

1565 (2022-2023) – Nr. 1
ingediend op 31 januari 2023 (2022-2023)

Verslag van de gedachtewisseling

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

over de ontwikkeling van de Vlaamse toetsen
door het Steunpunt Centrale Toetsen in 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 het Steunpunt Centrale Toetsen in Onderwijs	4
1.	Opdracht van het Steunpunt	4
2.	Wat houden de centrale toetsen in?	5
3.	Welke analyses worden uitgevoerd?	6
4.	Hoe kunnen centrale toetsen bijdragen aan onderwijsontwikkeling?	8
5.	Besluit.....	8
II.	Bespreking.....	9
1.	Tussenkomen van de leden	9
1.1.	Hannelore Goeman	9
1.2.	Jan Laeremans	9
1.3.	Koen Daniëls	9
1.4.	Loes Vandromme	10
1.5.	Elisabeth Meuleman	11
1.6.	Kathleen Krekels.....	12
2.	Antwoorden van het Steunpunt Centrale Toetsen in Onderwijs	12
3.	Replieken van de leden.....	16
	Gebruikte afkortingen	17
	Bijlage: zie dossierpagina van dit document op www.vlaamsparlement.be	

Op 12 januari 2023 organiseerde de Commissie voor Onderwijs een gedachtewisseling over de ontwikkeling van de Vlaamse toetsen door het Steunpunt Centrale Toetsen in Onderwijs (verderop: Steunpunt), met:

- Johan van Braak, promotor-coördinator van het Steunpunt en professor verbonden aan de vakgroep Onderwijskunde van de UGent;
- Carolien Frijns, strategisch coördinator van het Steunpunt en postdoctoraal onderzoeker verbonden aan de vakgroep Onderwijskunde van de UGent;
- Dimokritos Kavadias, werkdomeincoördinator Verwerking en Analyse van het Steunpunt en professor Politieke Wetenschappen aan de VUB;
- Ilse Laurijssen, teamleider Verwerking en Analyse van het Steunpunt en postdoctoraal onderzoeker bij het Departement Sociologie van de VUB;
- Rianne Janssen, werkdomeincoördinator Psychometrie van het Steunpunt en professor verbonden aan de Faculteit Psychologie en Pedagogische Wetenschappen van de KU Leuven;
- Lien Willem, teamleider Psychometrie van het Steunpunt en verbonden aan de eenheid Onderwijseffectiviteit en -evaluatie van de KU Leuven.

De nota 'Leerprestaties en leerwinst' is te vinden op de [dossierpagina](#) van dit document op www.vlaamsparlement.be.

I. Uiteenzetting door het Steunpunt Centrale Toetsen in Onderwijs

1. Opdracht van het Steunpunt

Johan van Braak roept kort de opdracht in herinnering die het Steunpunt van de Vlaamse Regering kreeg: wetenschappelijk onderbouwde beleidsinformatie leveren op verschillende niveaus (leerling, leerlingengroep, school en systeem) over het behalen van de eindtermen wiskunde en Nederlands. Daarvoor zijn vier momenten van de schoolloopbaan geselecteerd. De afname is populatiebreed, wat betekent dat alle leerlingen in het Vlaamse leerplichtonderwijs deelnemen, behalve de in het decreet beschreven uitzonderingen.

Verder staat het Steunpunt in voor de verwerking en analyse van de resultaten. Met het oog op snelle terugkoppeling wordt ingezet op automatisering, wat mogelijk wordt gemaakt door de afname op een digitaal platform. Vervolgens wordt ook in feedback voorzien op de verschillende genoemde niveaus, waartoe de actoren toegang zullen hebben via een platform. Het Steunpunt voert voorts trendanalyses uit, waartoe verschillende cohortes over de tijd worden gevolgd. Die opvolging maakt tot slot ook leerwinstmeting mogelijk.

Daarop gaat de spreker in op de toetskalender. De voorbereidingen zijn bezig en de eerste grote afname zal in het voorjaar van 2024 plaatsvinden in het vierde leerjaar lager onderwijs en het tweede leerjaar secundair onderwijs. Theoretisch kan in 2026 voor de eerste keer aan leerwinstmeting gedaan worden, al valt die datum wel buiten de eerste termijn van het Steunpunt.

Het is opvallend dat in het Steunpunt de expertise van alle universiteiten en twee hogescholen wordt samengebracht in een multidisciplinair team. Daarbij onderstreept Johan van Braak dat, met het oog op een draagvlak in de praktijk, de academische expertise voortdurend wordt aangevuld door mensen met ervaring in het onderwijs.

De spreker overloopt nogmaals de vijf kerntaken van het Steunpunt: centrale toetsen Nederlands en wiskunde ontwikkelen, scholen ondersteunen bij de afname van centrale toetsen, toetsresultaten verwerken en analyseren, feedback geven aan het onderwijsveld over de toetsresultaten en expertise ontwikkelen op het vlak van

duurzame onderwijseffectiviteit. Vandaag wordt gefocust op de gevraagde technische aspecten van de toetsen, met name eerste en derde kerntaak: centrale toetsen Nederlands en wiskunde ontwikkelen en toetsresultaten verwerken en analyseren.

2. Wat houden de centrale toetsen in?

Carolien Frijns gaat in op de vraag wat centrale toetsen inhouden en hoe ze ontwikkeld worden. Om te beginnen is men in de internationale literatuur op zoek gegaan naar krachtlijnen. Na gesprekken binnen het Steunpunt en met het onderwijsveld zijn er daar vier uit naar voren gekomen.

De eerste krachtlijn is dat toetsen betekenisvol moeten zijn, dat wil zeggen zo authentiek en functioneel mogelijke toetsen van wiskundige en talige vaardigheden. Vaardigheden als begrijpend lezen worden in kaart gebracht in een digitale toetsomgeving.

De tweede krachtlijn is toegankelijkheid. Welke redelijke aanpassingen nodig zijn om rekening te houden met leerlingen met specifieke onderwijsbehoeften, zoals het gebruik van voorleessoftware, werd bekeken in de literatuur en met de onderwijspraktijk. Concreet zal de ondersteuning waarop leerlingen recht hebben in de dagelijkse klaspraktijk zo veel mogelijk gefaciliteerd worden op het toetsplatform. Ook de toegankelijkheid van de digitale omgeving op zich wordt nagegaan in pilootonderzoek.

