

1116 (2021-2022) – Nr. 1
ingediend op 31 januari 2022 (2021-2022)

Verslag van de hoorzitting

namens de Commissie voor Onderwijs
uitgebracht door Koen Daniëls en Loes Vandromme

over centrale toetsen
voor het basis- en secundair onderwijs

Samenstelling van de Commissie voor Onderwijs:

Voorzitter: Karolien Grosemans.

Vaste leden:

Arnout Coel, Koen Daniëls, Karolien Grosemans, Kathleen Krekels, Annabel Tavernier;
Roosmarijn Beckers, Jan Laeremans, Kristof Slagmulder;
Loes Vandromme, Hilâl Yalçın;
Jean-Jacques De Gucht, Gwendolyn Rutten;
Johan Danen, Elisabeth Meuleman;
Hannelore Goeman.

Plaatsvervangers:

Marius Meremans, Axel Ronse, Nadia Sminate, Kris Van Dijck, Katja Verheyen;
Leo Pieters, Els Sterckx, Sam Van Rooy;
Katrien Schryvers, Brecht Warnez;
Els Ampe, Tom Ongena;
Stijn Bex;
Steve Vandenberghe.

Toegevoegde leden:

Kim De Witte.

INHOUD

I.	Uiteenzetting door Claire Shewbridge	5
1.	Vergelijking met andere OESO-landen	5
2.	Een evoluerende cultuur over gestandaardiseerde testen	6
3.	OESO-casestudy voor Vlaanderen	6
II.	Uiteenzetting door het Departement Onderwijs en Vorming.....	8
1.	Verschillende namen, dezelfde toetsen	8
2.	Regeerakkoord 2019-2024.....	8
3.	Startfase van het beleidsproject	8
4.	Aanleiding: prestaties Vlaamse leerlingen	9
5.	Doelstellingen Vlaamse toetsen	9
6.	Welke toetsen?.....	10
7.	Welke leerlingen?.....	10
8.	Timing.....	10
9.	Vergelijking met internationale toetsen en peilingen.....	10
10.	Impact van de resultaten	11
11.	Teaching to the test	11
12.	Terugkoppeling van de resultaten	11
13.	Openbaarheid van resultaten.....	11
14.	Planning 2022	11
III.	Uiteenzetting door Johan van Braak	12
1.	Van oproep naar oprichting	12
2.	Vijf kernopdrachten.....	12
2.1.	Kernopdracht 1: centrale toetsen voor Nederlands en wiskunde	12
2.2.	Kernopdracht 2: scholen ondersteuning bieden bij de afname van de centrale toetsen	13
2.3.	Kernopdracht 3: de toetsresultaten verwerken en analyseren ...	14
2.4.	Kernopdracht 4: het onderwijsveld feedback geven over de toetsresultaten	14
2.5.	Kernopdracht 5: expertise ontwikkelen op het vlak van duurzame onderwijs-effectiviteit	14
3.	Acht werkdomeinen.....	14
4.	Samenwerking met stakeholders	15
5.	Besluit.....	15

IV.	Uiteenzetting door Rianne Janssen	15
1.	Les 1: peilingsonderzoek 2002-2022	16
2.	Les 2: peilingen en centrale toetsen: gelijkenissen en verschillen.....	16
3.	Les 3: toets wiskunde en Nederlands.....	16
4.	Les 4: behalen van de onderwijsdoelen	17
5.	Les 5: verschillen tussen scholen en schoolfeedback	17
6.	Les 6: het belang van het systeemniveau	18
7.	Les 7: leerwinst	19
8.	Besluit.....	19
V.	Uiteenzetting door Hans Teegelbeckers.....	20
1.	Nederland is toetsgedreven	20
2.	Waarom toetsen?.....	20
3.	Centrale toetsing in Nederland.....	20
4.	Wat is er tegen centrale toetsing?.....	21
5.	Samenvattend advies	22
VI.	Bespreking.....	22
1.	Tussenkomensten van de leden	22
1.1.	Loes Vandromme	22
1.2.	Koen Daniëls	24
1.3.	Roosmarijn Beckers.....	25
1.4.	Hannelore Goeman	26
1.5.	Johan Danen	27
1.6.	Jan Laeremans	28
2.	Antwoorden van de sprekers	28
	Gebruikte afkortingen.....	36
	Bijlage: zie de dossierpagina van dit document op www.vlaamsparlement.be	

Op 16 december 2021 organiseerde de Commissie voor Onderwijs een hoorzitting over centrale toetsen voor het basis- en secundair onderwijs met:

- Claire Shewbridge, projectleider Strategic Education Governance project, Directorate for Education and Skills, OESO;
- Bieke De Fraine, projectleider Vlaamse Toetsen, afdeling Strategische Beleidsondersteuning, Departement Onderwijs en Vorming, en Jeroen Backs, afdelingshoofd Strategische Beleidsondersteuning, Departement Onderwijs en Vorming;
- Johan van Braak, promotor-coördinator van het Steunpunt Centrale Toetsen in het Onderwijs;
- Rianne Janssen, promotor-coördinator van het Steunpunt Toetsontwikkeling en Peilingen;
- Hans Teegelbeckers, directeur VOS/ABB, Nederland.

Het OESO-rapport Promoting Education Decision Makers' Use of Evidence in Flanders is terug te vinden op de [dossierpagina](#) van dit document op www.vlaamsparlement.be.

(Deze commissievergadering werd als videoconferentie georganiseerd.)

I. Uiteenzetting door Claire Shewbridge

Claire Shewbridge heeft recent onderzoek gevoerd in Vlaanderen naar het betrekken van belanghebbenden bij het project van de centrale toetsen. Het rapport van de OESO daarover is gepubliceerd op 22 november 2021. Het werd gepresenteerd aan de stakeholders; ook aan de Vlor gaf de spreker een presentatie.

1. Vergelijking met andere OESO-landen

Vlaanderen is een uitzondering onder de OESO-landen want het is een van de weinige dat geen centrale examens heeft op het einde van het leerplichtonderwijs. In sommige landen zijn die er zelfs op het einde van het basisonderwijs. Over de voorbije twintig jaar zijn er in veel landen algemene gestandaardiseerde proeven ingevoerd, waar elk kind en elke school die aan de testcriteria voldoen, bijvoorbeeld een bepaald leerjaar, uitgenodigd worden om aan deel te nemen, met voorzieningen voor kinderen met speciale behoeften, adaptaties voor leerproblemen enzovoort. In 2002 introduceerde Vlaanderen peilingen, op steekproeven gebaseerde beoordelingen, en vrijwillige parallelversies van die testen. Die combinatie komt in nogal wat andere landen voor.

Noorwegen introduceerde in 2004 een centrale test van de hele populatie. De aanleiding was de publicatie van de eerste PISA-resultaten. Voordien was er daar geen gemeenschappelijke testaanpak. Het doel van de test was vooral om de cultuur van feedback te veralgemenen. Toen Australië in 2008 testen introduceerde, was de context helemaal anders. Er waren al nationale steekproeftesten en ook verschillende staten en regio's hadden al proeven. Het doel was dan ook vooral de middelen efficiënter te besteden, maar ook nationale harmonisatie en de promotie van gelijkwaardigheid en professionalisering in testontwikkeling. Toen Denemarken in 2010 populatiebrede testen organiseerde, was er evenmin een gemeenschappelijke testaanpak, maar waren er wel verplichte leerlingvolgsystemen. Ook daar waren er discussies over onderwijskwaliteit naar aanleiding van de PISA-resultaten. Het Deense doel was daarnaast om een evaluatiecultuur te stimuleren, met leerlingassessment, schoolevaluatie en lerarenbeoordeling. Die waren er niet, in tegenstelling tot Vlaanderen waar de inspectie al veel langer ingesteld is. Kortom, de context, de startpunten en de doelen waren anders.

In Australië waar er voor de introductie al testen waren, ligt de nadruk sterk op aansprakelijkheid, verantwoording en transparantie. Het basisidee is professionalisering, feedback, kwaliteit van het leren. Het is een nationale evaluatie van leerprogramma's en onderwijsnetwerken. In grote delen van het land richten de gemeenten het onderwijs in. Het testen gebeurt tweejaarlijks op het einde van het schooljaar en gaat over verschillende gebieden van rekenvaardigheid en geletterdheid. Van bij het ontwerp is de mogelijkheid ingebouwd om de cohorte te volgen, en er is daarenboven een website en een rapportagesysteem.

In Denemarken wordt er op het einde van het schooljaar getest: afwisselend Deens, wiskunde, Engels, fysica en chemie. Er zijn ook optionele testen mogelijk, waar de scholen dus zelf voor kunnen kiezen. De wet bepaalt dat er geen resultaten per school openbaar worden gemaakt, maar wel algemene verslagen over kwaliteit en jaarrapporten. Voor de ouders zijn er individuele rapporten.

Centrale systemen opzetten vergt heel wat middelen, maar systemen evolueren ook met nieuwe inzichten. Zo heeft Noorwegen beslist om de testen te vervroegen naar het begin van het schooljaar zodat de leerkracht op basis van de resultaten nog actie kan ondernemen. Ook Australië besliste, na een uitgebreid debat ook in de academische wereld, de testen eerder op het schooljaar te houden, maar ook om niet langer tussen scholen te vergelijken maar enkel de schoolvoortgang in beeld te brengen. Voorts werden de testen van het derde naar het vierde middelbaar verschoven (grade 9 to 10) en de scope ruimer gemaakt dan enkel geletterdheid en rekenvaardigheid. Een en ander moet nog verfijnd worden.

In Denemarken wordt het testsysteem momenteel geëvalueerd door een raad van stakeholders, vergelijkbaar met het Vlaamse highlevelforum. Het voorstel is ook om de toetsen eerder op het jaar te houden en te beperken tot Deens en wiskunde, maar ook een verplichte geletterdheidstest en een test op dyslexie in het eerste leerjaar in te voeren. Het aanbod van optionele toetsen zou daarenboven worden uitgebreid. Er komt voorts een sterkere nadruk op hoogbegaafde leerlingen.

2. Een evoluerende cultuur over gestandaardiseerde testen

Het is voor de spreker duidelijk dat centrale toetsen maar kunnen worden geïntroduceerd als de geesten er rijp voor zijn. Vlaanderen maakte op dat vlak een evolutie door. In 2009 volgde één op vier leerlingen les in een school die een of andere vorm van gestandaardiseerde toets afnam. Een brede OESO-evaluatie van 2010 wees uit dat een grote meerderheid niet open stond voor centrale toetsen. In 2015 volgde al een op de twee leerlingen les in een school die een of andere vorm van gestandaardiseerde toets afnam. Een onderzoek van 2017 wees op een wijzigende mentaliteit. In 2018 zat twee van de drie leerlingen op een school die de evoluties mat met een gestandaardiseerde toets. Dat jaar ontstond ook de vereiste voor verplichte gevalideerde eindtoetsen in het zesde leerjaar.

3. OESO-casestudy voor Vlaanderen

De OESO-gevalsstudie heeft een drielagig perspectief, met de algemene Vlaamse context en de betrokkenheid van de verschillende actoren als uitersten. De brug tussen de beide bestaat uit capaciteit, kennis en verantwoording, dus de dataverzameling, de data, hoe die geïnterpreteerd worden en wat ermee gedaan wordt.

De onderzoekers hebben de stakeholders gevraagd welke bezorgdheden ze hebben over de introductie van gestandaardiseerde testen, of ze tevreden zijn over hun betrokkenheid bij de introductie en hoe hun professionele kennis of expertise kan bijdragen tot de succesvolle introductie ervan. Daaruit blijkt de brede steun voor wetenschappelijke onderbouwing van bij het begin van de ontwikkeling. Ze zien een belangrijke rol daarin voor het steunpunt dat een groot vertrouwen geniet. De

stakeholders vinden dat er te weinig overleg was in het beginstadium van de ontwikkeling. De keerzijde is dat ze gemotiveerd zijn voor een grotere en structurele betrokkenheid.

Voorts zijn er heel wat opmerkingen over de onduidelijke regeringscommunicatie over de bedoelingen van de gestandaardiseerde testen. Vlaanderen heeft daarop het highlevelforum opgericht, wat volgens de OESO het ideale platform is voor feedback en stakeholderbetrokkenheid. Kortom, uit de gevalsstudie blijkt de behoefte aan een duidelijke en actieve communicatie, aan een politiek engagement om de actoren te betrekken en aan georganiseerde deelname van het onderwijsveld aan het hele ontwikkelingsproces.

De stakeholders werd ook gevraagd om hun licht te laten schijnen op de volledige Vlaamse (onderwijs)context. Het is duidelijk dat ze allemaal bezorgd zijn over de achteruitgang van de onderwijskwaliteit. Ze wijzen wel op de complexiteit van de onderwijskwaliteit en van gelijke onderwijskansen.

Algemeen worden gestandaardiseerde testen gezien als instrumenten voor schoolontwikkeling. Het OESO-team noteerde als meest geciteerde bezorgdheden: het aantrekken en vasthouden van uitstekende leraren, de herwaardering van het lerarenberoep, de professionalisering van de leraar, de verschillende contexten en voorzieningen van scholen, de verschillen in leerlingenpopulatie. Daarbij komt nog dat scholen voorbehoud maken bij het publiceren van schoolresultaten. Verhoudingsgewijs zijn er in Vlaanderen hoge overheidsuitgaven voor onderwijs en is er een brede tevredenheid over salarissen. Er is echter de perceptie dat het beroep van leraar minder aantrekkelijk is geworden, onder meer door een bovengemiddeld niveau van stress, dat ook bevestigd wordt in enquêtes.

De OESO merkt dat de geesten ondertussen gerijpt zijn voor centrale toetsen, hoewel er zeker nog moet worden gewerkt aan het vertrouwen van het onderwijsveld. Uit onderzoek blijkt dat de meeste directies en leerkrachten openstaan voor verandering en innovatie, maar dat geldt niet voor het volledige onderwijssysteem. Leerkrachten en directies uit het lager onderwijs zijn er meer toe bereid dan die uit het lager secundair onderwijs.

Om het draagvlak te versterken, acht de OESO het nodig om gemeenschappelijke doelen te ontwikkelen en te verspreiden, een langetermijnperspectief op te stellen, de acties te coördineren, en concrete ervaringen van het onderwijsveld te delen en eruit te leren.

Wat de capaciteit betreft moeten systemen, maar ook scholen, klaar zijn voor digitale testen en dient de toetscultuur erop afgestemd te zijn. Heel wat scholen, vooral in het lager onderwijs, voldoen nu nog niet aan die eisen. Daarbij komt nog dat ze moeten gemotiveerd zijn om met testen aan de slag te gaan. De OESO peilde in het seminarie van juni naar die motivatie. Daaruit blijkt wel dat er behoefte is aan betrouwbare informatie over de resultaten van leerlingen, leerkrachten en scholen, aan vergelijkende feedback zodat scholen over hun ontwikkeling kunnen reflecteren, aan instrumenten om een feedbackcultuur te versterken en te promoten en aan een katalysator om de professionele vaardigheden om met wetenschappelijke gegevens en data om te gaan te boosten. Uit een OESO-enquête van 2018 blijkt dat heel wat leerlingen geen feedback krijgen van hun leerkracht. De data kunnen ook gebruikt worden om de schoolinspectie te ondersteunen. Kortom, de voornaamste vragen in het seminarie zijn een duidelijke communicatiestrategie, voldoende middelen en ondersteuning van scholen en leerkrachten.

