht op leerwinst*, Leuven, Acco.

SAUNDERS, L. (1999), "A Brief History of Educational "Value-Added": How Did We Get to Where We Are?", *School Effectiveness and School Improvement*, 10(2), 233-256.

SCHILDKAMP, K. en KUIPER, W. (2010), "Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors", *Teaching and Teacher Education*, 26(3), 482-496, geraadpleegd via <http://dx.doi.org/10.1016/j.tate.2009.06.007>.

SHEWBRIDGE, C., HULSHOF, M., NUSCHE, D. en STOLL, L. (2011), "School evaluation in the Flemish community of Belgium" in OECD (ed.), *OECD Reviews of Evaluation and Assessment in Education*, Parijs, OECD Publishing.

SLOANE, F.C., OLOFF-LEWIS, J. en KIM, S.H. (2013), "Value-added models of teacher and school effectiveness in Ireland: wise or otherwise?", *Irish Educational Studies*, 32(1), 37-67, geraadpleegd via <https://doi.org/10.1080/03323315.2013.773233>.

SMITH, W.C. (2014), "The Global Transformation toward Testing for Accountability", *Education Policy Analysis Archives*, 22(116), geraadpleegd via <http://doi:10.14507/epaa.v22.1571>.

STOBART, G. (2008), *Testing Times: The Uses and Abuses of Assessment*, Routledge.

VAN GASSE, R., VANHOOF, J., MAHIEU, P. en VAN PETEGEM, P. (2015), *Informatiegebruik door schoolleiders en leerkrachten*, Antwerpen, Garant Uitgevers.

VLOR (2021), *Nood aan debat over gevalideerde, gestandaardiseerde en genormeerde proeven. Advies naar aanleiding van de haalbaarheidsstudie 'naar de invoering van centrale toetsen in Vlaanderen'*, Brussel, Vlaamse Onderwijsraad.

WANG, L., BECKETT, G.H. en BROWN, L. (2006), "Controversies of Standardized Assessment in School Accountability Reform: A Critical Synthesis of Multidisciplinary Research Evidence", *Applied Measurement in Education*, 19(4), 305-328.

WEYTS, B. (2019), *Beleidsnota 2019-2024 Onderwijs*, geraadpleegd via www.vlaanderen.be/publicaties/beleidsnota-2019-2024-onderwijs.