De derde krachtlijn wordt gevormd door de toegankelijkheid en de betrouwbaarheid van de toetsen. Daarom wordt ingezet op de wetenschappelijke validiteit van de toetsen, opdat ze meten wat ze moeten meten.

De vierde krachtlijn van de toetsen is dat ze digitaal en adaptief zijn. Het eerste maakt snelle verwerking van de resultaten mogelijk en op termijn ook het tweede: toetsen die zich aanpassen aan het niveau van de leerling.

Inhoudelijk wordt gefocust op twee domeinen: Nederlands en wiskunde. Voor het eerste is dat al afgebakend in de opdracht zelf: begrijpend lezen en schrijfvaardigheid met inbegrip van grammatica. De relevante eindtermen voor de genoemde leerjaren werden dan ook geselecteerd.

Voor wiskunde ontbrak dergelijke afbakening in de opdracht en is het Steunpunt zelf op zoek gegaan naar een zo goed mogelijke vormgeving, in overleg met de opdrachtgever, de stuurgroep en het onderwijsveld. Resultaat was een tweevoudige aanpak: een breedtoets wiskundige problemen oplossen, waarin verschillende vaardigheden aan bod komen en die door alle leerlingen wordt afgelegd, naast een aantal focustoetsen over specifieke thema's als getallenleer of verzamelingen die groepen van leerlingen afleggen.

Vervolgens gaat de spreker in op de ontwikkeling van de toetsen. Na de eindtermselectie wordt een eerste toetsontwerp gemaakt met de eerste toetsvragen, die vervolgens worden afgenomen in pilootonderzoeken en tot slot in een groot-schalige kalibratie. In de verschillende stappen zijn ook denktanks aan het werk: praktijkreflectiegroepen uit het onderwijsveld, die bij elke stap kritisch meekijken. Vertegenwoordigd in die groepen zijn leraren, lerarenopleiders, pedagogisch begeleiders en inspecteurs voor de verschillende vakgebieden en onderwijsniveaus. Voor het lager onderwijs werden ook schoolleiders betrokken. Het hele proces van toetsontwikkeling vergt veel analyse en revisie, besluit de spreker. Naast de denktanks wordt dan ook veel ingezet op feedbackrondes in het Steunpunt met interne experts, met name de psychometrische. Het eindpunt is zoals gezegd de kalibratie.

3. Welke analyses worden uitgevoerd?

Rianne Janssen gaat nader in op de analyses, waarvan de eerste de toetsresultaten psychometrisch analyseert en de volgende (de data-analyse) op basis daarvan schoolverschillen onderzoekt.

Psychometrie hanteert statistische modellen voor het meten van leerlingresultaten. Die complexe techniek wordt gebruikt voor het optimaliseren van (a) de afname, (b) de analyse en (c) de interpretatie van grootschalige toetsen. Het is niet voor iedereen dezelfde proef en snelheid is geen belangrijk criterium. Wel gaat het zoals bij hoogspringen om zo nauwkeurig mogelijk het niveau van elke leerling te bepalen doorheen de centrale toetsen.

De meetschaal is opgebouwd uit toetsopgaven of items. Naarmate het item hoger op de schaal staat, is de opgave moeilijker. Doel is om ook de leerlingen zo precies mogelijk op diezelfde schaal te plaatsen. Gebruik van dergelijke meetschaal heeft als voordeel dat men – deels – kan toewerken naar toetsen op maat. In de plaats van identieke opgaven voor iedereen, kan men werken met bijvoorbeeld blokadaptieve toetsen. Hierbij worden aan de leerlingen opgaven voorgelegd, opgesplitst in groepen van telkens bijvoorbeeld tien items die elk een bepaald bereik van de meetschaal in kaart proberen te brengen. Leerlingen krijgen op basis van hun prestaties op een eerder deel een volgende deeltoets voorgelegd die beter aansluit bij hun vaardigheid. Daardoor wordt de meting nauwkeuriger maar kan zij toch plaatsvinden binnen dezelfde toetstijd.

Tweede reden voor het gebruik van de meetschaal is een zo nauwkeurig mogelijke bepaling van het niveau van elke leerling en van de mate van beheersing van de onderwijsdoelen. Dat laatste gebeurt aan de hand van cesuurbepaling op de meetschaal.

Doordat de Vlaamse toetsen jaarlijks herhaald zullen worden, kunnen de items elk jaar aangevuld worden en daardoor actueel blijven zonder dat men de equivalentie over de jaren heen uit het oog verliest. Erop toezien dat de meetschaal equivalent blijft over de verschillende afnamejaren, maakt het mogelijk om vast te stellen of er trends zijn in het Vlaamse onderwijs.

Daarnaast is er opvolging mogelijk op het niveau van de individuele leerling. Door de toetsen in het vierde en zesde leerjaar en in het tweede jaar van het secundair onderwijs te verbinden in een overkoepelende meetschaal, kan gemeten worden in welke mate leerlingen groei of leerwinst vertonen. Maar of dat in de toekomst ook gaat lukken van de eerste naar de derde graad in het secundair onderwijs, is nog niet duidelijk.

De bruikbaarheid van het geschetste model veronderstelt de vervulling van een aantal voorwaarden. Als eerste noemt ze de beschikbaarheid van een grote bank met items van verschillende moeilijkheidsgraad. Dat vergt een grote inspanning op het vlak van zowel toetsontwikkeling als kalibratie. De tweede is dat het model moet passen. Telkens opnieuw moet empirisch nagegaan worden in welke mate de eigenschappen van de meetschaal gelden voor de data. Dergelijke analyses zijn technisch en vragen tijd, waarmee rekening moet worden gehouden bij de afspraken over de snelheid van de feedback.

Ilse Laurijssen bespreekt de analyses die men wil bereiken op basis van de vaardigheidsscores en die uitgevoerd worden in functie van schoolfeedback en verder onderzoek. Hoe kunnen op grond van de toetsen uitspraken gedaan worden over de onderwijskwaliteit op schoolniveau? Doel van de feedback is immers om scholen bruikbare informatie te geven waarmee ze hun ontwikkeling kunnen opvolgen en hun beleid afstemmen. Daartoe hebben ze volgens het Steunpunt immers meer nodig dan weten waar hun leerlingen op een bepaald ogenblik staan.

Metingen zijn het resultaat van een – onvoltooid – proces. De feedback moet dan ook meer zijn dan de mededeling van een eindresultaat. Door ook het proces zelf mee in kaart te brengen, kunnen de scholen er meer greep op krijgen. Een dergelijk dynamisch perspectief is ook interessant voor het wetenschappelijke onderzoek, waar ze nu niet verder op ingaat.