Alle stakeholders voelen een weerstand tegen aansprakelijkheid (accountability), maar begrijpen uiteindelijk wel dat het enkel gaat om een interne verantwoordingsplicht om kwaliteitsvol onderwijs te garanderen. Er is eensgezindheid over

dat de testen moeten ingebed zijn in het onderwijskwaliteitskader. Stakeholders staan open voor professionele dialoog.

De spreker eindigt met de Vlaamse kwaliteitsdriehoek. De data die uit centrale proeven komen, kan elk deel en dus ook de volledige driehoek versterken, en een brug vormen tussen de inspectie, de pedagogische begeleidingsdiensten en de scholen.

II. Uiteenzetting door het Departement Onderwijs en Vorming

Jeroen Backs is verantwoordelijk voor de implementatie van de Vlaamse toetsen, een omvangrijk en complex beleidsproject. *Bieke De Fraine* is projectleider Vlaamse toetsen in het Departement Onderwijs en Vorming.

1. Verschillende namen, dezelfde toetsen

Vooreerst wijst hij op de verschillende namen die aan dezelfde toetsen gegeven worden. In het regeerakkoord staan gestandaardiseerde, genormeerde en gevalideerde net- en koepeloverschrijdende proeven, door sommigen afgekort als GGG- of 3G-proeven. Het universitaire steunpunt gebruikt de naam centrale toetsen, de minister – en dus ook de administratie – spreekt van Vlaamse toetsen.

2. Regeerakkoord 2019-2024

De passages over de Vlaamse toetsen in het regeerakkoord zijn uitgebreid en gedetailleerd. Die keuzes, onder meer dat ze gestandaardiseerd, genormeerd en gevalideerd moeten zijn, bepalen hoe de toetsen er zullen uitzien. De proeven moeten voorts net- en koepeloverschrijdend zijn. Ze dienen te meten of leerlingen de eindtermen bereiken, maar ook – en dat is nieuw – de leerwinst op leerling-, school- en systeemniveau in kaart te brengen.

Tot nog toe geven peilingen en internationaal onderzoek enkel informatie op systeemniveau. De toetsen dienen voorts ontwikkeld te worden door een onafhankelijke instantie in samenwerking met onderwijsverstrekkers. Het regeerakkoord schrijft toetsen voor Nederlands en wiskunde voor, en specificeert voor Nederlands zelfs in begrijpend lezen, schrijven en grammatica. Het regeerakkoord stelt voorts dat er vier afnamemomenten zijn in de leerloopbaan van een leerling: twee keer in het basisonderwijs en op het einde van de eerste en derde graad secundair onderwijs. De resultaten worden op leerling- en schoolniveau teruggekoppeld naar de scholen en op individueel niveau en geanonimiseerd naar de overheid, de inspectie, de koepels en onderzoekers. In het vet staat in het regeerakkoord dat er geen schoolrankings komen. Vlaanderen blijft voorts deelnemen aan internationaal vergelijkend onderzoek. Scholen die minder goede scores halen, moeten verplicht een vrij te kiezen begeleidingstraject volgen. Een laatste element van het regeerakkoord is dat de leerkrachten gevormd worden om de vaardigheden te ontwikkelen om de toetsresultaten te lezen en te gebruiken.

3. Startfase van het beleidsproject

De administratie heeft deze principes van het regeerakkoord niet fundamenteel in vraag gesteld waardoor ze een beslissende invloed hadden op de startfase van het project. Vooreerst is er een nieuw universitair steunpunt opgericht voor de ontwikkeling van toetsen en feedbackinstrumenten, de onafhankelijke instantie waar het regeerakkoord over spreekt. Een andere optie was een overheidsagentschap geweest zoals in Denemarken, of commerciële toetsontwikkelaars zoals in Nederland of het Verenigd Koninkrijk. Een universitair steunpunt biedt de Vlaamse hoger-

onderwijsinstellingen evenwel de kans hun expertise in te zetten en op te bouwen. Omdat dit een nieuw beleidsinitiatief is waarover nog veel openstaande kennisvragen waren, is er een haalbaarheidsstudie opgestart in drie onderdelen: pedagogisch-psychometrische, organisatorische en technisch aspecten. De derde beginbeslissing is de studieopdracht aan de OESO met als doel te komen tot een grotere stakeholdersbetrokkenheid.

In de volgende fase is, in opvolging van de OESO-studie, een highlevelforum opgericht waarbij alle onderwijsstakeholders betrokken worden. De vele beslissingen voor de implementatie van Vlaamse toetsen, worden daar besproken. Voorts worden de bevindingen van de haalbaarheidsstudie opgevolgd. Een communicatieplan is in uitvoering. Ietwat onderbelicht is dat er ook al aan de IT-architectuur gewerkt wordt. De gunningsprocedure voor een digitaal toetsenplatform loopt.

4. Aanleiding: prestaties Vlaamse leerlingen

Bieke De Fraine legt uit dat de aanleiding van de toetsen de dalende trend van prestaties van de Vlaamse leerlingen is. Die resultaten worden gedistilleerd uit Vlaams en internationaal onderzoek. De spreker illustreert dit met enkele voorbeelden. Zo zijn de resultaten voor begrijpend lezen in het vierde leerjaar, zoals geregistreerd in PIRLS, met 22 punten gedaald op 10 jaar tijd, wat ongeveer overeenkomt met een half jaar instructietijd. Ook de resultaten op TIMSS dalen in het vierde leerjaar basisonderwijs. PISA-resultaten bij 15-jarigen tonen aan dat Vlaamse leerlingen op wiskundige geletterdheid 35 punten achteruitgegaan zijn op ongeveer 15 jaar. Een jaar lestijd levert ongeveer 20 PISA-punten op. Ook de Vlaamse peilingen wijzen op de dalende trend, hoewel die enigszins verschilt naargelang het domein. De spreker illustreert dat met het voorbeeld van de peiling wiskunde op het einde van het zesde leraar lager onderwijs. Bij zes van de negen onderdelen daalt het aantal leerlingen dat de eindterm haalt. De peiling wiskunde in de eerste graad van de A-stroom toetst negen onderdelen van wiskunde. Op drie van die onderdelen daalt het aantal leerlingen dat de eindterm haalt, op twee onderdelen stijgt dat, bij de overige vier blijft het stabiel. De peiling wiskunde in de eerste graad van de B-stroom toetst elf onderdelen. Bij negen van die elf onderdelen daalt tussen 2008 en 2019 het aantal leerlingen dat de eindterm haalt. Kortom, zorgwekkende trends, stelt Bieke De Fraine.

5. Doelstellingen Vlaamse toetsen

Het primaire doel van de Vlaamse toetsen is dan ook de onderwijskwaliteit en de prestaties van leerlingen te verbeteren, aldus *Jeroen Backs*. Toetsen op zich zullen de kwaliteit niet weer verhogen, dat gebeurt wel als er goed met de resultaten aan de slag gegaan wordt. Dat de kwaliteit van het onderwijs systematisch en jaarlijks zal worden gemonitord, is een stap vooruit ten opzichte van het internationaal vergelijkend onderzoek. Met de data kan de interne kwaliteitszorg van alle scholen ondersteund worden, bij peilingen en internationaal onderzoek alleen maar van scholen die deelnemen. Ook de inspectie krijgt met de toetsen extra data om bepaalde indicatoren van het Referentiekader voor Onderwijskwaliteit, onder andere over het bereik van de eindtermen en over leerwinst, met cijfers te onderbouwen. Eigen data maken het daarenboven mogelijk om, als de leerlingenprestaties dalen, sneller in te grijpen op systeemniveau. Leerlingen krijgen informatie om hun leerproces bij te sturen. Leraren kunnen met de data reflecteren over hun pedagogisch-didactisch handelen en de kwaliteit van het leerproces versterken.

De Vlaamse toetsen komen niet in de plaats van de kwaliteitsdriehoek van inspectiesysteem, pedagogische begeleiding en eindtermen, die nu aan de basis van het Vlaamse onderwijs ligt, maar worden erin ingebed. De eindtermen vormen immers het kader voor de toetsen. De inspectie gebruikt de resultaten van de toetsen voor haar externe kwaliteitscontrole. De pedagogische begeleiding gebruikt ze om de

interne kwaliteitszorg van scholen te versterken. De toetsen vervangen evenmin het Referentiekader voor Onderwijskwaliteit, maar dienen om twee indicatoren van dat kader met data te onderbouwen.

6. Welke toetsen?

De toetsen zullen ontwikkeld worden voor de leergebieden en vakken wiskunde en Nederlands. Ze zullen gebaseerd zijn op de eindtermen en zullen dezelfde zijn voor alle netten en koepels. De toetsen worden gemaakt door het Steunpunt Centrale Toetsen in het Onderwijs in samenwerking met denktanks van directeurs, leraren en pedagogische begeleiders. De belangstelling van leraren en directies om deel te nemen aan focus- en klankbordgroepen is bijzonder groot, wat de grote betrokkenheid aantoont.

De toetsen zijn digitaal en adaptief, wat betekent dat ze aangepast worden aan de mate waarin een leerling juiste of foute antwoorden geeft. Dat is motiverender voor de leerling, verkort de toetstijd en levert meer betrouwbare resultaten op.

7. Welke leerlingen?

Conform het regeerakkoord komen er toetsen op het einde van het tweede en zesde jaar van het secundair onderwijs, voor het lager onderwijs wordt gekozen voor het einde van het vierde en van het zesde leerjaar. Het gaat om ongeveer 70.000 leerlingen per leerjaar. De grootste test nu is PISA, waarbij er 5500 leerlingen getoetst worden. De toetsen worden zo inclusief mogelijk afgenomen, dus zijn er voor bepaalde leerlingen redelijke aanpassingen. Over de deelname van het buitengewoon onderwijs wordt nog volop nagedacht.

8. Timing

Bieke De Fraine zegt dat, na een bijsturing, de timing ondertussen duidelijk is. De eerste Vlaamse toetsen worden in mei 2024 afgenomen in het vierde jaar lager onderwijs en het einde van de eerste graad secundair onderwijs. In mei 2023 gebeurt er een kalibratie, een generale repetitie om onder meer de moeilijkheidsgraad van de toetsvragen te evalueren. Het zesde leerjaar lager onderwijs komt aan bod vanaf mei 2026. Vanaf dat moment kan de leerwinst in het lager onderwijs worden beoordeeld. Het einde van de derde graad start in mei 2027.

9. Vergelijking met internationale toetsen en peilingen

Jeroen Backs benadrukt dat alle leerlingen van alle scholen de Vlaamse toetsen zullen afleggen, terwijl dat bij internationaal onderzoek en in peilingen een representatieve steekproef is. De Vlaamse toetsen worden vier keer afgenomen in een leerloopbaan, bij internationaal onderzoek en peilingen verschilt dat naargelang het onderzoek, en nemen sommige leerlingen nooit deel. Voor de Vlaamse toetsen worden Nederlands en wiskunde getoetst; voor internationaal onderzoek is dat variabel: geletterdheid, wetenschappen, wiskunde, burgerschap, digitale competenties. Voor peilingen is dat telkens één vak in het secundair onderwijs en één leergebied in het basisonderwijs.

De doelen die getoetst worden, zijn bij de Vlaamse toetsen en de peilingen de eindtermen, bij internationaal onderzoek het internationale referentiekader, wat linken leggen met het Vlaamse onderwijscurriculum bemoeilijkt. Internationaal onderzoek en peilingen geven enkel feedback op systeemniveau en op schoolniveau aan de deelnemende scholen; de Vlaamse toetsen op leerling-, klas-, school- en systeemniveau aan alle scholen.

Met de Vlaamse toetsen is er geen internationale vergelijkingsbasis, vandaar is het nodig te blijven deelnemen aan internationaal vergelijkend onderzoek. Met de Vlaamse toetsen wordt dan wel weer een jaarlijkse trendanalyse mogelijk.

10. Impact van de resultaten

Bieke De Fraine vervolgt met de impact van de resultaten, die zich op drie niveaus situeert. Ten eerste mag de klassenraad de toetsresultaten gebruiken om een leerling te beoordelen, maar het mag niet het enige beoordelingscriterium zijn. De inspectie gebruikt de schoolresultaten bij haar toezicht. In het brede Referentiekader voor Onderwijskwaliteit wordt dat een van de informatiebronnen. Scholen met resultaten beneden de verwachting zullen hoogstwaarschijnlijk sneller een doorlichting krijgen. Zo daar reden toe is, krijgt de school een verplicht begeleidingstraject, niet als straf maar om overal goed onderwijs te verzekeren.

11. Teaching to the test

Een van de vaak genoemde nadelen van centrale toetsen is teaching to the test. Dat containerbegrip is te dissecteren in twee verschijningsvormen. Het eerste is teaching the test of itemteaching, bijvoorbeeld toetsen van vorige jaren gebruiken om leerlingen te trainen of leerlingen trucjes aanleren om op een strategische manier meerkeuzevragen te beantwoorden. Dat herleidt leerlingen tot papegaaien en is dus geen goed onderwijs. De tweede vorm, curriculumteaching of teaching to the standard, bereidt leerlingen met goed onderwijs voor op de toets, bijvoorbeeld door te oefenen op het interpreteren van staafdiagrammen of door leerlingen te trainen in het begrijpen van teksten. Als de toetsen een grote inhoudsvaliditeit hebben, is een dergelijke aanpak wenselijk.

12. Terugkoppeling van de resultaten

De resultaten worden teruggekoppeld door feedback te geven over de leerling, de klas, de school en het onderwijssysteem. De feedback over de leerling kunnen de leerling zelf, de ouders en de leraren gebruiken om het leerproces van de leerlingen te ondersteunen. Voor de klassenraad is het een informatiebron. Leraren kunnen de feedback over de klas gebruiken om te reflecteren over hun pedagogisch-didactisch handelen. Het schoolteam kan de schoolfeedback gebruiken als een informatiebron voor kwaliteitszorg. Voor de onderwijsinspectie is de schoolfeedback een informatiebron voor doorlichtingen en voor snelle interventie bij lage prestaties. De feedback over het Vlaamse onderwijs, op systeemniveau dus, zal toegankelijk zijn voor het grote publiek dat zo trends in leerlingenprestaties kan volgen. Om fair te zijn, moet feedback gerelateerd worden aan leerlingenkenmerken of eerdere prestaties.

13. Openbaarheid van resultaten

Jeroen Backs vervolgt met het heikele punt van de openbaarheid van de resultaten en de vrees voor classificaties, zoals in andere landen gebeurt. Het regeerakkoord is evenwel heel duidelijk: geen rankings! In het ontwerp van decreet zal staan dat de resultaten van de scholen niet openbaar worden gemaakt. Rangschikkingen zijn immers een simplistische voorstelling van de complexiteit van schoolkwaliteit. De spreker wijst erop dat het publiceren ervan negatieve effecten heeft. Het gevolg is dat de toetsresultaten niet zullen kunnen worden gebruikt voor schoolkeuze. Uiteraard blijven doorlichtingsverslagen openbaar.

14. Planning 2022

Op de planning van 2022 staan: de gunning van het toetsenplatform en de start van de IT-infrastructuur, de ontwikkeling van toetsen en feedbackinstrumenten,

het besluitvormingsproces over een decreet over de Vlaamse toetsen, de verdere uitrol van de communicatiestrategie, de uitwerking van tal van operationele beslissingen in samenwerking met het highlevelforum, de draaiboeken voor afnamemomenten en nadenken over de situatie na 2024.

III. Uiteenzetting door Johan van Braak

Johan van Braak is verbonden aan de vakgroep Onderwijskunde van de UGent. Hij geeft verschillende onderwijskundige vakken en doet onderzoek naar effectiviteitsvragen en de integratie van digitale technologie in het onderwijs. Vandaag spreekt hij als promotor-coördinator van het Steunpunt Centrale Toetsen in Onderwijs.

