

Onderzoeksproject i.o.v. het Vlaamse Ministerie van Onderwijs en Vorming

Validering van IDP en de OVSG-toets: Eindrapport

Rianne Janssen
Daniël Van Nijlen
Lien De Mulder
Eef Ameel

Steunpunt Toetsontwikkeling en Peilingsonderzoek
KU Leuven

januari 2017

Inhoudsopgave

Samenvatting en beleidsaanbevelingen.....	4
Samenvatting van de bevindingen	4
Beleidsaanbevelingen.....	5
Inleiding.....	6
Situering onderzoek	6
Rapportering.....	6
Algemeen luik.....	7
Inleiding.....	7
Doelstellingen.....	7
Werkwijze.....	7
Rapportering.....	8
Algemene beschrijving van IDP en OVSG.....	9
Team van toetsontwikkelaars	9
Proces van toetsontwikkeling.....	9
Toetsinstrumentarium.....	9
Impact.....	11
Opstellen van normgegevens.....	11
Feedbackrapporten	12
Ondersteuning van de scholen.....	12
Reflecties vanuit de koepels.....	13
Evaluatie volgens het RCEC-beoordelingssysteem.....	13
Overzicht van de oordelen	13
Criterium 1: Doel en gebruik	13
Criterium 2: Toets- en examenmateriaal	14
Criterium 3: Representativiteit.....	15
Criterium 4: Betrouwbaarheid	17
Criterium 5: Standaardbepaling en normhandhaving.....	17
Criterium 6: Afname en beveiliging.....	18
Conclusies.....	18
Algemeen.....	18
Aandachtspunten vanuit statistische hoek	18
Beperkingen van dit onderzoeksluik	19
Gebruikersonderzoek.....	20
Inleiding.....	20
Doelstelling en werkwijze.....	20
Rapportering.....	20
Disclaimer	20
Bevindingen.....	21
Deelname	21
Organisatie	21
Gebruik van de toetsen op leerlingniveau	21
Gebruik van de toetsen op schoolniveau	22
Conclusies.....	24
Workshop praktische proeven	25
Beschrijving en rapportering	25
Conclusies.....	25
Specifieke luik.....	26
Inleiding.....	26

Doelstelling.....	26
Werkwijze.....	26
Rapportering.....	26
Bevindingen.....	27
Afnamevorm.....	27
Vraagvormen.....	27
Itemstatistieken.....	27
Beheersingstoetsen.....	27
Scoregroepanalyse.....	27
Gebruik van de toetsen op leerlingenniveau.....	27
Gebruik van de toetsen op schoolniveau.....	28
Gebruik van de OVSG-toets binnen GO!.....	28
Conclusies.....	28
Reflecties en aanbevelingen i.v.m. de ‘toolkit met gevalideerde toetsen’.....	29
Beknopte toelichting van de toolkit.....	29
Het toetsen van alle leerlingen.....	29
Welke scholen nog niet?.....	29
Alle leerlingen binnen een school.....	30
Welk soort toetsen?.....	30
Het gebruik van de toetsen op schoolniveau.....	30
Het gebruik van de toetsen op leerlingniveau.....	30
Toolkit gevalideerde toetsen.....	31
Referenties.....	32