Voor een correcte meting van schoolkwaliteit zijn een aantal duidelijk te onderscheiden elementen nodig, waarvan ook technische, die nauwkeuriger beschreven worden in de nota van het Steunpunt over leerprestaties en leerwinst. Een eerste belangrijk onderscheid is dat tussen status en groei.

Status verwijst naar vaardigheidsscores of leerprestaties op één meetmoment. Status op schoolniveau is het gemiddelde van alle leerlingen, of de mate waarin zij de eindtermen behalen. Status als indicator van onderwijskwaliteit zegt op zich niet zoveel over het onderwijseffect. Sommige scholen hebben immers een grotere instroom van sterkere leerlingen dan andere en mogelijk behalen zij goede resultaten zonder bijzondere inspanningen.

Daarom is het interessanter om te kijken naar groei: de vooruitgang die een leerling maakt doorheen de tijd of de leerwinst die zij of hij behaalt. Op schoolniveau drukt die uit hoeveel verder de leerlingen eindigen in vergelijking met hun aanvangspositie. De spreker geeft een fictief voorbeeld. Scholen kunnen hun leerlingen niet zomaar kiezen: de leerlingen van de ene school starten met minder bagage dan die van de andere. Toch slaagt de eerste school er bijvoorbeeld in om een grote groep tot de eindtermen te brengen, zij het niet veel verder. Niet naar de leerwinst kijken zou dan ook tekortdoen aan haar inspanningen. In de school met leerlingen met een grotere startbagage kunnen de prestaties eveneens relatief goed zijn, maar zij had dan, gegeven de startpositie van haar leerlingen, meer kunnen bereiken.

De vraag is nu of met de meting van de leerwinst het schooleffect voldoende wordt ingeschat: de bijdrage die een school biedt aan het leren door het onderwijs dat ze geeft. Leerwinst is belangrijk voor de bepaling van die toegevoegde waarde, maar onvoldoende. Er zijn immers andere mogelijke invloeden die mee de prestaties bepalen. Die invloed van de buitenwereld komt op drie manieren binnen: de input van de leerlingen (de eerdergenoemde aanvangspositie), het leren buiten de school (bijvoorbeeld in het gezin of tijdens buitenschoolse activiteiten) en de schoolkenmerken die (op korte termijn) moeilijk te veranderen zijn.

Doel is om zo goed mogelijk al die buitenschoolse factoren mee in rekening te brengen. Heel belangrijk daarbij is rekening houden met de achtergrondkenmerken. Dat is des te crucialer op het moment dat men nog niet beschikt over indicatoren van eerdere leerprestaties, zoals bij de eerste meting in het vierde leerjaar. Hoe dan ook worden scholen beter geïnformeerd over hun effectiviteit door al die factoren in rekening te brengen en niet alleen het bereiken van de einddoelen.

Dergelijke gecontextualiseerde leerwinst wil het Steunpunt op een begrijpelijke manier terugkoppelen naar de scholen. Een belangrijk nieuw concept is dat van – op basis van statistische berekeningen – vergelijkbare scholen. De spreker verwijst daarvoor naar de Australische schoolfeedback. Leerwinst wordt daar systematisch gemeten waarbij alle leerlingen op vier momenten in de schoolloopbaan worden getoetst in NAPLAN. Het resultaat van een school wordt niet alleen vergeleken met het gemiddelde van alle scholen, maar ook met dat van de leerlingen met dezelfde aanvangsscore en een vergelijkbare sociale achtergrond. Leerwinst moet men altijd in combinatie met status bekijken, besluit de spreker, zowel in de rapportering als bij de interpretatie. Op die manier levert men in de feedback relevante informatie.

De uitdagingen voor dergelijke gecontextualiseerde leerwinst worden beschreven in de eerdergenoemde nota. Afgezien van de meting zelf, is de statistische betrouwbaarheid van de uitspraken afhankelijk van de grootte van de groepen. Bij kleine aantallen is minder zeker of een resultaat echt afwijkt van het gemiddelde van alle of dat van vergelijkbare scholen. Dat probleem doet zich evenwel niet voor op systeemniveau, en evenmin voor de meeste scholen.

Ander punt is, zoals gezegd, dat achtergrondkenmerken van leerlingen moeten worden meegenomen. Daar zijn goede indicatoren voor nodig. Het Steunpunt is volop bezig met de bepaling van de meest krachtige, waarover ze een advies aankondigt.

Tot slot attendeert de spreker op het belang van niet-standaard schoolloopbanen ten gevolge van zittenblijven of verandering van school. Als men daarmee geen rekening houdt, vallen bepaalde groepen van leerlingen buiten het bereik van de berekening van leerwinst op schoolniveau, wat aanleiding kan geven tot strategisch gedrag bij scholen, zoals uit buitenlandse voorbeelden blijkt. Dat wil het Steunpunt voorkomen, maar de spreker erkent dat er nog werk is aan de ontwikkeling van een goede methode. De eerste leerwinstmeting is pas over drie jaar.

4. Hoe kunnen centrale toetsen bijdragen aan onderwijsontwikkeling?

Carolien Frijns gaat in op de bijdrage van centrale toetsen aan onderwijs- en schoolontwikkeling. Om te beginnen wordt daartoe ingezet op algemeen onderzoek naar onderwijs-effectiviteit: wat zijn de kenmerken van scholen of leraren die ertoe doen? Daarover brengt het Steunpunt verslag uit in rapporten en brochures.

Daarnaast wijdt het Steunpunt zich aan specifieke schoolfeedback na deelname aan een toets, waarmee de school aan de slag kan in een veilige context, gericht op interne kwaliteitszorg. Belangrijk daarbij is dat de resultaten toegankelijk zijn, zodat leraren en schoolleiders ermee aan de slag kunnen. Met het oog daarop wordt ook aan gebruikersonderzoek gedaan, waarin naar hun verwachtingen wordt gepeild en naar de gewenste visualisering. In dat verband vermeldt ze nog een onlinecursus over informatiegeletterdheid en de proeftuin die daarover in het voorjaar van 2023 van start gaat bij alle Vlaamse lerarenopleidingen lager onderwijs.