1. Van oproep naar oprichting

In de zomer van 2020 deed minister Ben Weyts een oproep voor de oprichting van een steunpunt voor gestandaardiseerde, genormeerde en gevalideerde toetsen. Dat de Vlaamse Regering de opdracht wou toekennen aan een academische partner, werd positief onthaald. Samenwerking met de academische gemeenschap is immers niet alleen een meerwaarde maar ook een voorwaarde voor een goed systeem van centrale toetsen. Op korte termijn is er dan een multidisciplinair team samengesteld met academisch geschoolden, vakinhoudelijke specialisten op het gebied van wiskunde en Nederlands, ervaren psychometrici, kwantitatieve analisten, experts in kwaliteitszorg en schoolfeedback. Op 15 december 2020 ging de gunning voor een periode van vijf jaar naar het steunpunt, met een beheersovereenkomst tussen de minister en Universiteit Gent en een samenwerkingsovereenkomst tussen de vijf Vlaamse universiteiten en twee hogescholen.

Het academische steunpunt is ervan overtuigd dat centrale toetsen in het leerplichtonderwijs een rol van betekenis kunnen spelen in een goed functionerend systeem van onderwijskwaliteit. Dat sluit aan bij het Referentiekader voor Onderwijskwaliteit, dat een set van kwaliteitsverwachtingen stelt. Centrale toetsen zullen de broodnodige evidentie leveren om uitspraken te doen over enkele van die kwaliteitsverwachtingen, zoals outputmaten op basis van de eindtermen op leerling- en schoolniveau maar ook, minstens even belangrijk, over kwaliteitszorg en kwaliteitsverbetering. Kortom, de centrale toetsen beantwoorden aan een informatie-behoefte. De resultaten zullen laten zien wat goed loopt en wat beter kan. Als aan een aantal voorwaarden wordt voldaan, zullen de centrale toetsen bijdragen tot onderwijsontwikkeling en zo leraren en schoolleiders ondersteunen bij hun waardevolle werk, meent de spreker.

2. Vijf kernopdrachten

De opdracht van het steunpunt is min of meer te herleiden tot vijf samenhangende kernopdrachten: toetsontwikkeling, ondersteuning bij de afname, analyse van de resultaten, feedback aan de gebruikers en expertiseontwikkeling in functie van duurzame onderwijs-effectiviteit.

2.1. Kernopdracht 1: centrale toetsen voor Nederlands en wiskunde

De eerste opdracht is dus toetsen ontwikkelen voor het vierde en zesde leerjaar van het basisonderwijs en voor het tweede en het zesde jaar van het secundair onderwijs. Dat de toetsen ontwikkeld worden op systeemniveau betekent dat in principe alle leerlingen in het gewoon onderwijs ze afleggen. In 2024 zal dit gebeuren bij een cohort van 920.000 leerlingen. In 2025 gebeurt voor het eerst een trendanalyse op systeemniveau. Die trendanalyse vergelijkt de resultaten van 2024 met die van 2025, voor het vierde jaar basisonderwijs en het tweede jaar secundair onderwijs. Vanaf 2026 worden dan ook progressie- of leerwinstanalyses

mogelijk van de leerlingen die in 2024 in het vierde leerjaar zaten, zowel op leerling- als op schoolniveau. Door vergelijking van de prestaties van dezelfde leerlingen in het vierde en het zesde leerjaar ontstaat een zuiver beeld van de relatieve bijdrage van de school in het leren van leerlingen. 2026 valt wel buiten de eerste beheerperiode van het steunpunt.

De toetsen worden gebaseerd op de eindtermen waardoor ze beter aansluiten bij het curriculum dan internationale toetsen, die per definitie niet gebonden zijn aan een curriculum van een land of regio. De eindtermen leggen sterk de nadruk op 21e-eeuwse vaardigheden: leerlingen moeten talig en wiskundig vaardig zijn, kritisch kunnen denken en allerlei problemen kunnen oplossen. De centrale toetsen sluiten nauw aan bij de Vlaamse onderwijsdoelen en dus bij datgene waarover eensgezindheid bestaat dat het belangrijk is om te leren.

Het steunpunt ontwikkelt met andere woorden op basis van de eindtermen rijke toetsen die niet gericht zijn op reproductie van kennis of vaardigheden waar leerlingen op een eenvoudige manier voor te bereiden zijn. Het zijn wel uitdagende levensechte en functionele toetsopdrachten waarvoor leerlingen inzicht, analytisch en redeneervermogen nodig hebben. Een dergelijke functionele insteek voor toetsen wordt ook aanbevolen in internationale testliteratuur. Die literatuur dient als basis om de toetsen wetenschappelijk te onderbouwen. Het staat buiten kijf dat de toetsen betrouwbaar en valide moeten zijn. De toetsresultaten mogen niet op toeval gebaseerd zijn en de toetsen moeten meten waarvoor ze ontworpen zijn. Een leerling die vandaag een toets begrijpend lezen aflegt zou hetzelfde resultaat moeten halen als op een toets een week later. Die toets moet ook echt de vaardigheid voor begrijpend lezen in kaart brengen en niet iets anders. Het multidisciplinaire team van het steunpunt staat in voor de wetenschappelijke ontwikkeling van de toetsen. Daarvoor lopen er al heel wat onderzoeken, straks volgen de kalibraties.

Toetsen moeten ook toegankelijk en eerlijk zijn, ook voor leerlingen die redelijke toetsaanpassingen nodig hebben. Daarover voert het steunpunt momenteel een onderzoek. De toetsen moeten ook eerlijk zijn in de zin dat geen enkele leerlingengroep benadeeld wordt. Dat betekent bijvoorbeeld dat een vraag even moeilijk of makkelijk moet zijn voor een jongen of een meisje van hetzelfde vaardigheidsniveau.

De toetsen zullen digitaal zijn, wat het mogelijk maakt ze adaptief te maken. Die toetsen passen zich dus automatisch aan aan het niveau van de leerling. Op basis van de antwoorden bepaalt het digitale toetssysteem welke volgende vragen gesteld worden. Adaptieve toetsen zijn wellicht ook motiverender, iets wat onderzocht zal worden. Leerlingen krijgen immers toetsenvragen die uitdagender zijn maar niet onbeantwoordbaar. Adaptieve toetsen zijn efficiënter en nauwkeuriger. Er is daarenboven minder toetstijd nodig om tot betrouwbare resultaten te komen.

Samenvattend, voor zijn eerste kernopdracht ontwikkelt het steunpunt rijke toetsen die digitaal worden afgenomen en die zo veel mogelijk aansluiten bij het niveau van de leerlingen. De toetsen zijn betrouwbaar, valide, eerlijk en toegankelijk. Ze zijn een hefboom voor onderwijsontwikkeling.

2.2. Kernopdracht 2: scholen ondersteuning bieden bij de afname van de centrale toetsen

Scholen en leerlingen moeten zich op het juiste moment voor de juiste toets kunnen aanmelden. Er komt een aanspreekpunt per school om het proces te coördineren. Er komen specifieke toetsaanpassingen voor bepaalde leerlingengroepen, zoals voorleessoftware voor leerlingen met dyslexie of extra toetstijd. Het nodige ondersteunende materiaal zoals draaiboeken, wordt ontwikkeld. Scholen worden ook vertrouwd gemaakt met de toetsomgeving. Leerlingen zijn digitale toetsen

weinig gewoon. De resultaten moeten de vaardigheden wiskunde of Nederlands reflecteren, niet de vertrouwdheid met digitale platformen of toetsen. Daarom zal het steunpunt ook kennismakingstoetsen maken.

2.3. Kernopdracht 3: de toetsresultaten verwerken en analyseren

De digitale afname heeft veel voordelen zoals de automatische scoring van resultaten en de snelle terugkoppeling naar gebruikers. Dat is ook nodig daar de toetsen aan het einde van het schooljaar afgenomen worden. Analyseren betekent ook nagaan in welke mate de leerlingen de doelen voor wiskunde en Nederlands bereiken. De data uit de toetsen zullen nadien ook gekoppeld worden aan data uit andere bronnen om op systeemniveau een inzicht te krijgen, wat verschillen in prestaties op niveau van leerlingen of van scholen verklaart.

2.4. Kernopdracht 4: het onderwijsveld feedback geven over de toetsresultaten

Door een vlotte en begrijpelijke terugkoppeling kunnen scholen met de toetsresultaten aan de slag. Feedback is een impuls voor interne kwaliteitszorg want helpt scholen om te zien wat goed loopt en wat beter kan. Het steunpunt zal feedback organiseren op vier niveaus: voor leerlingen, klassen, scholen en het systeem.

Het is belangrijk dat feedback aansluit bij de verwachtingen van de gebruikers. Om die reden organiseert het steunpunt gebruikersonderzoek bij bijvoorbeeld leraren en schoolleiders. Zo brengt het in beeld wat gebruikers verwachten van zowel de inhoud als de vorm van feedback. Op basis van dat onderzoek kan het steunpunt gebruiksvriendelijke feedback geven die scholen helpt de toetsresultaten te gebruiken voor kwaliteitsverbetering. Dat vergt evenwel datageletterdheid bij de gebruikers, niet alleen om de resultaten te interpreteren maar ook om ze in te zetten voor kwaliteitsverbetering. Het steunpunt organiseert dan ook professionalisering onder de vorm van e-courses over interpretatie van toetsresultaten en de aanwending ervan voor schoolontwikkeling. Dat gebeurt in samenwerking met andere actoren zoals pedagogische begeleidingsdiensten, legt de spreker uit.

2.5. Kernopdracht 5: expertise ontwikkelen op het vlak van duurzame onderwijs-effectiviteit

Deze opdracht is volgens Johan van Braak een complexe aangelegenheid. Voor eerst dienen de toetsen vergelijkingen in de tijd mogelijk te maken. Het steunpunt zal nagaan of de schoolresultaten veranderen doorheen de tijd. Het zal ook kijken naar eventuele fluctuaties en verklaringen ervan. Het zal onderzoeken hoe de scholen met de resultaten aan de slag gaan en welke acties die ze ondernemen, bijvoorbeeld wijzigingen in instructiepraktijken of extra onderwijstijd, verbetering teweegbrengen. Op termijn zal het nagaan of de resultaten van de scholen opschuiven in de gewenste richting. Het is niet alleen belangrijk gemiddelde scores te analyseren maar ook de kloof tussen de sterkste en de zwakste leerling. Op dat vlak is Vlaanderen de laatste jaren niet de beste van de klas. Met de centrale toetsen kan dat op systeemniveau gemonitord worden.

Samenvattend gaat het steunpunt rijke toetsen ontwikkelen, toetsafnames coördineren, toetsresultaten analyseren en terugkoppelen naar de gebruikers en onderzoeken in welke mate het nieuwe beleidsproject rendoert in functie van het doel van kwaliteitsverbetering.

3. Acht werkdomeinen

Om de kernopdrachten uit te voeren, werkt het steunpunt samen met vijf universiteiten en twee hogescholen. De werkzaamheden zijn opgedeeld in acht met elkaar verbonden werkdomeinen, verdeeld over zeven instellingen. De meeste

werkdomeinen zijn een samenwerkingsverband tussen verschillende instellingen. Zo wordt de toetsontwikkeling wiskunde gecoördineerd door de Vrije Universiteit Brussel, maar voor het lager onderwijs is de KU Leuven partner. De toetsontwikkeling Nederlands wordt gecoördineerd door de Universiteit Gent met de Arteveldehogeschool als partner voor het lager onderwijs. Feedback wordt gecoördineerd door de Universiteit Antwerpen met de universiteit Hasselt, de Vrije Universiteit Brussel en de AP Hogeschool Antwerpen als partners.

4. Samenwerking met stakeholders

De multidisciplinaire samenwerking tussen de instellingen is een voorwaarde en een meerwaarde om de centrale toetsen als vernieuwend beleidsproject in de steigers te zetten. De teams werken niet alleen samen met academici, maar ook met professionals die jarenlang ervaring hebben in het onderwijs. Bij de aanpak van het steunpunt vormt samenwerking met het onderwijs het DNA. Het ontwikkelt toetsen voor en met het onderwijsveld, in plaats van pas op het einde feedback te vragen over toetsen die al klaar zijn. Gedurende de hele rit worden de stakeholders betrokken. Het enthousiasme van het onderwijsveld blijkt uit het grote aantal deelnemers, namelijk meer dan vierhonderd.

Vooreerst zijn er denktanks over eindtermselectie en toetsenontwikkeling, tot nog toe acht, met verschillende doelgroepen uit het lager en het secundair onderwijs: leraren, schoolleiders, lerarenopleiders, pedagogische begeleiders en onderwijsinspecteurs. Daarnaast zijn er focusgroepen over de haalbaarheid en wenselijkheid van redelijke toetsaanpassingen, bijvoorbeeld of leerlingen die recht hebben op hulpmiddelen bij wiskunde die mogen gebruiken bij de centrale toetsen. Inmiddels hebben twintig focusgroepen plaatsgevonden van verschillende doelgroepen uit zowel lager als secundair onderwijs, bijvoorbeeld zorgcoördinatoren, CLB-medewerkers en vakleerkrachten. De resultaten van die gesprekken worden momenteel geanalyseerd. Ten derde zijn er gebruikersgroepen over feedback. In die gesprekken wordt gepeild naar de inhoud en de vorm die leraren en schoolleiders verwachten van de feedback. Tot nog toe waren er acht gesprekken met een twintigtal deelnemers en heel wat individuele interviews.

5. Besluit

Afsluitend vat de spreker de aanpak van het steunpunt samen: participatie en betrokkenheid van het onderwijsveld; fasering zodat het onderwijsveld tijd, vertrouwen en ondersteuning krijgt om ermee vertrouwd te geraken en expertise op te bouwen. De toetsen zijn gericht op onderwijsontwikkeling, een middel voor krachtig onderwijs voor elke leerling. De spreker benadrukt dat de toetsen niet vanuit verantwoordingsgericht perspectief bekeken worden.

IV. Uiteenzetting door Rianne Janssen

Rianne Janssen is promotor-coördinator van het Steunpunt Toetsontwikkeling en Peilingen (STEP). Voor het Steunpunt Centrale toetsen is ze werkdomeincoördinator voor psychometrie. Tot slot was ze ook promotor van het organisatorische onderdeel van de haalbaarheidsstudie. Vandaag geeft ze namens STEP, dat zijn laatste werkingsjaar ingaat, zeven reflecties over de centrale toetsen.

In juni 2021 zei minister Ben Weyts in zijn voorwoord bij het vragenuurtje over de Vlaamse toetsen: "(...) dat we niet langer moeten terugvallen op internationale onderzoeken, zoals PISA en PIRLS, om te weten hoe de kwaliteit van het Vlaamse onderwijs evolueert."

1. Les 1: peilingsonderzoek 2002-2022

De spreker herinnert aan het specifieke Vlaamse peilingsonderzoek dat al twintig jaar, naast het internationale vergelijkingsonderzoek, georganiseerd wordt. De peilingen zijn gestart in 2002 en vanaf 2004 werden ze jaarlijks afgenomen, zowel in het basisonderwijs als in het secundair onderwijs. In totaal waren er vijfendertig peilingen. De laatste twee peilingen over de domeinen Nederlands en wiskunde in de eerste graad overlappen al met het inhoudelijke domein en met een deel van het doelpubliek van de centrale toetsen. Van de zestien peilingen in het basisonderwijs gingen een kwart over wiskunde en een kwart over Nederlands. In het secundair onderwijs gingen acht van de negentien peilingen over wiskunde, twee over Nederlands en twee over project algemene vakken (PAV), dat zowel elementen van wiskunde als van Nederlands omvat.