Samenvatting en beleidsaanbevelingen

Samenvatting van de bevindingen

1. In juni 2016 namen bijna 9 op de 10 Vlaamse basisscholen deel aan IDP of de OVSG-toets met hun leerlingen van het zesde leerjaar. Specifiek nam 85 procent van de Vlaamse katholieke basisscholen deel aan IDP. Aan de OVSG-toets nam 98 procent van de OVSG-scholen en 92 procent van de basisscholen uit het GO! deel.
2. De IDP en de OVSG-toets zijn beheersingstoetsen die elk hun eigen invulling en traditie kennen. IDP biedt schriftelijke toetsen aan in de leergebieden wiskunde, Nederlands en wereldoriëntatie. De scholen kunnen deze toetsen ook online afnemen. Voor Nederlands en wereldoriëntatie is er ook een praktische proef. De OVSG-toets beoogt een 'allen-talenten-toets' te zijn, waarbij er schriftelijke toetsen en praktische proeven zijn voorzien voor alle leergebieden en domeinen waarvoor in het basisonderwijs eindtermen zijn geformuleerd.
3. Beide toetsinstrumentaria bieden snel na de afname van de schriftelijke toetsen schoolfeedback aan de deelnemende scholen. Daarbij wordt voor beide toetsinstrumentaria het school- en klasgemiddelde op elke toets vergeleken met het Vlaamse gemiddelde en met het gemiddelde van een referentiegroep van scholen, die een vergelijkbaar leerlingenpubliek kennen op het vlak van de onderwijskansenindicatoren.
4. De scholen vormen de doelgroep van IDP. Het gebruiksdoel is dat scholen IDP gebruiken voor zelfevaluatie in het kader van hun interne kwaliteitsontwikkeling. De OVSG-toets sluit aan bij dit gebruiksdoel op schoolniveau, maar heeft ook de leerlingen als doelgroep. De OVSG-toets wil een bijkomende bron van informatie zijn op leerlingniveau, naast de schooleigen informatie. Voor beide onderwijsverstrekkers ligt een bijkomend gebruiksdoel op het niveau van de koepels. Ze willen aan de hand van de toetsen wegen op de onderwijspraktijk door de concretisatie van leerplandoelen te illustreren en door het onder de aandacht brengen van bepaalde toetsbevindingen.
5. De communicatie vanuit de onderwijsverstrekkers maakt een verschil in het effectief gebruik van de toetsen door de scholen. In het gebruikersonderzoek, waarin een beperkte steekproef van directies bevestigd werd, geven nagenoeg alle bevestigde scholen aan dat hun motivatie om deel te nemen aan de toetsen op schoolniveau ligt. Wat het gebruik op leerlingniveau betreft, geven aan aantal scholen aan dat de leerlingresultaten worden meegedeeld aan de ouders en dat de toetsen mee een rol spelen bij het toekennen van het getuigschrift basisonderwijs naast de schooleigen informatie. Enkel bij de OVSG-scholen is dit een meerderheid van de bevestigde scholen. Ook waren er enkele OVSG-scholen waar de OVSG-toets een high-stakes karakter leek te hebben m.b.t. het al dan niet toekennen van het getuigschrift basisonderwijs. De onderwijsverstrekkers moeten blijven inzetten op hun communicatie dit oneigenlijk gebruik van de toetsen tegen te gaan.
6. De schriftelijke toetsen zijn van voldoende kwaliteit. Dit blijkt uit de kwaliteitszorg tijdens het toetsontwikkelingsproces en daarnaast uit de itemanalyses in het specifieke luik voor de toetsen Nederlands (IDP) en wereldoriëntatie (OVSG-toets). Voor beide toetsinstrumentaria werden enkele specifieke aandachtspunten geformuleerd.
7. Voor de onderzochte toetsen van IDP en de OVSG-toets is de betrouwbaarheid op leerlingniveau enkel voldoende op het niveau van de leergebieden. De betrouwbaarheid op schoolniveau is voldoende op het niveau van de afzonderlijke toetsen. Er zijn geen redenen om te twijfelen aan de veralgemeenbaarheid van deze uitspraken voor de toetsen die niet in het specifieke luik waren opgenomen. Het is belangrijk dat de koepels in de toekomst zelf de betrouwbaarheid staven aan de hand van bijkomende toets- en itemanalyses.

8. Wat betreft de inhoudsvaliditeit worden de toetsen afgestemd op de leerplandoelen. Gegeven de beperkte toetslengte staat de inhoudsvaliditeit noodgedwongen onder druk. Voor elke toets is er evenwel een relevante selectie van leerplandoelen en over de schooljaren heen streven de toetsontwikkelaars naar een brede dekking van de leerplandoelen. Voor de bestudeerde toetsen is er geen systematische vertekening van de selectie van eindtermen binnen een leergebied vastgesteld, zodat de toetsen ook een relevante selectie van de eindtermen beschouwd kunnen worden. Ook in de toekomst, dient per toets de concordantie tussen de items met leerplandoelen en met de eindtermen duidelijk aan de scholen gecommuniceerd te worden, die zo de inhoudelijke selectie binnen een toets mee kunnen nemen bij de interpretatie van de resultaten. Over de jaargangen heen blijven de toetsontwikkelaars streven naar een variatie in de dekking van de eindtermen.