De spreker verzekert nogmaals dat met het oog op voldoende draagvlak intensief wordt samengewerkt met het onderwijsveld. Ze vermeldt nogmaals de denktanks bij de toetsontwikkeling, naast de ontwikkeling van ondersteunende materialen voor het onderwijsveld zoals de e-course over informatiegeletterdheid. Verder is heldere communicatie belangrijk. Tot slot wijst de spreker op het belang dat het Steunpunt hecht aan samenwerking, met de actieve inzet van denktanks over eindtermenselectie en toetsen, focusgroepen over toegankelijke toetsen en gebruikersgroepen over feedbackgebruik.

5. Besluit

Johan van Braak vat samen. Centrale toetsen voor onderwijsontwikkeling zijn toetsen die betekenisvol, toegankelijk, valide en betrouwbaar, digitaal en adaptief zijn, in nauwe samenwerking met het onderwijsveld worden ontwikkeld, en niet alleen gericht zijn op status maar ook op leerwinstanalyses, en waar mogelijk als indicator van de relatieve bijdrage van scholen aan het leren. Die professionele aanpak gebeurt bovendien in een veilige context ter ondersteuning van – voornamelijk interne – kwaliteitszorg. De afname van toetsen garandeert op zich immers nog geen kwaliteit.

Tot slot wijst hij nog op het webadres www.steunpunttoetsen.be en op de visietekst van het Steunpunt die raadpleegbaar is via www.steunpunttoetsen.be/visietekst.

II. Bespreking

1. Tussenkomsten van de leden

1.1. Hannelore Goeman

Hannelore Goeman verzekert dat haar fractie de toetsen een interessant instrument vindt voor de kwaliteitsbewaking. Een dergelijke grote stap vereist wel de garantie van een goed begrip en van realistische beloften vanwege het beleid. De centrale vraag blijft voor haar die naar de mogelijkheden van de leerwinstmeting. Hoe zorgt men ervoor dat de niveaus die bereikt worden bij metingen in verschillende leerjaren vergelijkbaar zijn, aangezien die toch verschillende eindtermen hebben? Het lid is alvast tevreden dat de leerwinst gecontextualiseerd wordt in de feedback. Zij vraagt wel nadere toelichting bij het begrip 'vergelijkbare scholen'.

Hoe wordt in het lager onderwijs de vergelijking gemaakt tussen het vierde en het zesde leerjaar, terwijl er alleen voor dat laatste eindtermen zijn? In het secundair onderwijs komt daar nog bij dat die in elke graad veranderen, en dat er bovendien verschillen zijn volgens de richtingen. Of wordt alleen het basispakket getoetst? En wat met leerlingen die van richting, finaliteit of school veranderen? Hoe kan men in dergelijke gevallen leerwinst op het niveau van de leerling meten? In dat verband worstelt zij ook met het begrip focustoetsen en hun effect op de vergelijkbaarheid van leerlingen en scholen.

Tot slot informeert ze naar de verhouding tussen de adaptieve toetsen en de minima die in de eindtermen staan. Leerwinst meten is goed, maar een goede school moet toch minimaal de eindtermen halen? En zijn er ook vragen die verder gaan?

1.2. Jan Laeremans

De leerwinst lijkt *Jan Laeremans* makkelijk te meten in de vergelijking van twee leerjaren in een lagere school, maar die met het tweede jaar van het secundair onderwijs is een ander verhaal. Het gaat immers om een andere school en de leerlingen komen uit verschillende lagere scholen. Of gebeurt de vergelijking tussen lager en middelbaar niveau alleen voor individuele leerlingen? Wat is in dat verband de verklaring voor de grote sprong van vier jaar tussen het tweede en het zesde middelbaar? Hoe ziet het Steunpunt die keuze, die zij weliswaar niet zelf maakte?

Veel toetsen hebben te maken met taal. Het lid vraagt uitleg bij de deelname van de talrijke leerlingen die later instromen en thuis een andere taal spreken. Doen zij mee aan de toetsen of moeten ze eerst een minimumniveau bereiken? Hoe vermijdt men vertekening van de resultaten van hun klas?

Jan Laeremans weet dat er scholen zijn die het resultaat van de toetsen zullen laten meetellen bij de deliberatie en andere die dat niet zullen doen. Ze moeten dat opnemen in hun reglement. Houdt het Steunpunt rekening met het motivatie-effect daarvan bij de afname?

1.3. Koen Daniëls

Koen Daniëls ziet als voordeel voor het Steunpunt dat Vlaanderen een van de laatste regio's ter wereld is die dergelijke toetsen ontwikkelt. Er valt dan ook veel te leren van andere landen, zoals het goede voorbeeld van Australië. Wat dat betreft, wil hij weten wat de gecontextualiseerde benchmarking betekent voor uitspraken op Vlaams niveau. Blijven die mogelijk of alleen nog per profiel, dan wel in een soort van boxplot?

Het lid treedt bij dat samenwerking met het onderwijsveld belangrijk is, maar zij mag niet verlamd werken, waarschuwt hij. Het mag niet zo zijn dat een item dat gebaseerd is op de minimumdoelen wordt geweerd omdat bepaalde onderwijsverstrekkers dat vragen op grond van het handboek dat zij gebruiken. Komen dergelijke verzoeken naar boven in de denktanks? Hij benadrukt dat de onderwijskwaliteit getoetst moet worden en niet de leerplannen. Als de onderwijsverstrekkers vinden dat de pedagogische projecten getoetst moeten worden, moeten ze dat maar zeggen, maar hij denkt van niet.

Voor zijn fractie is het adaptieve karakter cruciaal. Te gemakkelijke items leiden voor sommige leerlingen tot verveling. Het komt erop aan de moeilijkheidsgraad geleidelijk op en af te bouwen om te achterhalen tot waar een leerling raakt. Mag hij ervan uitgaan dat men zal blijven items voorleggen tot men het maximum van een leerling bereikt, ondanks het feit dat er in de tabel maar 1 item staat boven de cesuur voor eindtermen van de A-stroom? Hij herinnert er daarbij aan dat er in de PISA-toetsen nog 4 niveaus boven de basisgeletterdheid van niveau 2 zijn. Wat is voor het Steunpunt het adaptieve plafond?