De doelstellingen van de peilingen liggen in de eerste plaats op systeemniveau. Ze zijn bedoeld om een criteriumgerichte uitspraak te doen over de mate waarin de leerlingen de eindtermen behalen. Daarbij wordt ook onderzocht of er in dat behalen van de eindtermen verschillen zijn tussen leerlingen, bijvoorbeeld tussen meisjes en jongens, tussen leerlingen met een bepaalde thuisachtergrond, tussen leerlingen van studierichtingen en eventueel onderwijsvormen in het secundair onderwijs. Als er een herhalingspeiling plaatsvindt, worden ook de evoluties in kaart gebracht. Naast de prestaties op systeemniveau, leveren de peilingen ook informatie over de prestaties van de school, zowel feitelijke prestaties als prestaties gecorrigeerd voor achtergrondgegevens. Bij elke peiling krijgt de school criteriumgerichte feedback met een samenvattend overzicht van de leerprestaties van hun leerlingen, maar ook van de prestaties van de school, namelijk hoe goed ze het doet in vergelijking met andere scholen en met vergelijkbare scholen.

Daarnaast is er een systeem van paralleltoetsen. De school kan zelf, in het kader van de interne kwaliteitszorg, vergelijkbare toetsen afnemen en zo vergelijkbare schoolfeedback als bij de peilingen verkrijgen. Alleen bij de paralleltoetsen is er feedback op leerlingniveau, dus over de mate waarin de leerlingen afzonderlijk de eindtermen halen.

2. Les 2: peilingen en centrale toetsen: gelijkenissen en verschillen

De tweede reflectie is dat peilingen en centrale toetsen allebei systemen van externe kwaliteitszorg zijn. Ze meten kwaliteit daarbij niet aan de hand van het proces maar op basis van toetsprestaties die verwijzen naar het behalen van de onderwijsdoelen. De peilingen gebeuren bij een steekproef van scholen, aan de centrale toetsen nemen alle scholen deel. Beide systemen vinden plaats in verschillende onderwijsniveaus en verschillende onderwijsvormen. De centrale toetsen gaan enkel over wiskunde en Nederlands, de peilingen over verschillende leer- en vakgebieden. De doelstellingen zijn deels gelijksoortig, namelijk nagaan of de onderwijsdoelen gehaald zijn, de evoluties daarin vaststellen en scholen vergelijken zonder rankings te maken. Centrale toetsen hebben daarnaast de expliciete doelstelling om leerwinst in kaart brengen. De focus van de peilingen is het systeemniveau, maar ze geven ook informatie op schoolniveau en, via de paralleltoetsen ook op leerlingniveau. Bij de centrale toetsen ligt de focus op school- en leerlingniveau. Toch zal ook informatie op systeemniveau belangrijk zijn bij de opvolging.

3. Les 3: toets wiskunde en Nederlands

De derde reflectie is dat de centrale toetsen enkel wiskunde en Nederlands toetsen. Ze dienen dus niet de kwaliteitszorg van de leerdoelen van het volledige curriculum, maar slechts van weliswaar twee belangrijke leer- en vakgebieden. Het is daarnaast een uitdaging om de omvangrijke curricula van Nederlands en wiskunde

te vertalen in toetsen. Daarvoor bestaan verschillende strategieën, vooreerst een selectie van wat uit het curriculum getoetst moet worden. De opdrachtgever heeft al enige selectie aangebracht in de toetsen Nederlands, hoewel het omvangrijke onderdelen blijven: leesvaardigheid, maar geen luistervaardigheid, schrijven en grammatica.

Een andere strategie is die van het peilingsonderzoek: de onderwijsdoelen opsplitsen in een groot aantal specifieke toetsen die telkens gaan over een beperkt aantal eindtermen. Bij de peilingen is het curriculum wiskunde in het basisonderwijs bijvoorbeeld gevat in 21 toetsen. Omdat het om steekproeven gaat, moeten niet alle leerlingen alle toetsen afleggen. Tot slot is een laatste strategie om te kiezen voor een globale toets die zo veel mogelijk de onderwijsdoelen dekt, maar die dan mogelijk minder informatieve waarde heeft.

Kortom, het steunpunt staat voor de uitdaging om een optimum te zoeken tussen een beperkte toetsduur, de uitspraken over verschillende niveaus en de leerwinst-meting.

4. Les 4: behalen van de onderwijsdoelen

Om het behalen van de onderwijsdoelen te beoordelen, werken de peilingen met de logica van een meetschaal waarin de opgaven op basis van psychometrische modellen omgezet worden in een soort toetsladder. Naargelang de moeilijkheidsgraad wordt een item hoger op die ladder geplaatst. De leerlingen worden op dezelfde toetsladder geplaatst in toenemende gradatie van vaardigheid. Om te bepalen wanneer leerlingen de eindtermen halen, trekken experts op basis van een inhoudelijke analyse aan de itemkant een lijn tussen de opgaven die leerlingen moeten beheersen om het minimumniveau te halen en die opgaven die de leerlingen nog niet moeten beheersen. Aan de leerlingkant komt er zo een tweedeling tussen leerlingen die de eindtermen halen en zij die ze nog niet halen. Die werkwijze is zeker transposeerbaar naar de nieuwe onderwijs-toetsen. Voor de laatste peiling wiskunde en Nederlands in de eerste graad secundair onderwijs gebaseerd op de nieuwe onderwijsdoelen, gelden daarbij niet één maar verschillende cesuren: basisgeletterdheid, eindtermen B-stroom, eindtermen A-stroom en uitbreidingsdoelen.

Een toets doet zo een algemene uitspraak over de mate waarin de leerling het beoogde onderwijsdoel haalt. Het is geen diagnostisch instrument voor remediëring, het is geen leerlingvolgsysteem. Door de lat constant te houden, kunnen scholen er wel evoluties mee vaststellen in de mate waarin hun leerlingen de eindtermen halen. De onderwijsminister denkt dat de delibererende klassenraad de resultaten van de centrale toetsen kan gebruiken bij het beoordelen van individuele leerlingen. Dat vereist evenwel een hoge nauwkeurigheid van de uitspraken over individuele leerlingen. Daarom werkt het steunpunt aan toetsen op maat, adaptieve toetsen die de meetnauwkeurigheid garanderen.

5. Les 5: verschillen tussen scholen en schoolfeedback

Een tweede belangrijke doelstelling naast het behalen van de onderwijsdoelen is verschillen tussen scholen vaststellen en de scholen daarover feedback geven. Bij het peilingsonderzoek wordt er op twee manieren feedback gegeven aan de scholen. Vooreerst krijgen de scholen de ruwe prestatieverschillen binnen een rangorde van schoolgemiddelden die niet publiek gemaakt wordt. Rond elk schoolgemiddelde is er een onzekerheids- of betrouwbaarheidsinterval. Dat interval wordt ook gebruikt om uitspraken te doen over de prestatie ten opzichte van het globale gemiddelde. Zo kan er een onderscheid gemaakt worden tussen scholen die goed presteren en scholen die minder goed presteren. Dat is bedoeld als een spiegel voor de school.

Een tweede element van schoolfeedback bij het peilingsonderzoek is om die spiegel te nuanceren of zelfs te corrigeren. STEP gebruikt daarvoor een nettomodel dat de toegevoegde waarde van een school in kaart brengt, bovenop de prestatie die verwacht kan worden, gegeven bepaalde kenmerken van school en populatie. Leerprestaties worden daarvoor gerelateerd aan achtergrondkenmerken van leerlingen, de compositiekenmerken van de school en de niet-beïnvloedbare schoolkarakteristieken. Het nettomodel is dus een meer informatieve spiegel. Het is zaak om voor de centrale toetsen de variabelen waarvoor gecorrigeerd wordt, in consensus te kiezen. Die variabelen bepalen immers de toegevoegde waarde van scholen.

Het peilingsonderzoek leert dat de schoolverschillen kleiner worden als de vergelijking eerlijker wordt, dus als er meer achtergrondkenmerken in rekening worden gebracht. STEP heeft weinig verklaringen gevonden voor de schoolverschillen in leerkracht- of schoolvariabelen. Omdat het over de resultaten van de hele populatie en niet slechts van een steekproef zal beschikken, kan dat misschien anders zijn voor de centrale toetsen.

De betrouwbaarheidsintervallen laten zien of een school al dan niet significant afwijkt van een populatie of een verwacht gemiddelde. Zelfs als er geen verschillen zijn in kwaliteit tussen de scholen, zal deze methode toch enkele scholen als afwijkend diagnosticeren, de zogenaamde type-1-fout. Ook als er geen schooleffecten zijn, is er 5 procent afwijking te verwachten. Vooraleer scholen te verplichten tot een begeleidingstraject naar keuze is het goed de stabiliteit van het vastgestelde schooleffect na te gaan. In Nederland wordt het resultaat van drie opeenvolgende jaren beoordeeld alvorens te besluiten dat een school afwijkende resultaten heeft. Terzijde merkt de spreker op dat het allicht goed is om, zoals de minister zegt, ook naar boven te kijken en de inspectie of de begeleidingsdienst te laten nagaan of er iets te leren valt van scholen die significant beter presteren dan verwacht.

Schoolverschillen zijn een signaal, maar verdere duiding, diagnose, handvatten voor bijsturing zijn nodig. In STEP, maar zeker ook door het nieuwe steunpunt, wordt onderzoek gedaan naar het begeleiden van scholen bij het omgaan met schoolfeedback.

6. Les 6: het belang van het systeemniveau

Het schoolniveau dient om een spiegel voor te houden en zo het onderwijs te verbeteren. Toch is volgens STEP het systeemniveau een belangrijk element in de omgang en het bewaken van de impact van de centrale toetsen. De spreker illustreert dat met het onderdeel 'rekenen met veeltermen' van de peiling wiskunde in de eerste graad secundair onderwijs van 2018. Een beperkt aantal scholen scoorde daarop inderdaad lager dan andere scholen, maar in totaal haalde slechts 28 procent van de leerlingen deze eindterm. De oorzaken van dit algemeen lage resultaat moet dus ook op systeemniveau gezocht worden. De verantwoordelijkheid kan niet alleen naar scholen toegeschoven worden.

De peilingen zijn slechts een onderdeel van de kwaliteitszorg, de check van de 'plan do check act'-logica, namelijk evaluatie en feedback. Soms heeft de spreker het gevoel dat de impact van de peilingen – ondanks studiedagen, colloquia, consensusconferenties enzovoort – op veranderingen in het onderwijs beperkt geweest zijn. Tot nog toe kende Vlaanderen bijvoorbeeld nog geen schokeffect zoals de PISA-resultaten in Duitsland teweegbrachten. Toch leidt het dalende niveau tot ongerustheid, wat precies de reden is waarom er nu centrale toetsen komen. Het is goed dat de commissie-Brinckman zoekt naar oorzaken. Om de resultaten van centrale toetsen te gebruiken om oorzaken te bepalen, is er nog bijkomend onderzoek nodig, zeker voor de verantwoordelijkheid op een hoger niveau dan de scholen.

7. Les 7: leerwinst

De zevende observatie is dat een belangrijke toevoeging van de centrale toetsen de focus op leerwinst is. Om vooruitgang te meten, is het nodig dezelfde meet-schaal te hanteren: if you want to measure change, don't change the measure. Ook de leerwinst van scholen kan gemeten worden: hoe ver brengen de scholen de leerlingen, gegeven van waar ze komen. Scholen met een sterke instroom halen doorgaans de eindtermen gemakkelijker. Scholen met een minder sterk leerlingen-publiek halen de maatstaf veel moeilijker, ook al bereiken ze een grotere leerwinst. Het is dus zaak om ook de verschillen in leerwinst tussen de scholen te analyseren en de toegevoegde waarde in leerwinst per school weer te geven.

Een dergelijke longitudinale opvolging is statistisch complexer, onder meer door leerlingen die van school veranderen en door de niet-normaalvorderende leerlingen. Een andere zorg is de haalbaarheid van een gemeenschappelijke meet-schaal voor de vier toetsmomenten in lager en secundair onderwijs. Het SiBO-onderzoek (een longitudinaal onderzoek in het basisonderwijs), PIRLS- en TIMSS-repeat bewijzen dat een gelijke meetschaal in het vierde en zesde leerjaar haalbaar is. In 2002 deden de KU Leuven en de UAntwerpen in opdracht van toenmalig onderwijsminister Marleen Vanderpoorten een onderzoek naar leerwinstmetingen in de eerste graad van het secundair onderwijs A- en B-stroom voor wiskunde en Nederlands. Dat leverde gemeenschappelijke meetschalen op voor de A- en de B-stroom, maar niet voor het einde van het basisonderwijs en het einde van de eerste graad. Nochtans zou het interessante informatie opleveren om de leerresultaten op het einde van de eerste graad te corrigeren met het niveau van de instroom uit het lager onderwijs. Of er een gemeenschappelijke meetschaal mogelijk is voor het einde van de eerste graad en het einde van de derde graad, is zeer onzeker. De longitudinale analyse van studieloopbanen in het secundair onderwijs (LiSO) is erin geslaagd om een overkoepelende meetschaal te maken, maar enkel omdat er elk jaar toetsen werden ontwikkeld zodat de ankerings op de gemeenschappelijke meetschaal jaar per jaar kon worden aangepast.

Zelfs als er geen gemeenschappelijke meetschaal gevonden wordt, is het vanuit het oogpunt van effectiviteitsonderzoek toch een winst om eerdere prestaties van leerlingen mee te nemen bij het bepalen van de toegevoegde waarde van scholen, omdat die indicator het dichtst bij de schooleigen bijdrage komt.

8. Besluit

Kortom, zowel centrale toetsen als peilingen zijn systemen van externe kwaliteitszorg, met het grote positieve verschil dat bij de centrale toetsen alle scholen onderzocht worden. Hoewel belangrijk, zijn Nederlands en wiskunde slechts onderdelen van het onderwijs. De centrale toetsen leveren informatie op leerling-, school-, en systeemniveau, maar hebben vooral een signaalwaarde. Duiding en een daaropvolgende bijsturing van het onderwijs zijn nodig. Leerwinst is een belangrijke toevoeging om de onderwijskwaliteit te bepalen. Als er geen gemeenschappelijke meetschaal is, dient bij het beoordelen van verschillen en het feedback geven, rekening gehouden te worden met voorafgaande prestaties. Centrale toetsen leiden op zich niet tot een betere onderwijskwaliteit, stelt Rianne Janssen. De echte revolutie zou zijn erin te slagen om de leerlingprestaties te verhogen. Daartoe vragen toetsprestaties zowel duiding als acties, dus borging en bijsturing. Op schoolniveau slaagt dit alleen maar als het systeem van externe kwaliteitszorg dat de centrale toetsen zijn, door scholen wordt gepercipieerd als een onderdeel van interne kwaliteitszorg. Ook op systeemniveau is een opvolging van de toetsresultaten nodig. De verantwoordelijkheid mag niet louter bij de scholen gelegd worden.

V. Uiteenzetting door Hans Teegelbeckers

Hans Teegelbeckers is directeur van de vereniging voor bestuur, management en medezeggenschap in het openbaar en algemeen toegankelijk onderwijs (VOS/ABB) in Nederland, die de belangen van het onderwijs in brede zin, het openbaar onderwijs, vertegenwoordigt, op politiek vlak maar ook bij de schoolbesturen zelf. Hij heeft 27 jaar lesgegeven, is daarna in het management terechtgekomen en is sinds 5 jaar ook directeur van VOS/ABB.