9. Zowel IDP als de OVSG-toets beogen geen absolute meting te geven van de beheersing van de leerplandoelen.. Statistisch gezien kan de relatieve meting op schoolniveau nog verder geoptimaliseerd worden door scholen niet alleen een vergelijking aan te reiken met het Vlaamse gemiddelde, maar ook het scorebereik te verduidelijken en door complexere statistische modellen te gebruiken voor schoolfeedback. Er is geen normhandhaving van de toetsen over verschillende schooljaren heen, wat binnen een relatieve meting ook niet noodzakelijk is.

10. Om de inhoudsvaliditeit van het toetsinstrumentarium te verhogen, wordt gebruik van praktische proeven en dit bij IDP en in nog grotere mate bij de OVSG-toets. De keuze voor deze toetsvorm is noodzakelijk voor het toetsen van bepaalde leerplandoelen (en bijhorende eindtermen). In dit project werden de praktische proeven niet in detail beoordeeld, ook omdat er op dit moment binnen de wetenschappelijke wereld nog geen pasklare psychometrische evaluatie voor praktische proeven is opgesteld. Enkel voor de praktische proeven lichamelijke opvoeding van de OVSG-toets is de relatieve meting grondig uitgewerkt. Het is raadzaam dat de koepels blijven inzetten op de ontwikkeling van praktische proeven.

Beleidsaanbevelingen

1. Zowel IDP als de OVSG-toets worden van voldoende kwaliteit geacht om opgenomen te worden in de 'toolkit met gevalideerde toetsen'.

2. De schriftelijke toetsen kunnen dus gebruikt worden om aan scholen informatie te bieden in het kader van hun interne kwaliteitsontwikkeling. Hoewel de koepels verschillen m.b.t het toevoegen van leerlingen als doelgroep van hun toetsen, wijzen de resultaten in dit onderzoek in de richting dat de toetsresultaten per leergebied ook op leerlingniveau gebruikt kunnen worden. Vanuit wetenschappelijke hoek wordt het gebruik van standaardtoetsen als bijkomende bron van informatie naast de schooleigen informatie aanbevolen voor belangrijke beslissingen op leerlingniveau, zoals het al dan niet toekennen van het getuigschrift basisonderwijs. Ook de praktische proeven kunnen op leerlingniveau worden ingezet in het kader van een brede evaluatie.

3. In navolging van wat het RCEC-beoordelingsinstrument aangeeft, kan deze erkenning als gevalideerde toetsen gelden voor vijf jaar, waarbij de koepels in de tussentijd de kwaliteit van de nieuw ontwikkelde toetsen jaar na jaar systematisch opvolgen.

Inleiding

Situering onderzoek

In Vlaanderen bieden twee koepels toetsen aan die de leerprestaties van kinderen in het basisonderwijs in kaart brengen. Katholiek Onderwijs Vlaanderen organiseert de interdiocesane proeven (IDP), die worden afgenomen in het vierde en het zesde leerjaar. Bij OVSG gaat het om toetsen die bestemd zijn voor leerlingen van het zesde leerjaar. Op vraag van de overheid beoordeelde het Steunpunt Toetsontwikkeling en Peilingen de toetsen voor het zesde leerjaar op hun kwaliteiten. Daarbij werd gekeken naar de toetsafnames van juni 2016.

Het onderzoek bestond uit twee luiken. In het algemene luik werd de ontwikkelings- en aanwendingscyclus van IDP¹ en de OVSG-toets² doorgelicht vanuit wetenschappelijk oogpunt, rekening houdend met de doelen van de instrumenten. Daarbij werd ook een korte bevraging gehouden bij 20 schoolhoofden over het gebruik van de toetsen in de scholen. Ook werd een workshop georganiseerd over principes en problemen van praktische proeven. In het specifieke luik werden psychometrische analyses uitgevoerd op de toetsen van het leergebied Nederlands van IDP en wereldoriëntatie van de OVSG-toets. Omdat GO! een grote gebruiker is van de OVSG-toets, namen ze met een steekproef van hun scholen ook deel aan het gebruikersonderzoek en aan het specifieke luik.