Het longitudinale interesseert hem sterk. Zal het opnieuw mogelijk worden om een leerling te volgen zoals in het LOSO-onderzoek van professor Van Damme? Wordt daarin rekening gehouden met de adviezen die een leerling kreeg maar mogelijk niet volgde? Wat dat betreft, suggereert hij om de achtergrondvariabelen uit de genoemde studie te bekijken. Het verdient immers de voorkeur de dataset vanaf het begin ruim genoeg te nemen. In dat verband pleit hij ervoor om de thuistaal in elk geval als aparte variabele op te nemen. Werkt het Steunpunt in het algemeen zoveel mogelijk met losse variabelen of met clusters? Wordt de in handboeken gebruikte methode als variabele opgenomen? Waarom (niet)? Ook die kan immers de verklaring blijken te zijn voor verschillen in resultaten.

In welke mate zal de coronacrisis de kalibratie beïnvloeden of is dat effect te verwaarlozen omdat het voor alle leerlingen geldt?

Wordt de schrijfvaardigheid getoetst met pen en papier of een toetsenbord? De ene school oefent immers met het eerste en de andere met het tweede. Hij herinnert in dat verband aan de paralleltoetsen die destijds bestonden bij de peilingsproeven, die scholen ook als oefening konden gebruiken. Ook in de PISA-toetsen bleek de overgang van papier naar digitaal een effect te hebben. Hoe kunnen scholen daarin begeleid worden? Aanvullend merkt hij op dat de mate waarin leerlingen al ervaring hebben met de gecontextualiseerde opgaven die men wil gebruiken, een rol kan spelen in de resultaten. Hoe kan de taalbeheersing, die een grotere rol speelt dan bij zuivere rekenoefeningen, eruit gefilterd worden, zodat men wel de gelijk de reken- en niet de taalvaardigheid test?

In een school met maar één vierde leerjaar is het klaseffect gelijk aan het school-effect, maar dat is niet zo in het geval van parallelle klassen. Kunnen dergelijke scholen uit de resultaten ook informatie halen op klasniveau? Is het statistisch relevant om de praktijk van de leerkrachten te vergelijken?

Itemverversing zonder de plaats in het adaptief model te wijzigen, lijkt hem wetenschappelijk vrijwel onmogelijk, tenzij men louter getallen vervangt, en dan nog. Hoe wordt het longitudinale bewaakt?

1.4. Loes Vandromme

Ook voor de fractie van *Loes Vandromme* is gecontextualiseerde leerwinst een belangrijke kwestie. Er zijn inderdaad veel data beschikbaar, maar hoe wordt de co-horte bepaald? Welke kenmerken worden gebundeld? Beschikt het Steunpunt over alle informatie, nu gebleken is dat ook de inspectie niet alles doorstuurt naar de

schoolbesturen? Wie is eigenaar van de data en wie beschikt erover? Het samenstellen van groepen lijkt haar een grote uitdaging in het licht van bijvoorbeeld het verschil tussen A- en B-stroom met hun verschillende eindtermen.

Haar fractie blijft erbij dat geen enkele regel bepaalt welke eindterm in welk leerjaar van de lagere school behaald moet worden. Dat scholen hun pedagogisch project zouden moeten aanpassen aan de keuze van onderwijsdoelen per leerjaar die in de toets gemaakt wordt, is niet volgens de Vlaamse manier van onderwijsorganisatie.

Toetsen moeten in se adaptief zijn, want er zijn eindtermen die gelijkwaardig bevonden zijn en aanvaard worden als vervangend. Hoe ziet het Steunpunt dat? Verder staat in de ontwerptekst van het decreet dat de eindtermen de lat vormen en dat de toetsen dienen om te weten te komen of ze gehaald worden. Wat Koen Daniëls voorstaat, namelijk toetsen tot een leerling het niet meer kan, gaat veel verder. Zij ziet niet hoe een en ander verzoend kan worden.

Wat de veilige context betreft, is in de haalbaarheidsstudie gebleken dat de centrale toetsen de peilingstoetsen vrijwel niet kunnen vervangen op systeemniveau. Hoe reageert het Steunpunt? Bij de leerwinst op klasniveau hoopt ze dat die gemeten is over verschillende jaren en dat de leerkracht dus niet als enige verantwoordelijk is maar wel het hele team. Het lid heeft begrip voor de bezorgdheid van leerkrachten dat de informatie gebruikt zou worden om hen te evalueren. Wat antwoordt het Steunpunt?

De mogelijkheid van oefentoetsen staat volgens haar al in de huidige tekst van het ontwerp van decreet, de breedtetoetsen en dergelijke evenwel niet. Waar worden die geregeld? Op welke basis werden die keuzes gemaakt? Bestaan daar al afspraken over? Het lid is in het algemeen bezorgd over de stappen die het Steunpunt al heeft gezet zonder decretale basis. Ook de Vlor stelt vragen over de betrokkenheid. Wat is de mening van het Steunpunt?

Loes Vandromme vraagt nog uitleg bij de oefentoetsen die in het huidige ontwerp van decreet staan. Wie krijgt er toegang toe? Hoe worden ze opgemaakt? Wat is hun band met de breedtetoetsen?

1.5. Elisabeth Meuleman

Elisabeth Meuleman benadrukt de grote impact van de invoering van de centrale toetsen, maar de vele bezorgdheden waarmee het Steunpunt rekening houdt en de ernst waarmee het dat doet, stellen haar gerust. Zij treedt bij dat de verhoging van de onderwijskwaliteit het doel moet zijn, al spreekt het Steunpunt zelf van effectiviteit. Wat verstaat het daar precies onder?

Haar fractie wil niet dat de toetsen Vlaanderen doen inboeten op het vlak van gelijke onderwijskansen, waarop het nu al niet goed scoort. Het Steunpunt antwoordt daarop met zijn gecontextualiseerde leerwinst, maar die is niet evident. Zij sluit zich dan ook aan bij de vraag hoe men die precies gaat meten. De vrees blijft dat mensen in hun perceptie geen rekening zullen houden met die context en dat er een verdere segregatie komt tussen wat men als betere scholen ziet en andere. Komt er longitudinaal onderzoek over de gelijke onderwijskansen?

Een andere vrees is dat het onderwijs voortaan eenzijdig focust op wat getoetst wordt. Misschien is de itemverversing bedoeld om de toetsen onvoorspelbaarder te maken, maar op de duur zal men wel weten wat ongeveer getest zal worden. Hoe kijkt het Steunpunt aan tegen 'teaching to the test'?

Verder sluit ze zich aan bij de vragen over de eindtermen die pas aan het einde van het lager onderwijs gehaald moeten worden. Zij worden trouwens binnenkort vernieuwd, waardoor de vraag rijst of de toetsen meteen grondig aangepast moeten worden.