1. Nederland is toetsgedreven

Velen in Nederland zouden het eens zijn met wat hij zal zeggen over de zin en onzin van toetsing, maar velen zouden het er ook helemaal niet mee eens zijn. Dat zal in Vlaanderen niet anders zijn. In Nederland heerst een soort toetsvirus. De Nederlandse controledrang komt tot uiting in een brede obsessie voor testen en toetsen. Vanaf groep 3 van de basisschool worden kinderen 50 tot 75 keer langs de meetlat gelegd. Iedere school moet een leerlingvolgsysteem hebben en de inspectie ziet toe op de hantering ervan.

Dat levert een grote werkdruk op voor de leraar, iets waar Vlaanderen ook mee kampt. Ook in Nederland is er een groot lerarentekort dat de komende jaren alleen maar zal toenemen. Steeds meer weerklinkt de roep om meer professionele ruimte voor de leraar om betekenisvolle lessen te maken en te geven. Dat gaat niet alleen over kwalificatie maar ook over andere dingen die leerlingen aangaan.

De toetsenkalender van de niet-lesgebonden toetsen van groep 5 tot en met 8 toont volgens de spreker de doorgeslagen Nederlandse toetsdrang aan. Ook bij kleuters worden er allerhande toetsen afgenomen.

2. Waarom toetsen?

De hoofdvraag is waarom er getoetst wordt. In de eerste plaats dienen toetsen ter objectivering van wat een kind kent en kan, onder andere om keuzeprocessen voor het vervolgonderwijs te begeleiden. In dergelijke toets spelen vooroordelen van ouders en leraren geen rol. In die zin geeft een objectieve centrale toets minder kans op sociale ongelijkheid. In Nederland worden de toetsen echter ook gebruikt om rankings te maken en om scholen, via de inspectie, de stempel zwak of zeer zwak te geven. Die lijsten staan ook op het internet. Scholen die zeer zwak zijn, moeten een programma opstellen, een plan van aanpak, dat ze ook aan de ouders moeten meedelen. Die scholen worden dus aan de schandpaal genageld.

3. Centrale toetsing in Nederland

In Nederland is er een centrale toetsing in groep 8, het einde van de basisschool, voor Nederlandse taal en rekenen. Dat kan schriftelijk, maar sinds kort ook digitaal. Scholen kunnen beslissen om een facultatief onderdeel over wereldoriëntatie af te nemen. Vooraf, voor 1 maart, geeft de leraar die de kinderen al acht jaar kent, die de film van hun schoolloopbaan zag, een schooladvies over het vervolgonderwijs of het voortgezet onderwijs, wat in Vlaanderen het secundair onderwijs genoemd wordt. Vervolgens wordt er in april, of als het digitaal is in mei, een centrale toets afgenomen. Dat is een momentopname, een foto van hoe het kind die twee dagen presteert. Zaken als werkhouding, attitude, zelfverzekerdheid, zelfstandigheid, sociale en emotionele ontwikkeling worden niet getoetst. Het is louter een kwalificerende toets op rekenen en taal. Als de toets beter is dan het schooladvies kunnen de ouders een hoger advies vragen voor de vervolgschool, een zogenaamde heroverweging, waarna er een definitief advies volgt. Tussen 1 maart en april kunnen de meer bemiddelde ouders huiswerkbureaus inhuren om hun kinderen bij te

schaven. Ze kunnen zelf aan teaching to the test doen of dat uitbesteden. Dat vergroot de kans op afwijkingen tussen schooladvies en toetsresultaten. Met een beter resultaat kunnen ze de leraar onder druk zetten.

Voorheen stond het CITO in voor de centrale toets. Scholen kunnen inmiddels kiezen uit vijf verschillende centrale toetsen, wat vergelijken moeilijk maakt. Dat is evenwel afgedwongen door het onderwijsveld omdat de inrichting van onderwijs en dus ook de toetsafname in Nederland vrij zijn. Voor het voortgezet onderwijs zijn er centrale examens, tentamens en centrale toetsen.

4. Wat is er tegen centrale toetsing?

Centrale toetsen zijn zoals gezegd momentopnames, foto's. De school volgt de leerling vele jaren, acht in Nederland en zes in Vlaanderen. De leraren kennen de leerlingen, hun attitudes, emotionele toestand en doorzettingsvermogen. Al die zaken zijn van invloed op de carrière van de leerling.

Als een school drie jaar na elkaar een lager resultaat heeft, is dat te zien op de internetlijstjes. Hoewel het er niet zo veel zijn, worden ze erop afgerekend, want ouders sturen hun kinderen niet gemakkelijk naar een school die zwak of zeer zwak scoort.

Een ander nadeel is dat toetsen geld kosten en heel veel tijd vergen: voor aankoop, afname, onderzoek en implementatie. De spreker raadt de parlementsleden aan de kosten te berekenen en daar nog eens goed over na te denken.

In Nederland is er een discussie om de centrale toetsing te verlaten, omdat de selectie aan het einde van groep 8 veel te vroeg is. Misschien is het zelfs pas in een eerste of tweede brugjaar zinvol. In Nederland komt er allicht een brede brugklas waardoor toetsmomenten uitgesteld worden tot beter kan worden bepaald wat leerlingen kennen en kunnen. Kinderen verschillen immers enorm: snelle en langzame, vroeg- en laatbloeiers, hoog- en onderpresteerders enzovoort. Het kan best zijn dat een kind laag scoort op een toets en toch goed terecht komt. De spreker ziet zichzelf als laatbloeier. Hij komt uit een arbeidersgezin, had moeite met lezen en kreeg het advies om naar een lagere technische school te gaan. Hij volgde echter met succes middelbaar algemeen voortgezet onderwijs (MAVO) en is leraar geworden.

Methodegebonden toetsen zijn formatieve toetsen en geven de leraar een beeld op de vorderingen in de aangeboden lesstof. Een goede leraar kent zijn kinderen en houdt hun ontwikkeling in de gaten. Niet-methodegebonden toetsen houden het risico in dat er leerstof getoetst wordt die nog niet in de klas aan bod kwam. Methoden verschillen immers, ook al voldoen ze allemaal aan de eindtermen. Methodegebonden toetsen die voldoen aan de eindtermen, geven de leraar voldoende informatie om de kinderen te volgen. Zelfs een centrale toets is overbodig. Zo nodig kan de leraar pedagogisch-didactisch begeleid worden. Vlaamse en Nederlandse leerlingen doen het steeds minder goed in begrijpend lezen, leesvaardigheid en leesbereidheid. VOS/ABB is tot de conclusie gekomen dat de methodes en lessen voor begrijpend lezen verworden zijn tot structuurtaallesjes waar kinderen niet alleen een hekel aan hebben maar die er daarenboven niet voor zorgen dat ze beter leren lezen of begrijpend lezen. Het lesmateriaal kent arme teksten, terwijl rijke teksten en voorlezen het leesplezier verhogen, wat de lees kwaliteit zou bevorderen.

De Algemene Onderwijsbond, een van de grootste onderwijsvakbonden, vroeg in 2017 aan meer dan 4000 leraren uit het primair onderwijs en bijna 4000 uit het voortgezet onderwijs naar hun activiteiten en tijdsverdeling. Leraren in het primair onderwijs met een voltijdse lesopdracht hebben een normjaartaak van 1659 uur in

41,3 weken van 40 uur, lesuren en voorbereidingstijd inbegrepen. In werkelijkheid presteren leraren veel meer uren, gemiddeld 46,9 uur per week. In het primair onderwijs is een leraar gemiddeld ongeveer 8,2 uur per week met toetsing bezig: voorbereiden, afnemen, nakijken en analyseren. De spreker vermenigvuldigt die toetstijd per week met het aantal werkweken en het aantal jaren primair onderwijs, acht in Nederland, zes in Vlaanderen, wat een totaal van 2709 uur toetsen in Nederland en 2032 uur toetsen in Vlaanderen oplevert. Met die tijd zijn heel veel nuttige andere dingen te doen, waar leerlingen misschien veel beter van zouden worden.

5. Samenvattend advies

Doel van de Vlaamse toetsen is blijkbaar de onderwijskwaliteit te verbeteren. Toetsen afnemen en analyseren levert evenwel geen informatie op over de onderwijsprocessen. Dat vergt een diepgaandere analyse of de aanpak van de leraar, maar ook een methode die voldoende afgestemd is op de leerdoelen. Met de toetsen wil Vlaanderen ook de leerwinst in kaart brengen, maar dat is iets wat een school al moet doen. Of de leerlingen de doelen bereiken, kan ook gemeten worden met methodegebonden toetsen. Volgens de spreker is het belangrijker de interne kwaliteitszorg te verbeteren dan scholen met de toetsen een extra extern hulpmiddel te bieden. Hij raadt een goede evaluatie van de processen, van de kwaliteitszorg en goede communicatie aan. Daar groeien leraren van en zo leren ze keuzes maken. Als de toetsresultaten beneden de verwachtingen zijn, wil Vlaanderen de scholen een begeleidingsprogramma laten volgen. Dat is de taak van de directies, de interne begeleidingsdienst maar ook van externe ondersteuning.

Kortom, toetsen zijn een hulpmiddel en geen doel op zich. Hans Teegelbeckers hoorde vandaag veel mooie woorden over het project, over onderzoek, maar het kind lijkt wat verloren gegaan te zijn. Wat levert het werkelijk op? Het kost in ieder geval veel tijd en veel geld. Centrale toetsing levert sowieso rankings op. Het publiek maken ervan wordt dan wel door de huidige minister en in een decreet verboden, maar wie zegt dat het zo blijft? Lekken er geen gegevens uit? Ook bij het feit dat de centrale toetsen direct bijdragen tot kwaliteitsverbetering heeft hij vragen.

Het Nederlandse regeerakkoord stelt zelfs voor om toetsen voor kleuters te schrappen. Ook uit verschillende onderzoeken van de Nederlandse Onderwijsraad blijkt dat er veel te vaak wordt getoetst en te veel met gesloten vragen, maar ook dat de resultaten van kinderen beter op andere manieren getoetst en gevolgd worden. De spreker vat zijn ideeën samen: stel niet de PISA-lijst maar het kind centraal en baseer je niet op een foto, een momentopname, maar op de volledige schoolloopbaan, de film.

VI. Bespreking

1. Tussenkomsten van de leden

1.1. Loes Vandromme

Loes Vandromme uit vooraf haar verwondering over de perstoelichting, waar ook enkele van de sprekers van deze hoorzitting aan deelnamen. Dat is een voorafname op de parlementaire discussies. Is dat een onderdeel van het communicatieplan? Hoe schatten de deelnemers hun participatie aan dat persmoment in? Wat was het doel precies? CD&V heeft sterk aangedrongen op deze hoorzitting, vooral wegens de bezorgdheden die in het Vlor-advies staan.

Kwaliteit hoort tot de hoogste onderwijsprioriteiten, naast bijvoorbeeld het leraar-
rentekort. In de presentaties mist het parlementslid aandacht voor wat al bestaat,
bijvoorbeeld de peilingen. Het lijkt wel of Vlaanderen helemaal geen toetsen kent.
Visies veranderen en de OESO suggereert die veranderingen te inventariseren.
Maatregelen kunnen maar een effect hebben als ze ingebed zijn in het beleid van
de school zelf. Ze staat vooreerst stil bij de autonomie van het evaluatiebeleid. In
de presentatie van de OESO werd het belang van een evaluatiecultuur benadrukt.
Hoe belangrijk is dat? En zijn er daarin grote verschillen tussen de Europese lan-
den? Welke uitdagingen ziet de OESO in de Vlaamse evaluatiecultuur?

Voor CD&V is de autonomie van de klassenraad cruciaal. De minister heeft het zelfs
over het herstel van die autonomie. Hoe verhinderen dat een beslissing van de
klassenraad aangevochten wordt door ouders omdat de leerling wel slaagt voor
een Vlaamse toets? Hoe zien de sprekers het risico op juridisering? Zijn er in andere
Europese landen lessen te leren over hoe centrale toetsen de relatie tussen school
en ouders veranderen?

De Vlor mist in de plannen een ontwikkelingsgerichte visie op kwaliteit. De raad
vindt dat scholen bovenal moeten inzetten op interne kwaliteitszorg. Hoe zien de
sprekers de verhouding tussen een ontwikkelingsgerichte aanpak en responsabili-
serend werken? Klopt het dat het onderwijsveld de prioriteit niet kent? Hoe kunnen
de schoolteams geholpen worden met interne kwaliteitszorg? Hoe kunnen school-
of methodegebonden toetsen daartoe bijdragen? Wat vinden de sprekers van de
opmerking dat de opzet van sommige toetsen te breed is, bijvoorbeeld zowel kwa-
liteitsborging als studieoriëntatie.

De Vlaamse toetsen gaan dan wel over de eindtermen, maar slechts over twee van
de zestien sleutelcompetenties. En die twee competenties zijn zo breed dat het
moeilijk wordt om er de volledige scope van te testen. Hoe wordt dat het best
ondervangen? Daarbij komt nog dat onderwijs gericht is op een evenwicht tussen
kwalificeren, socialiseren en subjectiveren. Hoe curriculumvernaauwing tegengaan?

De haalbaarheidsstudie stelt dat een juridisch kader dat verbiedt dat data over
scholen publiek worden gemaakt, rankings niet zal tegenhouden. De Vlor en de
onderwijsverstrekkers dringen er sterk op aan om het eigenaarschap van de data
bij de scholen te leggen. Wat is de mening van de sprekers daarover? Is dat een
appellering aan de interne kwaliteitszorg?

Het onderwijsveld vraagt uitdrukkelijk om ervoor te zorgen dat de toetsen vol-
doende gedragen zijn. Ook de OESO pleit daar sterk voor, maar ook de vele mid-
delen, mensen en tijd die eraan gespendeerd worden, maken dat noodzakelijk.
Daartoe moeten de betrokkenheid en de participatie verhogen. Het parlement moet
kritische vragen stellen zodat de plannen goed en bruikbaar zijn voor het onder-
wijsveld. De spreker vraagt om overleg tussen de parlementscommissie en de Vlor.
Nu heeft ze het gevoel dat er op ramkoers gevaren wordt.

Het regeerakkoord heeft het expliciet over een decretale basis. Het parlement is
niet gemaakt om wat al in voege is te regulariseren. De operationalisering hoort
na de goedkeuring van het decreet te komen. Ze mist een duidelijke visienota met
doelen en fundamentele keuzes waar het highlevelforum zich over kan buigen.
Momenteel gaat het forum al in op concrete lijnen, over fragmenten, zoals toetsen
in het buitengewoon onderwijs en de gunning van een digitaal platform, terwijl er
zelfs nog geen decreet is. De OESO benadrukt het belang van een goede proces-
bewaking en een geregelde evaluatie. Wanneer en hoe kan er bijgestuurd worden?
In Vlaanderen zijn er al netoverschrijdende toetsen, bijvoorbeeld de OVSG-toets.
Hoe kan wat er al is en goed is, meegenomen worden?

Niet alle landen hebben een inspectie of eindtermen, waardoor de overheid allicht andere manieren van kwaliteitsborging zoekt. Hoe heeft dat een invloed op de manier waarop er gekeken wordt naar centrale toetsen? Hoe kijkt het forum naar de schoolevaluatie? Ziet het brood in het Nederlandse voorbeeld waarbij een school pas extra onderzoek of een begeleidingstraject krijgt na drie opeenvolgende negatieve jaarscores?