Rapportering

In dit overkoepelend rapport wordt een samenvatting geboden van de bevindingen uit het algemene en specifieke luik. Daarnaast worden ook een aantal beleidsadviezen geformuleerd. De resultaten uit het algemene en specifieke luik werden ook teruggekoppeld aan Katholiek Onderwijs Vlaanderen, OVSG en deels ook aan het GO!.

In de bijlagen zijn de presentaties opgenomen van de workshop over de praktische proeven en van de mondelinge rapporteringen van de resultaten aan IDP, OVSG en GO!. Tot slot worden ook de bestanden met de analyses uit het specifieke luik en met de concordantie tussen de toetsen en de eindtermen opgeleverd, evenals de bestanden met de geanonimiseerde antwoorden van de scholen uit het gebruikersonderzoek.

¹ In de rapportering wordt de naamgeving van Katholiek Onderwijs Vlaanderen overgenomen die IDP enkel als acroniem gebruikt en die daarmee dus in het enkelvoud refereert naar de verschillende toetsen uit het IDP-instrumentarium.

² Ook dit is de naamgeving van de koepel zelf, waarbij de OVSG-toets een paraplueterm is voor alle toetsen die daaronder vallen.

Algemeen luik

Inleiding

Doelstellingen

In het eerste deel van het algemene luik van het valideringsonderzoek werd de ontwikkelings- en aanwendingscyclus van IDP en de OVSG-toets van het schooljaar 2015-'16 bekeken vanuit wetenschappelijk oogpunt, rekening houdend met de vooropgezette doelen van het instrument. De focus lag hierbij op de algemene aanpak en was dus onafhankelijk van een specifieke toets in een specifiek leergebied. Er werden twee resultaten van dit algemene onderzoek vooropgesteld: (a) feedback over de gevolgde procedures van ontwikkeling en aanwending en (b) suggesties ter verbetering van deze procedure.

Werkwijze

Om dit te realiseren werd gebruik gemaakt van een bezoek ter plaatse bij elk van beide koepels waarin de onderzoekers op basis van een reeks richtvragen (zie Tabel 1) met het team van toetsontwikkelaars praatte. In de marge van dit bezoek werd heel wat informatie uitgewisseld.

Tabel 1

Richtvragen voor het eerste bezoek aan het ontwikkelteam

Onderwerp	Richtvragen
Toetskader	<ul style="list-style-type: none">- Wat is het doel van de toetsen?- Welke toetsen worden aangeboden?- Wat zijn de sterke punten van de toetsen, waar is verbetering mogelijk?- Welke leerplandoelen worden getoetst?- Wat is de band met de eindtermen? In welke mate omvat de toets de breedte van de eindtermen (voor dat bepaald leergebied)?- Wat is het afnamesdesign (welke combinaties van toetsen worden afgenomen? in hoeveel scholen, bij hoeveel leerlingen?)?- Hoe wordt er gestreefd naar vergelijkbaarheid over de jaren heen?
Tijdspad	<ul style="list-style-type: none">- Welk tijdspad wordt er gevolgd bij de ontwikkeling van een nieuwe toetsjaargang?
Toetsontwikkeling	<ul style="list-style-type: none">- Wordt er gebruik gemaakt van een toetsmatrijs? Wordt er rekening gehouden met bepaalde itemkenmerken die beschouwd worden als determinanten van de moeilijkheidsgraad?- Hoe worden de items opgesteld (proces)?- Wie is hierbij betrokken en met welke expertise?- Wat zijn interne regels en aandachtspunten hierbij?- Is er een vorm van intervisie tussen de toetsontwikkelaars?- Hoe worden bij meerkeuzevragen de antwoordalternatieven (juiste antwoord + afleiders) opgesteld? Wat zijn de sterke punten in de formulering van de antwoordalternatieven, waar is verbetering mogelijk?- Is er een itembank beschikbaar met de items van voorgaande jaargangen? Welke itemkenmerken worden daarin meegenomen?
Vooronderzoek	<ul style="list-style-type: none">- Gebeurt er voor de grootschalige afname een screening van de kwaliteit van de ontwikkelde items?- Zijn er pilootafnames bij leerlingen?- Wordt er gebruik gemaakt van expertbeoordelingen?- Hoe wordt de informatie uit een eventueel vooronderzoek verwerkt?
Samenstelling toetsen	<ul style="list-style-type: none">- Hoe worden de toetsen samengesteld?- Wordt er een cesuur bepaald?- Hoeveel items worden er per toets voorzien?- Hoe wordt de afnameduur bepaald?