Hoe overtuigd zijn de onderzoekers zelf dat de centrale toetsen een positief effect zullen hebben op de onderwijskwaliteit en waarom?

1.6. Kathleen Krekels

Kathleen Krekels is enthousiast en gelooft dat de toetsen inzichten zullen opleveren over welke verbeteringen mogelijk zijn. Zij heeft mogen vaststellen dat uit de OVSG-toetsen al veel kan worden gehaald, maar bij de centrale toetsen zal dat tot de tiende macht zijn. De toegankelijkheid is wel een belangrijk aspect, zowel voor leerkrachten als voor leerlingen, in het bijzonder die met specifieke leerbehoeften. Wat kan men op dat vlak leren uit de ervaring van andere landen?

Voor de fractie van Hannelore Goeman moet het doel van de toetsen zijn om de onderwijskwaliteit te monitoren op zowel Vlaams als schoolniveau. Klopt, wat het eerste betreft, dat men zal kunnen weergeven hoeveel leerlingen de eindtermen halen en hoe dat evolueert, maar dat het moeilijker is voor de contextfactoren op dat niveau? Wat het tweede betreft, benadrukt ze nogmaals hoe belangrijk het is dat de leerwinst wordt getest, maar evenzeer hoeveel leerlingen de eindtermen halen. Komen de scholen ook dat laatste te weten in de feedback? Kwaliteitsvol onderwijs impliceert voor haar immers dat de leerlingen de minimumdoelen halen, wat helaas vandaag niet in alle scholen lukt.

Tot slot vraagt ze waarom het Steunpunt spreekt van adaptieve toetsen terwijl de minister het altijd had over semi-adaptieve. Wat betekent dat verschil?

2. Antwoorden van het Steunpunt Centrale Toetsen in Onderwijs

Johan van Braak verzekert dat het Steunpunt enthousiast is over de centrale toetsen en veel verwacht van hun effecten. Hij wijst er wel op dat zij in het buitenland soms met andere bedoelingen in de markt gezet zijn, de zogenaamde 'high stakes'-toetsen, waarop ook kritiek is gekomen. In Vlaanderen wordt ingezet op kwaliteitszorg op schoolniveau. Het succes daarvan is afhankelijk van de bereidheid van scholen om met de data aan de slag te gaan, met de tools die door het Steunpunt worden aangereikt in samenwerking met de pedagogische begeleidingsdiensten. De beweging naar meer uitkomst- en datagestuurd onderwijsbeleid is al in de jaren 80 van de vorige eeuw ingezet. Scholen hanteren daarin kwantitatieve data om processen vorm te geven. Wat er uitkomt, wordt opgevolgd met onderwijs-effectiviteitsonderzoek.

Vanaf dag 1 een model uittekenen voor de meting van leerwinst op vier momenten en op vier aggregatieniveaus is onhaalbaar. Daarom is geen operationalisering gepresenteerd, want dat is ook een beleidsbeslissing. Wat wel naar voren is gebracht, zijn de technische contouren die het mogelijk maken om de leerwinst in beeld te brengen. Het is geen voorafname van welke operationalisering dan ook, verzekert hij.

Dat het eenvoudiger zal zijn van het vierde naar het zesde leerjaar dan van dat laatste naar het zesde middelbaar, is evident. Verder is duidelijk dat in de gefaseerde implementatie in het begin voorrang wordt gegeven aan grote domeinen zoals leesbegrip en wiskundige problemen oplossen als breedtedomein, wat gemakkelijker in kaart te brengen zal zijn dan de leerwinst in focustoetsen.

Wat de oefentoetsen betreft, is het voor het Steunpunt belangrijk dat leerlingen voldoende vertrouwd raken met de manier waarop ze zijn opgesteld en hoe de

opgaven eruitzien. Ze zijn inderdaad digitaal en niet eerst met pen en papier. Doel van de oefentoetsen is dan ook de vertrouwdsheid met het digitale platform zo groot mogelijk te maken. Het Steunpunt wil met het oog op de validiteit immers vermijden dat de meetresultaten beïnvloed worden door het nieuwe karakter van de toetsen. De spreker brengt nog de nuance aan dat de oefentoetsen, ook wel kennismakingstoetsen genoemd, eigenlijk geen toetsen zijn maar een bundeling van opgaven ter kennismaking, en dat het dus niet de bedoeling is hun resultaten te aggregeren en terug te koppelen.

Carolien Frijns antwoordt dat alle toetsvragen op de eindtermen gebaseerd zijn. Vervolgens wordt een gekalibreerde meetschaal ontwikkeld, waarop op een wetenschappelijke manier samen met het onderwijsveld een cesuur wordt aangebracht. De eindtermen vormen dus wel degelijk een blauwdruk van de toetsmatrijs. Begrijpend lezen en wiskundige problemen oplossen worden aan alle leerlingen voorgeschoteld. Op die vlakken zal men dus in elk geval met leerwinst aan de slag kunnen gaan.

Ilse Laurijssen legt uit dat de meting van leerwinst een toetsdomein veronderstelt met een stabiele inhoudelijke basis. Die vindt men eerder in ontwikkelingsgerichte vaardigheden zoals de twee genoemde. Zij kunnen op elk moment van de schoolloopbaan getoetst worden. Zij zijn ook zo generiek dat ze een gemeenschappelijke kern hebben in de eindtermen van de verschillende leerjaren. Doel is te werken met ankeritems zodat de toetsen geëquivalet kunnen worden.

Het klopt dat het moeilijker kan worden om gemeenschappelijke items op te nemen als er meerdere schooljaren tussen zitten, zoals tussen het tweede en het zesde jaar secundair. Dat zal men moeten bekijken en het is mogelijk dat het niet lukt. Een tussenliggende meting kan dan een optie zijn maar die vergt bijkomend onderzoek en investeringen. Dat spoor ligt nog open.

De overgang tussen lager en middelbaar onderwijs wordt opgelost door de meting in het zesde leerjaar te zien als een aanvangsmeting bij de start van het secundair onderwijs. Dat de leerlingen uit verschillende scholen komen is dan niet zo'n probleem.

Dimokritos Kavadias vult aan dat de bepaling van vergelijkbare scholen nog bezig is.