Metten is weten, dat klopt, maar een varken wordt niet vetter door het te wegen. Er is dus ook bijkomend onderzoek nodig naar de oorzaak van de dalende scores. Wat zijn de plannen van het steunpunt op dat vlak? Wat de professionele ruimte van de leerkracht en de werkdruk betreft, herhaalt Loes Vandromme wat ze al vaker zei, dat niet controle en wantrouwen, maar een appel aan de beroepstrots, het vakmanschap, de zeggenschap en de autonomie, eerder zelfs dan een hoger salaris, een goede leerkracht motiveren. Leerkrachten horen architecten te zijn van het onderwijs, niet louter uitvoerders.

Loes Vandromme informeert nog of er een decretale basis nodig is om de toetsen digitaal te maken. Wordt er bij de interpretatie van de resultaten rekening gehouden met eventuele ongewenste effecten van die digitale versies? Voorts vraagt ze tips over hoe de toetsen een toegevoegde waarde geven aan de scholen.

1.2. Koen Daniëls

Koen Daniëls herinnert eraan dat de daling van de onderwijskwaliteit nogmaals bleek zowel bij de Vlaamse peilingproeven als bij de internationale PISA. Gewoon voortdoen, is geen optie. De wijze woorden uit het regeerakkoord, voldoende besdiscuteerd door de regeringspartijen, vormen een goede basis voor verdere actie. Hij beaamt dat een varken niet vetter wordt door het te wegen, maar gelukkig zijn de kinderen geen varkens, moeten ze niet vetter worden, noch geslacht. Als pedagoog weet hij dat toetsen een belangrijk leereffect hebben, wat in veel fora vergeten wordt. Hij vraagt de parlementsleden zich te beperken tot vragen, en opinies, al dan niet ingefluisterd door belangengroepen, voor later te houden.

Toetsen moeten meerlagig testen, want als ze maar één aspect (eenpuntscesuur) toetsen, wordt teaching to the test wel en excelleren niet aangemoedigd. Uit de uiteenzettingen van wie met de Vlaamse toetsen bezig is, leidt hij af dat dat hier niet het doel is. Door hun adaptiviteit dagen ze de leerlingen uit. Een tweede verschil met de Nederlandse testen is dat de Vlaamse toetsen ook leerwinst meten. Toetsen vergen inderdaad tijd, energie en middelen. Het Vlaamse onderwijs krijgt 14 miljard euro belastinggeld en de parlementsleden zijn daar de behoeders van. Sinds ettelijke decennia gaan de scores erop achteruit, toch is er geen enkele school zwaar op de vingers getikt, noch gesloten wegens ondermaatse onderwijskwaliteit of gebrek aan leerwinst, wel kregen ze aandachtspunten voor eindtermen die niet gehaald zijn.

Een extra toets waar niets mee gedaan wordt, is verloren tijd. De Vlaamse toetsen zijn dat niet daar ze informatie op leerling-, school- en systeemniveau opleveren. In Nederland is dat niet het geval. De Vlaamse toetsen moeten ook valide en betrouwbaar zijn, zeker om de leerwinst te meten. Hoe zal het steunpunt daarvoor zorgen?

Bij heel wat aspecten van de plannen zijn er bedenkingen te maken, bijvoorbeeld bij de rankings, maar met alleen maar negativiteit wordt er geen vooruitgang geboekt. Slagen andere landen erin rankings uit de openbaarheid te houden? Zo ja, hoe? De spreker informeert of Vlaamse toetsen in itemontwikkeling gelijkaardig zijn als de huidige peilingproeven en parallelproeven in het secundair onderwijs. Kunnen die ervaringen gebruikt worden voor de ontwikkeling van de Vlaamse toetsen?

Vlaanderen zal blijven deelnemen aan internationale onderzoeken. Wat zijn de gelijkenissen en verschillen met de Vlaamse toetsen? Zullen er bij de toetsen ook open vragen zijn of enkel meerkeuze- en gesloten vragen? Leerkrachten en scholen willen weten waar ze staan. Hoe kunnen ze zichzelf zinvol vergelijken gezien de verschillen in school en methode? Hoe kan bepaald worden wat de oorzaak is van leerwinst: de leerling, de leerkracht, de methode? Het leerlingnummer maakt het mogelijk om leerlingen met bijvoorbeeld leerstoornissen of een verslag te identificeren. Vlaanderen laat leerlingen van het buitengewoon onderwijs (special needs) deelnemen aan PISA en TIMSS, in andere regio's is dat niet het geval. Wat een vertekend resultaat geeft. Zullen zij wel deelnemen aan de Vlaamse toetsen?

Hoe moet het mild summatieve karakter van de Vlaamse toetsen begrepen worden? Terloops meldt hij dat ook hij geen uitnodiging kreeg voor de persconferentie. Vlaamse toetsen zonder terugkoppeling naar de leerling, zijn verspilling, maar als eindexamen, zijn ze evenmin bedoeld. Ze zijn slechts een onderdeel van de evaluatie van de leerling. Wat gebeurt er met de resultaten als de leerling echt duidelijk een slechte dag heeft en de rest van het jaar goed presteert? Worden statistische uitersten uit de totale evaluatie van de leerling gehaald?

In het highlevelforum is zowat iedereen vertegenwoordigd, ook praktijkmensen. Zijn er daar nog blinde vlekken? Krijgt het parlement daar voldoende rechtstreekse feedback van? Het parlamentslid verwijst voorts naar een tweet van Jan Vanhoof waarin hij alle interpretaties van teaching to the test op een rijtje zet. Hoe kan het de onderwijskwaliteit bedreigen dat een leerkracht zijn lessen afstemt op de Vlaamse toetsen? Welke interpretatie van Jan Vanhoof is fout?

Zullen de groepen in het secundair onderwijs, met de verschillende basisopties in de eerste graad, de onderwijsvormen en de richtingen in de hogere graden voldoende groot zijn voor statistisch relevante conclusies? Blijven de OVSG-toetsen en de interdiocesane proeven bestaan? Vlaanderen blijft deelnemen aan PISA, TIMSS en PIRLS. Kunnen de Vlaamse toetsen evenwel gebruikt worden om data aan te leveren voor die internationale studies?

De spreker heeft zelf ooit onderzoek gedaan over schooldoorlichting als middel voor onderwijsverbetering. De schooldoorlichting is dan wel stressvol, doorgaans gaat een school erna snel terug naar de baseline, dus zonder al te veel bijschaving. Hoe zit dat in de nieuwe onderwijsdoorlichting van het onderwijskwaliteitskader? Hoe wordt leerkrachten en directies geleerd om te gaan met de data en de informatie, de zogenaamde datageletterdheid?

1.3. Roosmarijn Beckers

Dat er landen zijn met gestandaardiseerde centrale toetsen die het goed doen op PISA, zoals Estland, Polen, Portugal, Ierland, maar andere met dergelijke toetsen die het slecht doen, zoals Nederland en Frankrijk, doet *Roosmarijn Beckers* besluiten dat het niet de toetsen zelf zijn, maar wat ze teweegbrengen op de klasvloer dat het verschil maakt. Wat zijn naast de kwaliteit, de validiteit en de betrouwbaarheid van de toetsen de essentiële randvoorwaarden uit de best practices waar Vlaanderen vooralsnog geen aandacht voor heeft? Het OESO-rapport pleit voor voldoende draagvlak via het activeren van het engagement door duidelijke communicatie, maar ook voor het horen van sleutelfiguren waaronder de schooldirecteurs. Dat komt ook naar voren in het recente Vlor-advies. Valt er op dat vlak ook iets te leren van andere landen die centrale proeven hebben ingevoerd?

Data interpretern en er lessen uit trekken en die toepassen, is voor veel leerkrachten en directies, zeker in de lagere scholen, niet eenvoudig. Dat heeft te maken met een beperkte datageletterdheid van het personeel maar ook met de beperktere beleidsondersteuning. Hoe wordt de professionalisering aangepakt in

andere landen, bijvoorbeeld in Nederland voor de CITO-toetsen? Dat meten dat wat wordt gemeten, wijzigt (measurement changes the measured), kan zowel positief als negatief zijn. Hoe kan curriculumverschraling vermeden worden terwijl scholen nog heel veel vrijheid hebben in de omgang met de toetsen? Teaching to the test moet worden voorkomen. Daarnaast is er het risico van teaching to the rating, namelijk dat leraren hun lessen richten op leerlingen waarvoor er nog snel winst te boeken is, doorgaans de middengroepen, wat minder goed is voor de zwakkere of sterkere leerlingen. Hoe kan dat worden voorkomen, zowel op micro- als op macroniveau?

De Vlaamse toetsen zijn vooral formatieve testen en voor de leerlingen dus eerder low stake. Zullen ze dan wel voldoende gemotiveerd zijn en de testen serieus nemen? Hoe de motivatie en de prestatiebereidheid van de leerlingen verhogen en de toets toch formatief houden? Of is het beter de test summatief te maken? De scholen kunnen nu zelf beslissen om de uitkomst deels te laten meetellen in de evaluatie, wat zal leiden tot verschillen tussen scholen. In de haalbaarheidsstudie staat dat in het secundair onderwijs de afstand tussen de proef aan het einde van de eerste graad en het einde van de derde graad groot is. Leerlingen maken op die vier jaar een grote evolutie door. De eindtermen en de leerdoelen van de eerste en de derde graad zijn moeilijk met elkaar te vergelijken. Er is meer kans op leerlingenverschuiving. Moet er op termijn niet gedacht worden aan een proef op het einde van de tweede graad secundair onderwijs?

De proef op het einde van de derde graad heeft weinig formatieve waarde voor de leerling, maar zal voor hen meer summatief aanvoelen dan de drie eerdere proeven. Moet de voorbereiding dan anders aangepakt worden? Wat zal er gebeuren met de andere toetsvormen, de Vlaamse peilingproeven, de net- en koepelgebonden testen? Blijven die behouden of is integratie in de Vlaamse toetsen zinvol, met een standaarddeel en een specifiek deel per net of koepel?

1.4. Hannelore Goeman

Ook Vooruit is voorstander van een goede monitoring die gericht is op het verbeteren van de onderwijskwaliteit, aldus *Hannelore Goeman*. De GGG-proeven kunnen daarin een belangrijk instrument zijn, op voorwaarde dat ze deel uitmaken van een bredere aanpak van kwaliteit. Scholen moeten in staat gesteld worden om lessen te trekken uit de resultaten, om hun methodieken bij te schaven. Vandaag is er nog heel wat onduidelijk: hoe zullen de toetsen eruitzien, wat zullen de gevolgen zijn enzovoort. Dat zorgt voor ongerustheid. Het is goed dat het geïnformeerde debat vandaag start. Het parlement moet inderdaad vooraf het wetgevende kader scheppen. De voorafnames op het debat, ook in de pers, betreurt zij.

Wat er gemeten wordt, heeft verregaande consequenties voor het toetsdesign. Volgens de beleidsnota zijn er drie doelen, maar is het überhaupt mogelijk om leerwinst van de leerling, kwaliteitswinst van de school en van het systeem met één toets op een betrouwbare en gevalideerde manier te meten? Hoe kan de leerwinst tussen het tweede en zesde jaar secundair onderwijs betrouwbaar gemeten worden gezien de diverse richtingen en niveaus en de diverse wegen die leerlingen inslaan? De leerwinst op leerlingniveau wordt gecorrigeerd op leerlingkenmerken. Wordt er met de omgevingselementen en kenmerken van de populatie rekening gehouden bij de feedback aan de scholen?

Van de experts is een deel gewonnen voor een brede, grofmazige toets, waarbij enkele eindtermen en onderdelen van het vak getoetst worden. Een ander deel is voorstander van specifieke en gedetailleerde toetsen, waarbij niet elke school alle toetsen jaarlijks afneemt. De laatste werkwijze verhindert dat voortgang gemonitord wordt. Sowieso is het belangrijk dat alle toetsen, zowel in het lager als in het

secundair onderwijs, dezelfde methodiek en design hanteren. Hoe zal het steunpunt daarvoor zorgen?

Ook Hannelore Goeman vraagt nog eens te herhalen wat het verschil precies is tussen peilingproeven en centrale toetsen. Waarom is er niet voor gekozen om de bestaande peilingproeven voor Nederlands en wiskunde te veralgemenen? Wijzigt de methodiek? Lopen de OVSG- en interdiocesane toetsen gewoon door? Of wordt eraan gedacht ze complementair te maken aan de GGG-toetsen? Daarvoor is overleg met de onderwijspartners cruciaal. Ook dit parlementslid is benieuwd hoe dat overleg loopt. Hoe zit het met het draagvlak op dit moment? Wat zijn de reacties van de grote onderwijsverstrekkers?

Ook zij merkt heel wat verwarring over de term semisummatief. In het regeerakkoord staat dat scholen die niet goed genoeg scoren een zelf te kiezen maar wel verplicht begeleidingstraject krijgen. De Vlor ziet dat als een rechtstreeks gevolg, als high stake, wat de intrinsieke motivatie van onderwijsverstrekkers zal aantasten en het risico inhoudt van onder meer ongewenste vormen van teaching to the test. Volgens de sprekers zijn de toetsresultaten veeleer een onderdeel van een bredere kwaliteitscontrole door de inspectie. Zal de inspectie dan scholen die enkele keren slechter scoren prioritair doorlichten? Is er al beslist dat de inspectie de resultaten op schoolniveau krijgt? Voor leerlingen is een toets een momentopname. Leerkrachten kunnen het resultaat niet naast zich neerleggen, maar het mag zeker niet bepalend zijn voor oriëntering of attestering.

Omdat er nog fundamentele keuzes moeten worden gemaakt en het ondersteuningsbeleid nog moet worden opgezet, maar ook omdat scholen drukdoend zijn met de invoering van de eindtermen en de Digisprong, lijken de invoerdata haar ambitieus. Zijn ze volgens de sprekers wel haalbaar? Wat zijn de plannen met de ondersteuning door de administratie?

De GGG-toetsen die nu ontwikkeld worden voor het basisonderwijs, zijn allicht gebaseerd op de oude eindtermen. Dat betekent dat het werk moet worden overgedaan zodra er nieuwe zijn. Dat is niet alleen weggesmeten geld, het maakt het moeilijk om resultaten voor en na de nieuwe eindtermen te vergelijken. De spreker begrijpt dat de meerderheid de toetsen zo snel mogelijk wil lanceren, maar kwaliteitsvol werk primeert. Tot slot vraagt ze of het mogelijk is, gezien de wetgeving op de openbaarheid van bestuur, te garanderen dat de schoolresultaten niet kunnen worden gebruikt voor publieke rankings. Zijn er goede voorbeelden uit het buitenland? Hoe waterdicht zijn de garanties?

1.5. Johan Danen

Johan Danen merkt dat Vlaanderen een van de laatste is om dergelijke proeven in te voeren, maar ook dat bepaalde landen ervan terugkomen, onder andere Nederland. Komt dat omdat de doelstellingen niet gehaald zijn of zijn er andere redenen? De minister en de regering deden al uitspraken over wat de toetsen moeten zijn, nog voor het maatschappelijke debat van start ging. De minister wil toetsen om niet afhankelijk te zijn van buitenlandse of internationale rankings. Wat wil hij er precies mee bereiken? In zijn antwoord op het Vlor-advies vernoemt de minister het versterken van en de feedback aan scholen niet. Is dat een bewuste vergetelheid? Of zijn de doelstellingen voor de scholen nog niet bepaald? Kennen de ontwikkelaars de precieze beleidsdoelstellingen?