Tabel 1 (vervolg)

Onderwerp	- Richtvragen
Afname	<ul style="list-style-type: none"> - Zijn er duidelijke afname-instructies voor de scholen? - Is er bij open vragen een duidelijke antwoordsleutel voorzien? - Worden er garanties voor het bewaren van de objectiviteit ingebouwd bij afname en bij scoring? - Als er een online en een papieren afname is, hoe worden beide afnamemodaliteiten voorbereid en vergeleken? - Zijn er speciale afnamemodaliteiten voor kinderen met specifieke onderwijsbehoeften? - Hoe en door wie worden de toetsen gescoord?
Analyse resultaten	<ul style="list-style-type: none"> - Wordt er gebruik gemaakt van centraal beschikbare achtergrondgegevens? - Welke analyses worden op de data van de leerlingen en mogelijke achtergrondgegevens uitgevoerd? - Gebeuren er psychometrische analyses op de toetsen? - Hoe worden de resultaten teruggekoppeld aan de toetsontwikkelaars?
Gebruik toetsen door de koepels en voorziene schoolfeedback	<ul style="list-style-type: none"> - Welke informatie wordt op systeem- of koepelniveau geanalyseerd? - Wat is de participatiegraad van de scholen aan de toetsen? - Welke feedback krijgen scholen over de prestaties van hun leerlingen? <ul style="list-style-type: none"> o Is er een uitspraak over het behalen van de leerplandoelen en/of van de eindtermen? o Zijn er punten waarop deze feedback nog geoptimaliseerd kan worden? o Worden er soms bepaalde (groepen) leerlingen uitgesloten van de toetsen?

Daarnaast werden de gebruikte toetsen geëvalueerd aan de hand van een evaluatiesysteem voor de beoordeling van de kwaliteit van toetsen en examens³ dat recent ontwikkeld werd door Sanders e.a. (2016) binnen het Nederlandse Research Center voor Examinering en Certificering (RCEC). Het RCEC-systeem werkt met zes criteria (zie Tabel 2) voor de beoordeling van toetsen en examens. Deze criteria sluiten eerder aan bij een product- dan een procesevaluatie. De focus van de evaluatie van IDP en de OVSG-toets volgens het beoordelingssysteem lag uitsluitend op de schriftelijke toetsen.

Tabel 2
Overzicht van de criteria uit het RCEC-beoordelingssysteem

criterium	Omschrijving
1	Doel en gebruik
2	Toets- en examenmateriaal
3	Representativiteit
4	Betrouwbaarheid
5	Standaardbepaling en normhandhaving
6	Afname en beveiliging

Rapportering

In wat volgt wordt een samenvatting gegeven van de belangrijkste punten uit de aparte deelrapporten over het algemene luik die voor IDP en de OVSG-toets werden opgesteld. Op vraag van de opdrachtgever wordt de informatie voor beide toetsinstrumentaria daarbij naast mekaar geplaatst.

³ Te verkrijgen via <http://www.rcec.nl/beoordelingssysteem/>

Algemene beschrijving van IDP en OVSG

Team van toetsontwikkelaars

In beide koepels wordt het toetsinstrumentarium voor het einde van het zesde leerjaar opgesteld door een team van toetsontwikkelaars die dit naast hun andere taken behartigen. De aansturing en organisatorische structuur is licht verschillend tussen beide koepels. Voor IDP is er een kernteam van vier personen dat bijgestaan wordt door werkgroepen die bestaan uit pedagogisch begeleiders uit de regio's, (zorg)leerkrachten, beleidsondersteuners en lerarenopleiders, die allen op vrijwillige basis deelnemen. Bij OVSG werken een twintigtal pedagogisch begeleiders aan de ontwikkeling van de OVSG-toets. Er is een coördinator voor het inhoudelijke gedeelte en voor alle organisatorische en logistieke processen. Bij beide koepels is er daarnaast ook administratieve ondersteuning en is er ook een ICT-medewerker met digitaal-organisatorische taken.