Verder zijn de data van LiSO, de opvolger van LOSO, wel degelijk opgevraagd en is men die aan het verrijken met administratieve gegevens om uit te maken wat de meest krachtige indicatoren zijn. Het is niet de bedoeling er één cluster van te maken. Dat is nuttig voor administratief gebruik maar niet voor begrip, want daarvoor moet men naar de componenten kijken. De demografische dynamiek van Brussel en de Rand schuift inderdaad anderstalige nieuwkomer als kenmerk naar voren. De uiteindelijke keuze is aan het parlement. Om analyses zoals die van LiSO te herhalen, lijkt een steekproef van de deelnemende scholen hem overigens nuttiger dan de voorliggende zware machinerie, want daarmee belast men niet alle scholen tegelijk.

Ilse Laurijssen bevestigt dat het Steunpunt volop aan het bekijken is welke kenmerken men hoe meeneemt. Het principe is om maximaal gebruik te maken van degene die relatief gemakkelijk beschikbaar zijn, bijvoorbeeld in de databanken van het Departement Onderwijs en Vorming. Wat daarnaast nodig is, wordt bepaald op basis van literatuurstudie, de analyse van bestaande data en de koppeling van extra gegevens aan de kalibratiedata. Dergelijke bijkomende gegevens verzamelen op populatieniveau is niet evident, waarschuwt ze. Er is ook niet in een algemene bevraging van de scholen voorzien, in het licht van de taakbelasting. Aan de afname van de toetsen wordt wel een leerlingensurvey gekoppeld, waarin

nog noodzakelijke kenmerken verzameld zouden kunnen worden, in aanvulling op bijvoorbeeld het gebruik van hulpmiddelen en de motivatie.

Johan van Braak gaat in op de toetsduur. Wiskunde is een breed domein en er is alvast het genoemde onderscheid gemaakt tussen een breedtetoets en focustoetsen. Het volledige curriculum peilen, vergt immers bij wijze van spreken twee weken. Binnen de toetsduur is inderdaad ook een beperkte tijd voorzien voor vragenlijsten. Scholen gaan immers de toetsen op uiteenlopende wijze kaderen, zoals dat ook het geval is met de opname in het schoolreglement. Dat heeft zijn effect op de motivatie, die vanzelfsprekend op haar beurt weer een effect heeft op het toetsresultaat.

Wat het Steunpunt ook wil te weten komen, is het onderwijsaanbod. Maakt de keuze van het handboek een verschil? Welke onderwijstijd wordt besteed aan specifieke leergebieden? In welke mate doen scholen en leerkrachten inspanningen om leerlingen wiskundige problemen te leren oplossen? Instructietijd blijkt immers een belangrijke verklarende factor voor resultaten. Na aggregatie op systeemniveau is de uitkomst daarvan ook instrumenteel voor de decreetgever voor beleidsbeslissingen.

Rianne Janssen behandelt de psychometrische vragen. Itemverversing is niet een item vervangen door een ander maar een groep van items toevoegen en bepalen waar ze op de meetschaal komen, waarna ze tijdens de volgende adaptieve toets worden afgenomen.

Wat zij tijdens haar presentatie voorstelde, is inderdaad semi-adaptief. Er wordt immers gewerkt met blokken of groepen, terwijl maximaal adaptief inhoudt dat na elk item een nieuw item wordt bepaald. Het afnameplatform moet geschikt zijn voor een dergelijk complex interactief systeem, zeker met zoveel scholen en leerlingen tegelijk. Blokadaptief is een minder risicovol alternatief. Het verlies aan gewenste nauwkeurigheid valt volgens haar gunstig uit.

De spreker bevestigt dat het Steunpunt zich houdt aan de onderwijsdoelen, met inbegrip van de uitbreidingsdoelen, die voor Nederlands verder gaan dan de eindtermen voor de A-stroom in de eerste graad secundair onderwijs. Voor wiskunde zijn die laatste er niet. Variatie op de meetschaal is er in elk geval, verzekert ze, want opgaven die over dezelfde eindterm gaan, kunnen makkelijker of moeilijker blijken. Daarom is er ook die de cesuurbepaling door experts. Het is eigen aan de meetschaal met items, dat de lat pas achteraf wordt gelegd en dat leerlingen dus verder kan opgaan op de meetschaal dan die lat, maar het is niet de intentie om toetsen te ontwikkelen die verder gaan dan de eindtermen.

Het blijft onduidelijk voor *Hannelore Goeman* hoe die cesuur wordt bepaald. Wat vindt men voldoende in het semi-adaptieve parcours voor het halen van de eindtermen?

Rianne Janssen geeft toe dat het een moeilijke kwestie is. Experts (onder anderen leerkrachten, inspecteurs en pedagogische begeleiders) beoordelen welke vertalingen van de abstracte eindtermen in vragen op het minimumniveau zitten en welke dat overschrijden. De lat wordt gelegd in overleg met het onderwijsveld. Eens ze gelegd is, kan het Steunpunt haar wel behouden.

De grote schaal van de toetsafname verhindert dat men alle mogelijke vragen kan voorleggen. Het systeem vereist immers snelle en automatische scores. Hoe complexer de context, hoe lastiger dat wordt. De toetsontwikkeling is gericht op contextrijke vragen, maar men moet afwegen met de grootschaligheid van de afnames. Er wordt in de kwaliteitsbewaking ook rekening mee gehouden dat de taalvaardigheid geen al te grote rol speelt.

De logica van de kalibratiestudie is dat de verschillen in moeilijkheidsgraad tussen de items worden bepaald. Zelfs als leerlingen lager presteren ten gevolge van de coronacrisis, blijft dat mogelijk.

De regeling voor leerlingen van scholen met gelijkwaardige eindtermen, welke items zij voorgelegd krijgen en hoe zij passen in de meetschaal, is nog niet in detail besproken.

Leerwinst kan alleen gemeten worden voor die toetsen waarvoor dat voorzien wordt (probleemoplossen en lezen, maar niet de focustoetsen wiskunde). Het kan alleen gemeten worden als het model past, en dat moet nagegaan worden. De afstanden zijn inderdaad niet gunstig. De toetsontwikkeling voor het vierde leerjaar is wel gebaseerd op die van het zesde, wat longitudinale meting mogelijk maakt, maar ook dat moet nog blijken. Verandering van school compliceert inderdaad longitudinale opvolging.