De spreker heeft begrepen dat er te weinig uit de Vlaamse peilingproeven gehaald is. De bestaande proeven van de netten, koepels, peilingproeven, de PIRLS, PISA en TIMSS-proeven, leggen de pijnpunten van het Vlaamse onderwijs bloot: achteruitgang in wiskunde en Nederlands. Waarom daar niet meteen iets aan doen? Is het echt nodig om veel geld, tijd, manuren en onrust in een systeem te stoppen

dat enkel zal bevestigen wat al bekend is? Ook hij maakt zich zorgen over teaching to the test, zeker omdat de geteste stof beperkt is, tenzij de vragen heel rijk zijn, zoals blijkbaar de bedoeling is. Opdat ze aan de slag kunnen met rijke toetsresultaten, moet er in de school een cultuur van reflectie en omgang met feedback en kritiek heersen en een goede interne kwaliteitscontrole zijn. Hoe ervoor zorgen dat scholen op een goede manier aan de slag gaan met de resultaten, ook als die minder goed zijn?

Heel wat andere landen en regio's worstelen met onderwijskwaliteit. Centrale toetsen hebben niet overal de neerwaartse trend kunnen keren. Wat is er daarvoor wel meer nodig?

1.6. Jan Laeremans

Ook *Jan Laeremans* heeft vragen over hoe, gezien de wet op de openbaarheid van bestuur, rankings geheim kunnen worden gehouden. Er wordt dan wel vaak op de negatieve gevolgen gewezen, maar openbaarheid, ook van rankings, zorgt ook voor transparantie, dus misschien wel voor een vorm van responsabilisering van de scholen. Onderwijseconoom Kristof De Witte maakt niet zo'n probleem van schoolrankings, op voorwaarde dat ze goed geduid worden. Hoe zien de sprekers die suggestie, zeker daar ze in Nederland al gebeuren?

2. Antwoorden van de sprekers

Claire Shewbridge raadt de parlementsleden de lectuur aan van het volledige OESO-rapport, waarin heel wat internationale voorbeelden staan, hoewel het niet bedoeld is als benchmark. Het spitst zich immers toe op de vervolgstappen in de Vlaamse context. Uit de enquête blijkt dat geen enkele centrale actor gekant is tegen de toetsen, hoewel er heel wat bezorgdheden zijn. Scholieren hebben een divers standpunt over gestandaardiseerde testen: sommige kijken uit naar bijkomende feedback, andere krijgen er stress van. Allemaal hebben ze evenwel opmerkingen bij de huidige feedback. Ze hebben vragen over hoe hun eigen prestaties beoordeeld worden in vergelijking met andere scholieren, maar ook bij verschillen tussen beoordelingen van leerkrachten, scholen en netten. Het gebrek aan objectiviteit is een constante bezorgdheid. Al bij al vinden ze centrale toetsen betrouwbaarder en objectiever. Heel wat onderwijsstakeholders verklaren te hopen dat de centrale toetsen hen zouden helpen bij het professionaliseren van zowel toetsings- als feedbackvaardigheden.

Ervaringen uit andere landen wijzen uit dat leerkrachten het best van bij de ontwikkelingsfase betrokken worden bij de centrale toetsen. Daarbij komt nog dat inzichten evolueren, zoals de ontwikkelingen in Denemarken tussen 2012 en 2018 bewijzen. Leerkrachten, ook in andere landen, ervaren in eerste instantie centrale toetsen als betuttelend, als een politieke beïnvloeding zelfs. Wetenschappelijke studies wijzen evenwel uit dat ze een goede werkwijze zijn. De resultaten van de toetsen zijn slechts een basis om het gesprek op gang te brengen. Het is inderdaad een momentopname. Daarbij komt nog dat ze slechts over Nederlands en wiskunde gaan, wat er slechts een van de informatiebronnen over het functioneren van de leerling van maakt.

De spreker refereert voorts aan een studie van haar hand over de kwaliteit van de toetsen. Wetenschappelijke ondersteuning ervan is cruciaal. Het onderwijsveld heeft hoge verwachtingen over de statistische kwaliteit en de betrouwbaarheid van gestandaardiseerde of centrale toetsen. Ze moeten uiteraard afgestemd zijn op de onderwijsdoelen, in Vlaanderen de eindtermen. Het toetsingsinstrument op zich volstaat niet om de onderwijskwaliteit op te krikken. Er moet ook iets gedaan worden met de resultaten, wat bepaalde vaardigheden maar ook een bereidheid vergt. Wat de accountability betreft, kan een school zich vrijwillig laten begeleiden door

de pedagogische begeleidingsdienst. Ook de onderwijsinspectie kan de kwaliteit helpen verbeteren, maar de controlecycli zijn lang en ze is slechts beperkt in de school aanwezig. Vanuit internationaal perspectief is het Vlaamse Referentiekader voor Onderwijskwaliteit een goede manier om de vele aspecten van onderwijskwaliteit gemeenschappelijk te maken voor het volledige onderwijsveld.

Een gemeenschappelijke en professionele evaluatiecultuur is inderdaad belangrijk. Vlaanderen heeft heel wat mogelijkheden om de scholen op een goede manier te begeleiden en te ondersteunen bij datageletterdheid, maar ook bij de interpretatie van de juiste draagwijdte van de toetsen, één element van evaluatie. De Nederlandse spreker zegt terecht dat toetsen proportioneel moeten zijn, stipt Claire Shewbridge aan.

Ze roept de overheid op te blijven luisteren naar alle onderwijsactoren en zeker de scholieren niet te vergeten. Centrale toetsen zijn een ideale kans om objectiviteit te verzekeren en het schoolniveau van evaluaties te overstijgen. De ervaringen in andere landen hebben bewezen dat centrale toetsen scholen actie doen ondernemen, onder meer voor de professionalisering van hun lerarenteam. Centrale toetsen zijn een stimulans om het debat aan te gaan, beter samen te werken. Contextualiseren en rekening houden met de samenstelling van de leerlingpopulatie zijn belangrijk om leerwinst te meten, waarmee dan aangetoond kan worden dat de inspanningen van het personeel vruchten afwerpen.

Politici hebben verschillende meningen, leggen andere klemtonen. Voor de spreker is het evenwel duidelijk dat de data per school niet openbaar mogen worden gemaakt. Ook Denemarken kent openbaarheid van bestuur, maar toch worden de resultaten van de scholen er niet bekendgemaakt. Het is voldoende om in de wet te bepalen dat de data dienen voor vergelijkingen en wetenschappelijk onderzoek en dat enkel de school en eventueel de inrichtende macht ze zullen krijgen. Een databank met een beschermde toegang is ICT-gewijs niet moeilijk.

Van een wijzigende mentaliteit is Australië een goed voorbeeld. Transparantie en verantwoordingsplicht zijn ook daar heel belangrijk. Het land had eerst gekozen voor een openbaar platform, maar is daarvan teruggekomen en vergelijkt nu enkel de scholen met hun eerdere prestaties. Zelfs met sociaaleconomische correcties is er geen goede vergelijking tussen scholen mogelijk. De spreker verwacht dat centrale toetsing niet alleen objectiever en betrouwbaarder is, maar op langere termijn ook tot een efficiënter gebruik van de middelen zal leiden omdat het fragmentatie beëindigt. De verschillende Vlaamse onderwijsnetten beschikken al over goede toetsen om hun leerplannen te evalueren, maar de centrale toetsen kunnen aanvullend zijn. Zo is het bijvoorbeeld mogelijk om met netgebonden toetsen andere vakken dan Nederlands en wiskunde te testen. De Vlaamse peilingen kunnen misschien herwerkt worden tot centrale toetsen. Het grote verschil tussen peilingen en centrale proeven is het aantal leerlingen dat eraan deelneemt. De peiling werkt met een steekproef, waardoor het mogelijk is nieuwe procedures uit te proberen. Ze kunnen dus ook fungeren als proefprojecten.

Het is zeker nuttig trends en evoluties te monitoren zoals ook met PISA gebeurt. Centrale proeven zijn geen statisch gegeven. Het is zaak om voortdurend open te staan voor verbeteringen, in het belang van het leerproces, van het voorbereiden van leerlingen op de immer veranderende realiteit. Er zullen echter altijd keuzes en compromissen moeten worden gemaakt. Tot slot raadt ze de parlementsleden aan de Deense websites over dit onderwerp te consulteren. De openheid is daar heel groot, het doel duidelijk. De standpunten van de diverse stakeholders, waaronder de scholieren, staan er vermeld.

Voor de Vlaamse administratie is het draagvlak een absolute prioriteit, aldus *Jeroen Backs*. Het publieke debat en de hoorzitting juicht hij toe. Het communicatieplan

past in de initiatieven om draagvlak te creëren. Ondertussen heeft de administratie een vijftigtal sessies gehouden over de Vlaamse toetsen, een website en informatie-uurtjes voor leraren en schoolleiders. Ook het persmoment paste in het communicatieplan. Misconcepties over de toetsen voeden immers allerlei indianenverhalen en creëren onnodig ongerustheid. Het is zaak om de juiste informatie te verspreiden, onder meer aan journalisten die over de hoorzitting van vandaag berichten.

In het highlevelforum zijn alle onderwijsverstrekkers, de vakbonden, de Vlaamse Scholierenkoepel en het steunpunt vertegenwoordigd. De rechtstreekse dialoog met de schoolleiders is evenwel nog onvoldoende uitgebouwd. Ook het OESO-rapport kaartte dat aan. De administratie denkt na over hoe daar een tandje bij te steken. Het highlevelforum is vertrokken van enkele hangende kwesties uit de haalbaarheidsstudie. Als onderdeel van de beleidsvoorbereiding wordt het draagvlak van enkele scenario's getest, bijvoorbeeld of leerlingen uit het buitengewoon onderwijs zullen deelnemen, vooraleer de administratie de minister een voorstel voor een ontwerp van decreet doet. Het is uiteraard aan de wetgever om de finale beslissing te nemen.

Het leerlingperspectief blijft inderdaad soms wat onderbelicht. Het risico op juridisering is reëel. Het is zaak om de ouders goed te informeren dat de resultaten slechts een onderdeel van de evaluatie zijn. Dat zit in de planning vanaf 2023. Het doel is de impact voor de leerling low stakes te maken. De klassenraad zal de resultaten van de Vlaamse toets als onderdeel van de evaluatie kunnen gebruiken. Het is niet zo dat, zoals in Nederland, de toegang tot vervolgoopleidingen rechtstreeks afhankelijk worden van de resultaten. Wat evenwel ook te vermijden is, zijn no-staketoetsen. Uit het PISA-onderzoek blijkt immers dat testmotivatie wel degelijk de resultaten beïnvloedt.

De Franse Gemeenschap heeft sinds 2006 centrale toetsen. Het is er bij decreet verboden de resultaten van de scholen openbaar te maken. Tot nog toe zijn er geen rankings publiek gemaakt. Anderzijds is transparantie inderdaad belangrijk. Het zal moeilijk zijn beide aspecten in balans te houden. Rankings zoals in het Verenigd Koninkrijk jaarlijks gepubliceerd worden, dus op basis van ruwe resultaten zonder rekening te houden met standaardfouten, met achtergrondkenmerken van de school en de leerlingen, geven een fout of simplistisch beeld, wat wetenschappelijk bewezen is schadelijk te zijn. Portugal maakt echter schoolresultaten bekend zonder rankings, waarbij de school data kan contextualiseren en toelichten. Het klopt dat er nog verder onderzoek naar nodig is.

De Vlaamse administratie apprecieert het werk van OVSG en Katholiek Onderwijs Vlaanderen om eigen toetsen te ontwikkelen. Het zijn belangrijke instrumenten voor de kwaliteitszorg van het beleidsdomein Onderwijs en Vorming. De OVSG-toetsen en de interdiocesane proeven beantwoorden evenwel niet aan de criteria van gestandaardiseerd, genormeerd en gevalideerd, net- en koepeloverschrijdend. Dat is de reden waarom de Vlaamse toetsen er niet op gebaseerd zijn. Het beleidsdomein Onderwijs en Vorming voerde al gesprekken over de toekomst van de netgebonden toetsen, met wisselende graad van vlotheid. Het is aan de netten om erover te beslissen, maar de administratie ziet complementariteit mogelijk als ze over andere leergebieden gaan dan de Vlaamse toetsen.

Vertrekken van de Vlaamse peilingproeven was een optie. Dat er niet voor gekozen is, heeft te maken met timing maar ook met verschillen in opzet. Het nieuwe universitaire steunpunt neemt evenwel de expertise van de afgelopen twintig jaar mee. Rianne Janssen blijft meebouwen aan het onderdeel psychometrie.

Uiteraard zullen de data die de toetsen opleveren, niet de oorzaken van de problemen blootleggen. Investeren in verklarend onderzoek zal een noodzakelijke voorwaarde zijn om van de toetsen een kwaliteitsinstrument te maken.

Dat werken met een cesuur, een minimumlat, zal zorgen voor lagere kwaliteit en een plafondeffect, is een terechte bezorgdheid. Het beleidsdomein denkt eerder aan een rapportagesysteem vergelijkbaar met dat over PISA, met verschillende vaardigheidsniveaus. De items in een adaptieve toets gaan verder dan het niveau van de eindterm, waardoor de verschillende vaardigheidsniveaus gemeten worden. Zo kan bijvoorbeeld bepaald worden hoeveel leerlingen excelleren en wordt een plafondeffect tegengegaan.

De Vlaamse toetsen geven een beeld van het bereik van het Vlaamse onderwijscurriculum maar zijn geen internationale vergelijkingsbasis, hoewel het wel mogelijk is bepaalde peilingsitems op de internationale schaal van PIRLS of TIMSS te zetten. Omdat internationale onderzoeken andere referentiekaders hanteren, geven ze complementaire informatie over het bereik van de Vlaamse onderwijsdoelen.

Denemarken ontwikkelde hoogkwalitatieve toetsen, maar de data waren te complex voor leerkrachten om ze te gebruiken in de klas. Goede feedbackrapporten en datageletterdheid van leerkrachten zullen cruciaal zijn. Vanaf 2023 wordt datageletterdheid een belangrijk thema voor de Vlaamse pedagogische begeleidingsdiensten, maar ook voor het steunpunt.

Centrale toetsen leiden niet automatisch tot kwaliteit. Cruciaal is dat elke betrokkene, van beleidsmaker tot uitvoerder, bereid is feedback te aanvaarden en zijn rol in vraag te stellen. Kritische zelfreflectie dus, wat een cultuuromslag vergt, waarin de pedagogische begeleidingsdiensten een belangrijke rol zullen spelen. Dat de data gebruikt worden in de interne kwaliteitszorg is evenzeer cruciaal.

De timing is volgens Jeroen Backs inderdaad ambitieus, maar toch haalbaar. Dat is mede omdat de toetsontwikkeling snel na de start van de regeerperiode gestart is, wat ook wrevel veroorzaakt heeft omdat er geen debat aan voorafging. De doorlooptijd van toetsontwikkeling is evenwel twee tot drie jaar. Om de eerste toetsen van start te laten gaan voor het einde van de regeerperiode, was snel ageren noodzakelijk. Over voldoende ondersteuning en personeelskader bij de administratie voor de periode na 2024 moet zeker worden nagedacht.

Initieel zouden er tegen 2024 nieuwe eindtermen basisonderwijs zijn. Dat nu de oude eindtermen gebruikt worden, houdt inderdaad het risico in dat de toets grondig moet worden herwerkt en dat trends en de voortgang meten moeilijk wordt. Eindtermen wiskunde en Nederlands worden allicht niet helemaal omgegooid, de ankeritems blijven mogelijk gelijk. Het is aan de toetsontwikkelaars om daarmee om te gaan.