Proces van toetsontwikkeling

De ontwikkeling van de toetsen die in juni 2016 werden afgenomen, startte bij beide koepels al ruim vooraf. Voor IDP duurt het proces van toetsontwikkeling twee jaar, voor de OVSG-toets is dat anderhalf jaar. Voor beide instrumentaria wordt eerst een thema vastgelegd dat voor een deel de inhoud van de toetsen bepaalt. Bij de OVSG-toets wordt binnen dat thema ook eerst een bronnenboek opgesteld dat leerlingen bij het oplossen van bepaalde toetsvragen moeten raadplegen. Dit bronnenboek bestond in 2016 uit 12 rijk geïllustreerde pagina's met teksten van uiteenlopende teksttypes (o.a. informatieve teksten, interviews, tabellen en grafieken en een korte woordenlijst Frans).

Bij beide koepels worden ontwikkelde vragen uitgebreid voorgelegd aan collega's. De OVSG-toets wordt ook voorgelegd aan de pedagogische begeleidingsdienst van het GO!, aangezien de scholen van het GO! een grote gebruikersgroep vormen van de OVSG-toets. Ook zijn er bij beide koepels voor elke schriftelijke toets pilootafnames in enkele klassen van leerlingen. Daarnaast kent IDP ook een grote pilootafname van nieuw ontwikkelde vragen. De scholen die online deelnemen aan IDP moeten in mei een proefafname organiseren om de leerlingen vertrouwd te laten worden met de digitale toetsomgeving en om eventuele ICT-problemen bij de eigenlijke toetsafname te vermijden. Bij die proefafnames worden de nieuwe toetsvragen voor de IDP-toetsen van het volgende schooljaar uitgetest. De praktische proeven worden bij beide koepels in verschillende try-outs op punt gesteld.

Toetsinstrumentarium

IDP

Tabel 3 biedt een overzicht van de toetsen die binnen IDP werden aangeboden aan de scholen in het schooljaar 2015-'16 voor het zesde leerjaar⁴. Voor wiskunde en Nederlands werd IDP afgenomen op twee halve dagen (met in totaal vier delen) naar keuze tussen 6 en 17 juni 2016. De scholen konden kiezen voor een online afname of een afname op papier. Voor wereldoriëntatie was een online omnibustoets⁵ voorzien van een lestijd waarin alle bestaansdimensies aan bod komen. De vragen uit de schriftelijke toetsen zijn voor het merendeel meerkeuzevragen, maar daarnaast zijn er ook korte, open vragen, vooral bij wiskunde en wereldoriëntatie en natuurlijk ook bij het kerndictee voor Nederlands. IDP bevatte twee praktische proeven. De observatieopdracht voor wereldoriëntatie en de spreekopdracht voor Nederlands waren al vanaf februari beschikbaar voor individuele afname, waarbij beide opdrachten eventueel geïntegreerd konden worden.

⁴ IDP wordt ook georganiseerd voor het vierde leerjaar.

⁵ Een omnibustoets is een toets die heterogeen is van inhoud omdat ze gaat over een ruim domein dat niet verder werd opgesplitst in deeldomeinen (en deoltoetsen).

Tabel 3
Overzicht van IDP van juni 2016

Leergebied	Toetsen	Aantal items	Afname
Wiskunde	Getallenkennis	10	Deel 1
	Bewerkingen:		
	hoofdrekenen	10	Deel 1
	zonder zakrekenmachine	5	Deel 1
	met zakrekenmachine	5	Deel 3
	Meten en metend rekenen	10	Deel 3
Nederlands	Meetskunde	10	Deel 3
	Spelling	20	Deel 2
	Luisteren	10	Deel 2
	Begrijpend lezen	10	Deel 4
	Taalbeschouwing	10	Deel 4
	Spreekproef	--	vanaf februari
Wereldoriëntatie	Online proef	30	één lestijd
	Observatieopdracht	--	vanaf februari