De centrale toetsen zullen wel degelijk informatie opleveren op systeemniveau, wat ook belangrijk is, bijvoorbeeld in geval van problemen die zich over heel Vlaanderen blijken voor te doen, zoals al bij peilingsresultaten is gebleken. De steekproeven zijn zelfs groter dan bij peilingsonderzoek. Bij sommige toetsen is de afname populatiebreed. Verder bevestigt zij dat bij peilingsonderzoek wel degelijk gekeken wordt naar verschillen tussen handboeken. Allicht zal dat ook zo zijn in het Steunpunt.

Carolien Frijns bevestigt andermaal dat het Steunpunt in de literatuur heeft onderzocht op welke manier men kan zorgen voor redelijke toetsaanpassingen. Over toegankelijkheid zijn, ten eerste, focusgroepen ingericht over heel Vlaanderen met onder anderen leraren en pedagogische begeleiders over hun wensen en praktijken. Binnenkort verschijnt daarover een rapport op de website. De uitkomst komt erop neer dat men, als men leerlingen kan bieden wat ze gewend zijn in de dagelijkse klaspraktijk door het na te bootsen op het digitale toetsplatform, de afname daarmee ondersteunt. Daarom zal er bijvoorbeeld voor gezorgd worden dat leerlingen de hun bekende voorleessoftware kunnen koppelen aan het toetsplatform.

Kennismakings- of oefeningstoetsen zijn, ten tweede, uiteraard belangrijk. De spreker benadrukt dat het niet om drillen gaat maar om het vinden van de weg in de omgeving, zodat alle leerlingen met gelijke bagage kunnen beginnen. Daarom ook wordt voorzien in ondersteuning van de scholen om de kennismakingstoetsen in te richten in de klas. Niet iedereen krijgt immers thuis die kans.

Een derde aspect van de toegankelijkheid is de eerlijkheid van de items. Door middel van psychometrische DIF-analyses wordt bijvoorbeeld voorkomen dat bepaalde toetsvragen moeilijker zijn voor jongens dan voor meisjes met hetzelfde niveau voor bijvoorbeeld begrijpend lezen.

Vierde element is het afnemen van think alouds in de scholen, 'hardopdenkprotocollen', met interviewvragen aan de leerlingen, om te zien hoe ze de toetsen kwalitatief ervaren. Het Steunpunt begint daarover overigens een pilootstudie.

Verder is het logisch dat leraren benieuwd zijn naar de feedback over hun lesgroepen of klassen. De spreker benadrukt dat die resultaten geen evaluatie vormen van de leerkracht. Zij worden namelijk ook bepaald door eerdere leraren en door achtergrond- en schoolkenmerken. Het gaat wel degelijk om een team effort, treedt ze bij.

Volgens haar bieden de toetsresultaten – naast andere gegevens – juist kansen voor de verbetering van de onderwijskansen van kwetsbare leerlingen op bepaalde voorwaarden. Het laatste rapport van het Kinderrechtencommissariaat wijst in

dezelfde richting, net als het advies van de Taalunie. Voorwaarden zijn wel interne kwaliteitszorg, een veilige context en een draagvlak in het onderwijsveld. Men kan die nalezen in de visietekst van het Steunpunt. Feedback en onderzoek vormen volgens haar de hefboomen.

Op de vraag, in het licht van de effectiviteit, naar het in kaart brengen van de methodes die scholen hanteren, antwoordt *Johan van Braak* dat men het onderwijsaanbod op verschillende manieren kan benaderen, maar het Steunpunt moet hoe dan ook bescheiden zijn in zijn bevraging van leerkrachten en leerlingen. Hij herhaalt daarom dat een selectie bevraagd zal worden over onderwijsmethodes en instructiepraktijken. Daarnaast zijn studies mogelijk op kleinere schaal in scholen zelf, aan de hand van observaties en interviews. Naast het Steunpunt wordt ook flankerend onderzoek opgezet over de genoemde motivatie van leerlingen en het kaderen van de toetsen door de scholen.

Lien Willem erkent dat er inderdaad geen eindtermen zijn voor leerlingen van het vierde leerjaar lagere school. Voor hen wordt dan ook geen cesuur bepaald en wordt er evenmin aan scholen teruggekoppeld over het behalen ervan. Er zal wel worden gekeken naar vaardigheidsniveaus, maar de concretisering daarvan wordt nog uitgezocht. De collega's van toetsontwikkeling hebben gezocht naar gemeenschappelijke elementen in scholen, omdat er inderdaad verschillen zitten in het pad dat ze afleggen. Dat is ook nageslagen in handboeken en leerplannen, en er is overlegd met AHOVOKS. Het klopt evenwel dat de eindtermen geen duidelijke houvast bieden.

3. Replieken van de leden

Hannelore Goeman blijft erbij dat de timing zorgwekkend is. De kalibratie voor het vierde leerjaar is bezig, terwijl in het licht van de afname volgend schooljaar eigenlijk alles al op punt zou moeten staan. De vraag is ook hoe men leerwinst kan meten vanaf het vierde leerjaar als men daar geen cesuur bepaalt. Volgens haar is het hoog tijd om het politieke debat te voeren, want er zijn beloften gemaakt die niet realistisch zijn, zoals die over het longitudinaal onderzoek in het secundair onderwijs, dat men blijkbaar nog aan het bekijken is en misschien niet eens mogelijk is. Het lid wil zekerheid voor zij over een dergelijk kader stemt.

Johan Danen roept de vraag op wie met de rijke waaier aan data die uit de toetsen voortkomt, aan de slag gaat. Welke ideeën heeft het Steunpunt daarover, wat het geaggregeerde niveau betreft? En hoe worden de scholen voorbereid op de omgang met data op hun niveau? Niet alle scholen hebben immers hetzelfde beleidsvoerend vermogen.

Karolien GROSEMANS,
voorzitter

Loes VANDROMME
Koen DANIÉLS
Elisabeth MEULEMAN,
verslaggevers

Gebruikte afkortingen

AHOVOKS	Agentschap voor Hoger Onderwijs, Volwassenenonderwijs, Kwalificaties en Studietoelagen
DIF	differential item functioning
KU Leuven	Katholieke Universiteit Leuven
LiSO	Loopbanen in het Secundair Onderwijs
LOSO	Longitudinaal Onderzoek Secundair Onderwijs
NAPLAN	National Assessment Program – Literacy and Numeracy
OVSG	Onderwijsvereniging van Steden en Gemeenten
PISA	Programme for International Student Assessment (van de OESO)
UGent	Universiteit Gent
Vlor	Vlaamse Onderwijsraad
VUB	Vrije Universiteit Brussel