Hij zal geen politieke afweging maken van de wenselijkheid van centrale toetsen, maar die zullen in elk geval relevante informatie opleveren voor evidence informed onderwijs. Het grote verschil met de Vlaamse peilingen en het internationaal vergelijkend onderzoek is dat alle scholen en alle leerlingen in Vlaanderen deelnemen. De toetsen zullen ook rijk zijn, en dus rijkere informatie opleveren. Persoonlijk is de spreker overtuigd van het nut van centrale toetsen, alsook dat de kosten en de inspanningen opwegen tegen het versterkende karakter ervan voor het Vlaamse onderwijs, hoewel heel wat afhangt van de randvoorwaarden. Dat de toetsen digitaal zijn, is nodig gezien de omvang van de onderzochte populatie: 560.000 toetsen per jaar op kruissnelheid. Het risico op een testmode-effect door digitale toetsen is er enkel als er zowel toetsen op papier als digitaal zijn.

Bieke De Fraine bevestigt dat de resultaten van de leerling kunnen worden geaggregeerd naar klas-, school- en Vlaams niveau, maar ook kunnen worden gebruikt om leerwinst te berekenen. De Vlaamse toetsen zijn evenwel niet bedoeld voor schoolkeuze en evenmin voor studieoriëntering. De autonomie van de klassenraden blijft overeind. De klassenraad kan rekening houden met de toetsresultaten, maar die mogen geen bindend element zijn voor studieoriëntatie.

Het highlevelforum heeft afgesproken om de termen mild of semisummatief niet meer te gebruiken wegens verwarrend, maar wel de Engelstalige term low stakes, hoewel een goede Nederlandse term welkom is. Het is niet zo dat de toets no stake is, noch voor de leerling, noch voor de school. De onderwijsinspectie krijgt jaarlijks de toetsresultaten van alle scholen. Hoe ze er precies mee zal omgaan, moet nog uitgetekend worden, maar bij de voorbereiding van de doorlichting neemt de onderwijsinspectie ook vandaag al een heleboel gegevens van de school mee in rekening: over zittenblijven, over attestering, over spijbelen, over doorstroming, over verloop van leerlingen en leerkrachten. De Vlaamse toetsen worden gewoon een extra element in het Referentiekader van Onderwijskwaliteit. Lage resultaten zullen dus niet onmiddellijk leiden tot een begeleidingstraject. De onderwijsinspectie zal wel nagaan wat er precies aan de hand is en pas dan bepalen of de school een begeleidingstraject nodig heeft.

De brede vorming van de leerlingen blijft overeind, een curriculumvernauwing is er niet. Nederlands en wiskunde zijn essentiële basisvaardigheden. Tal van studies wijzen uit dat grote groepen leerlingen basisgeletterdheid en rekenvaardigheid missen. Dat ondermijnt de toekomstkansen van die kinderen. De onderwijsinspectie zal in haar doorlichting altijd ruimer kijken dan alleen naar Nederlands en wiskunde.

Johan van Braak zegt dat Vlaanderen niet kiest voor highstaketesting zoals Nederland. De toetsresultaten worden immers niet gebruikt voor studieoriëntering of attestering, noch worden de leraren afgerekend op de resultaten van hun leerlingen. Het steunpunt erkent dat de centrale toetsen tot een hogere druk zullen leiden. Er wordt immers een nieuw element van kwaliteitsbewaking toegevoegd aan een bestaand kader. Veel onderwijsactoren zijn daar bezorgd over. Het netwerk is er zich van bewust dat de scholen voor een uitdagende opdracht staan, en dat net in een al heel uitdagend tijdsgewricht. Het steunpunt houdt daar zo veel mogelijk rekening mee.

Op metaniveau staat de ontwikkelingsgerichtheid, een van de pijlers van de visie van het steunpunt, niet de verantwoordingsgerichtheid waarin leraren en scholen op een bestraffende manier benaderd worden. De acht werkdomeincoördinatoren en een team van meer dan veertig universitaire promotoren staan klaar om het netwerk in zijn opdracht te begeleiden. Kortom, de academische wereld zet zijn schouders mee onder het project. Het steunpunt wil vooreerst bijdragen aan een versterking van de onderwijskwaliteit. Er zijn niet alleen de dalende trends, maar internationale studies wijzen ook uit dat de groep kwetsbare leerlingen steeds groter wordt. Informatie over hoe leerlingen het doen, voegt een belangrijk element toe aan het proces van kwaliteitszorg.

De naam van het steunpunt is in een democratische verkiezing bepaald. De toetsen zijn evenwel geen doel op zich, maar een middel om andere doelen te bereiken. Het werkelijke potentieel ligt in de manier waarop de scholen met de toetsdata aan de slag gaan: de resultaten analyseren, interpreteren en op basis daarvan hun werking bijstellen.

Het zit in het DNA van het steunpunt om voor draagvlak te zorgen. De denktanks selecteren de eindtermen die getoetst worden. Voor Nederlands maakt het regeerakkoord al een selectie: begrijpend lezen, grammatica en schrijven. Voor wiskunde

moeten er nog keuzes worden gemaakt: de volle breedte, enkel specifieke elementen of een combinatie van beide. De voor- en nadelen van breedtemeting versus dieptemeting worden nu tegen elkaar afgewogen, waarna er een scenario zal worden voorgelegd aan de stuurgroep. In die groep zijn de belangrijke onderwijsspelers vertegenwoordigd.

De focusgroepen voor redelijke aanpassingen en de gebruikersgroepen over feedback, zijn een andere manier om draagvlak te creëren. Het steunpunt is daarnaast ook gestart met de ontwikkeling van een wetenschappelijke opvolging van dat draagvlak. Binnenkort wordt er ook een survey uitgestuurd naar het bredere onderwijsveld, die daar specifiek op ingaat. Een ander belangrijk punt is scholen ondersteunen om met de data aan de slag te gaan. Het steunpunt zal daar een traject voor uitwerken, in overleg met het onderwijsveld, niet het minst met de pedagogische begeleidingsdiensten. Zo komen er onder meer opleidingen voor informatiecoaches, die leren omgaan met de bredere informatie die scholen krijgen om hun evaluatiebeleid in beeld te brengen. Het steunpunt erkent de complexiteit van het beleidsproject en het belang van het draagvlak, maar vertaalt dat ook in zijn werkwijze met een sterke betrokkenheid en samenwerking. Die is een belangrijke voorwaarde om samen op korte termijn tot succes te komen.

Rianne Janssen herinnert eraan dat ze vanuit STEP slechts een van de perspectieven gaf. Bij het beantwoorden van de vragen focust ze zich op de beperkingen en de impact van de centrale toetsen. Met enkel centrale toetsen kan nooit de volledige kwaliteit van het onderwijs en de scholen in kaart worden gebracht. De Vlaamse peilingen wijzen bijvoorbeeld uit dat er op bepaalde onderdelen van wiskunde weinig leerlingen goed scoren. De grootte van de getoetste populatie, impliceert dat de toetsen digitaal en de vragen zo veel mogelijk automatisch scorebaar moeten zijn. De ontwikkelaars zullen erover waken dat de digitalisering een zo klein mogelijke impact heeft op validiteit en geen vernauwing met zich meebrengt met betrekking tot de getoetste verwerkingsniveaus. Uit onderzoek blijkt alvast dat afname op computer of papier geen grote effecten heeft. Adaptief toetsen, is een streefdoel van het steunpunt centrale toetsen, maar vergt een grote itembank, een grote set van gekalibreerde vragen. Alle vaardigheidsniveaus omvatten, is een ander doel. Het steunpunt zal er alles aan doen om die doelen meteen te halen, maar allicht zal de kwaliteit mettertijd toenemen.

De tegenstelling tussen analytisch en globaal toetsen leeft inderdaad. De toetsduur is beperkt en het uiteindelijke design is sowieso een compromis. Misschien zal niet elke leerling op alle afnamemomenten over het hele curriculum getoetst worden. Het is aan de stuurgroep om de voorstellen van het steunpunt daarover te beoordelen. Als de meting betrouwbaar is op het laagste niveau, namelijk dat van de leerling, zal ze ook zo zijn voor de hogere niveaus van school en systeem.

De ervaringen van STEP en andere onderzoekers wijzen uit dat het niet zeker is dat leerwinst zal kunnen worden gemeten. Voor het vierde leerjaar zijn er geen onderwijsdoelen, dus kan er geen criteriumgerichte feedback aan de scholen gegeven worden. De meting in het vierde leerjaar kan belangrijk zijn omdat de toetsen de voormeting zijn voor onder andere de prestaties in het zesde leerjaar. Daarbij komt nog dat de prestaties in het vierde leerjaar al beïnvloed zijn door de school. Voor de spreker is de leerwinstmeting van leerlingen maar ook van scholen net een belangrijke toevoeging van de centrale toetsen.

Dat bij centrale toetsen een toetseffect ook een leereffect heeft, is voor de spreker niet helemaal zeker. Bij formatieve toetsen is dat een doel, bij toetsen aan het einde van het curriculum veel minder. Ze benadrukt dat het aangewezen is als er leerlingfeedback aan leerlingen en ouders wordt gegeven, die dan via de scholen wordt overgemaakt.

Ze heeft retrospectief inderdaad wel wat bedenkingen of alles uit de Vlaamse peilingproeven is gehaald, of ze wel voldoende gewogen hebben op de onderwijspraktijk. Misschien kwamen niet alle ministers graag naar buiten met peilingsresultaten bereikt onder andere ministers. Sowieso hebben de peilingen bijgedragen tot het insijpelen van de alarmerende berichten over het onderwijsniveau. De net- en koepelgebonden toetsen geven geen informatie over de daling in de prestaties omdat ze geen constantie in de meetschalen hebben. Ze bieden aanvullende informatie, zonder dat er criteriumgerichte uitspraken en trendanalyses mee mogelijk zijn.

Metten is weten, maar het weten lijkt soms vlug vergeten. Het is belangrijk dat de resultaten effecten hebben, niet alleen op de scholen maar ook op het onderwijsstelsel. Een laag resultaat in een Vlaamse peilingproef is soms beter in een volgende peiling. Pessimisme is niet nodig, maar centrale toetsen zijn geen gouden koetsen waarmee we ineens door een beter onderwijslandschap rijden. Daarvoor dient er heel wat werk verzet te worden. Ook de verwachtingen dat toetsen het leerproces van leerlingen zullen ondersteunen, moeten enigszins getemperd worden. Het is in ieder geval zo dat met de peilingen de verschillende vakgebieden ruimer in kaart worden gebracht. De academici engageren zich om het proces zo goed mogelijk te ondersteunen.

Datageletterdheid omvat niet alleen het lezen van de grafieken maar ook het duiden van de resultaten, er betekenis aan geven. Het bestaande onderzoek daarover kan een basis vormen voor maatregelen.

Hans Teegelbeckers zegt dat, ondanks alle jaren toetservaring, de aanscherping van de diverse toetsingskaders en de talrijke raden die toezien op de kwaliteit van de bestuurder en de scholen, de kwaliteit in Nederland blijft achteruitgaan. Onderwijs draait niet enkel om kwalificatie, maar ook om socialisatie en personificatie, identiteitsontwikkeling. Nederland zoekt nu naar een goede balans. De vraag is hoe toetsen daarin passen.

De Nederlandse ministerraad heeft beslist de toetsen bij de kleuters vanaf 2021 af te schaffen. Eigenlijk was het al de bedoeling om ze te schrappen in 2013. Volgens experts sluit de toets niet aan bij de ontwikkeling van kinderen tussen 4 en 6 jaar. Het is een momentopname en de uitslag zegt niets over wat de kleuter kan. Experts pleiten voor het observeren en het bijhouden van observatielijsten.

In Nederland zijn er op elke school interne begeleiders, die samen met de leerkrachten de resultaten van alle toetsen analyseren en de leerkrachten begeleiden bij eventuele remediëring en handelingsplannen. Die begeleiders volgen vaak een specifieke opleiding en houden onderlinge interviews. Ook worden er in veel scholen coördinatoren aangesteld voor bijvoorbeeld taal, lezen of rekenen, afhankelijk van de speerpunten van de school. Vlaanderen heeft vast ook wel experts die een dergelijke rol kunnen spelen.

Teaching to the test kan leiden tot kansenongelijkheid. In Nederland wordt het advies van de leerkracht naast het resultaat van de centrale toets gelegd. De leerkracht kan zelf het voordeel van de twijfel geven: als de toets wat minder goed is, kan hij zijn oorspronkelijk hoger advies aanhouden op basis van de schoolcarrière van het kind. Ouders die vinden dat hun kind te laag geadviseerd is door de leerkracht, gaan daar wel eens tegen in verweer, wat soms tot vervelende juridische procedures leidt. Ouders sturen hun kind ook wel eens naar privé bureautjes om te oefenen. Dat is enkel weggelegd voor ouders van een hogere sociale klasse. Samen met een lagere inschatting door vooroordelen leidt dit tot kansenongelijkheid.

In heel wat landen hebben centrale toetsen de kwaliteit niet doen stijgen. De leraar is de sleutel. Voor hem zijn goede leerstof, een goede begeleiding en kwaliteitszorg en een stimulerende inspectie wezenlijk. Het is maar zeer de vraag of tijd inves-

teren in systemen als leerlingvolgsystemen de kerntaak van de leraar, namelijk kinderen dingen leren, ten goede komt. Natuurlijk is meten goed, maar het mag geen hoofdbezigheid worden. Dat leidt tot werkuitval van leerkrachten. Dat het huidige Nederlandse regeerakkoord aankondigt dat de salarissen van primair en secundair onderwijs gelijkgetrokken worden, zal de kwaliteit niet onmiddellijk doen toenemen. Allicht zal het enkel, en pas over vijf jaar, een deel van het lerarentekort oplossen. De spreker acht een andere fundamentele wijziging nodig: mede door het lerarentekort moeten er misschien eerder minder lessen gegeven worden, met een grotere focus op de basiszaken, waardoor de leraar tijd overhoudt voor professionele gesprekken, goede lesvoorbereiding en analyse. Hij begrijpt dat de Vlaamse centraaltoetsinfrastructuur er al is, maar dat er nog heel wat moet gebeuren om voor draagvlak te zorgen. Het is zaak om te vermijden dat de centrale toetsen verworden tot het vliegveld van Lelystad dat volledig aangelegd is, maar nooit in gebruik genomen is door protest van omwonenden.

Karolien GROSEMANS,
voorzitter

Koen DANIELS
Loes VANDROMME,
verslaggevers

Gebruikte afkortingen

CITO	Centraal Instituut voor Toets Ontwikkeling
CLB	centrum voor leerlingenbegeleiding
GGG	gevalideerde, gestandaardiseerde en genormeerde proeven
KU Leuven	Katholieke Universiteit Leuven
LiSO	Loopbanen in het Secundair Onderwijs
OESO	Organisatie voor Economische Samenwerking en Ontwikkeling
OVSG	Onderwijsvereniging van Steden en Gemeenten
PAV	project algemene vakken
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment (van de OESO)
SiBO	Schoolloopbanen in het basisonderwijs
STEP	Steunpunt Toetsontwikkeling en Peilingen
TIMSS	Trends in International Mathematics and Science Study
UAntwerpen	Universiteit Antwerpen
UGent	Universiteit Gent
Vlor	Vlaamse Onderwijsraad
VOS/ABB	vereniging voor openbare en algemeen toegankelijke scholen
3G	gevalideerde, gestandaardiseerde en genormeerde proeven