OVSG-toets

De OVSG-toets beoogt een ‘alle-talenten-toets’ aan te bieden aan de scholen voor hun leerlingen van het zesde leerjaar. Daarbij worden zowel praktische proeven (zie Tabel 4) als pen-en-papier toetsen (zie Tabel 5) ontwikkeld voor alle leergebieden en domeinen waarvoor in het basisonderwijs eindtermen zijn geformuleerd. Daarnaast wordt voor muzische vorming jaarlijks één domein getoetst. Het is de bedoeling dat de praktische proeven tijdens het schooljaar al worden afgenomen. De praktische proeven voor Frans worden pas later vrijgegeven om zo voldoende onderwijstijd voor alle leerlingen te voorzien. De schriftelijke toetsen werden afgenomen op maandag 20, dinsdag 21 en woensdag 22 juni telkens in de voormiddag. Elke toetsvoormiddag was verdeeld in twee delen met tussenin een pauze. In totaal zijn er dus zes delen, waarvoor telkens een aparte toetsbundel is voorzien. De leerlingen startten de eerste dag met een verkenning van het bronnenboek. In vergelijking met IDP telt de OVSG-toets minder meerkeuzevragen. Op basis van de schriftelijke en de praktische proeven samen krijgt elke leerling een

Tabel 4
Overzicht van de praktische proeven in de OVSG-toets van 2016

Leergebied	Toetsen	Afname vanaf	Evaluatie
Nederlands	Luisteren	november	10 punten ^a
	Spreken	november	10 punten
Frans	Spreken	april	10 punten
	Mondelinge interactie	april	10 punten
Wetenschappen en Techniek	Constructie Techniek	november	15 punten
Mens en maatschappij	Fietsexamen	november	Brevet
Muzische vorming	Beweging (flashmob)	november	Zelfevaluatie
Lichamelijke opvoeding ^b	Zwemmen	november	Score
	Atletiek	november	Score
	Gymnastiek	november	Score
	Rollend en glijdend materiaal	november	Score
	Ritmiek en expressie	november	Score
	Balvaardigheid en sportspel	november	Score

^a Ook zelfevaluatie. ^b De leerlingen zelf konden een brevet van ‘Olympisch sporter’ halen.

Tabel 5

Overzicht van de schriftelijke toetsen in de OVSG-toets van 2016

Leergebied	Toetsen	Werkbundel	Richttijd	Vragen	Deelvragen	Meerkeuze	Bronnenboek	Punten
Wiskunde	Getal­len­kennis	1	50'	18	4	1	1	20
	Hoofdrekenen	2	25'	10	0	1	2	10
	Cijferen	5	50'	10	0	2	1	10
	Meetkunde	5	50'	10	6	3	0	10
	Meten	6	50'	10	1	0	4	10
Nederlands	Schrijven	2	100'	6	5	0	2	10
	met schrijfo­p­dracht				4 stappen			
	Taal­be­schou­wing	3	30'	8	5	1	2	10
Wereldoriëntatie	Lezen	4	30'	8	3	3	7	10
	Maatschappij	1	30'	10	0	5	3	10
	Ruimte	3	30'	10 ^a	4	3	3	10
	Natuur	4	30'	10	5	3	3	10
	Tijd	6	40'	10	2	5	6	10
Frans	Techniek	6	15'	5	0	1	1	5
	Luisteren	3	40'	6 ^b	5	5	1	10
	Schrijven	4	30'	7	4	0	0	10
	Lezen	5	30'	4 ^b	3	6	1	10

^aInclusief één doe-opdracht (globe).

^bEen itemnummer verwijst hier naar een reeks van items die horen bij een bepaalde tekst.

score op vijf toetsen voor de leergebieden wiskunde, Nederlands, wereldoriëntatie en Frans. Elke toets wordt, eventueel na omzetting, gescoord met een maximum totaal van 10 punten.

Impact

In totaal namen in het schooljaar 2015- 1096 scholen deel aan IDP voor het zesde leerjaar. IDP bereikt daarmee 85 percent van de scholen uit het Katholieke Onderwijs Vlaanderen.

In totaal namen 857 scholen deel aan de OVSG-toets: 483 gemeentelijke scholen, 354 scholen uit het gemeenschapsonderwijs en 20 andere scholen. Dit komt overeen met 98,4 percent van de OVSG-scholen en 92 percent van het aantal lagere scholen binnen het GO!. Ook 13 scholen uit het buitengewoon onderwijs van OVSG namen deel aan (delen) van de OVSG-toets met een of meerdere leerlingen met een handelingsplan dat nauw aansluit bij het gemeenschappelijk curriculum in het gewone basisonderwijs. De leerlingen volgen meestal les in Type 8/basisaanbod, Type 3, Type 9 en 7.

Opstellen van normgegevens

Voor het opstellen van normgegevens heeft IDP de data uit de digitale toetsafnames ter beschikking. De scholen die op papier deelnemen worden gevraagd om de antwoorden van de leerlingen op elke vraag na de afname via een online database door te geven. Ook aan de scholen die deelnamen aan de OVSG-toets

werd gevraagd om na de afname van de toetsen de leerlingresultaten door te geven, maar dan niet de antwoorden op de items, maar wel de behaalde scores op de toetsen.

Naast de berekening van het Vlaamse gemiddelde voor verschillende variabelen, berekent men zowel bij IDP als bij OVSG ook het gemiddelde voor verschillende groepen van scholen, die worden onderscheiden in functie van de onderwijskansenindicatoren van hun leerlingen. De opdeling van de scholen gebeurt bij beide toetsinstrumenten op basis van een andere samenvattende variabele, namelijk het gemiddelde aantal OKI-criteria per leerling voor IDP en de verhouding van het aantal SES-lestijden t.o.v. het aantal lestijden van de school voor de OVSG-toets. Dit betekent dat telkens verschillen tussen scholen op basis van een continue variabele worden omgezet in een categorische opdeling van de scholen. Bij zowel IDP als de OVSG-toets kunnen scholen ook kiezen om hun resultaten te vergelijken met een andere referentiegroep.

Feedbackrapporten

De schoolfeedbackrapporten waren vanaf 24 juni beschikbaar voor IDP en vanaf 27 juni voor de OVSG-toets.

De schoolfeedback van IDP is zeer uitgebreid. Naast het rapport voor de school, zijn er ook rapporten per deelnemende klas. De meeste informatie wordt zowel in tabel als in grafiek weergegeven. De feedbackrapporten bevatten de ruwe resultaten per item, per toets en per leergebied met daarbij ook telkens het gemiddelde resultaat in Vlaanderen en in de referentiegroep van scholen. De resultaten worden berekend zonder de leerlingen die redelijke aanpassingen (RA) gebruikten of een individueel aangepast curriculum (IAC) volgden. Een rapport op de website bevat een overzicht van de resultaten van die leerlingen per referentiegroep. De scholen krijgen ook een trendanalyse met hun schoolresultaten per leergebied en per toets voor de afgelopen vier schooljaren. Ook wordt aan elke school per klas een leerlingenlijst toegestuurd met daarin de toetsresultaten, waarbij telkens ook het Vlaamse gemiddelde wordt vermeld.

De scholen die deelnamen aan de OVSG-toets krijgen op een erg overzichtelijke wijze feedback over hun klas- en schoolresultaten door in te loggen op www.ovsgtoetsen.net. Per klas en voor de school in zijn geheel wordt het gemiddelde weergegeven per leergebied en voor elke toets binnen een leergebied voor wiskunde, Nederlands, wereldoriëntatie en Frans. Voor lichamelijke opvoeding worden de gemiddelden gegeven voor de zes activiteitsgebieden, maar zonder totaalresultaat. Telkens krijgen de scholen bij elk resultaat ook het Vlaamse gemiddelde, het gemiddelde van het net waartoe ze behoren (OVSG, GO! of ander) en hun SES-referentiegroep. De webpagina is interactief: scholen kunnen kiezen tussen de resultaten uit de tabel ook laten weergegeven in een grafiek. Ook kunnen scholen de resultaten van de afgelopen zeven jaren opvragen (voor de afname van juni 2016 dus tot en met 2009) en zich vergelijken met een andere referentiegroep. De gemiddelde resultaten van deelnemende scholen uit het buitengewoon basisonderwijs zijn apart opvraagbaar.

Ondersteuning van de scholen

Zowel bij IDP als bij de OVSG-toets kunnen de scholen een beroep doen op een telefonische en online helpdesk.

Scholen die dat wensen kunnen bij beide koepels en ook bij GO! om pedagogische begeleiding vragen bij hun proces van interne kwaliteitsontwikkeling op basis van de resultaten op de toetsen en/of de analyses van het schoolteam. Bij Katholiek Onderwijs Vlaanderen werd daarnaast in het schooljaar 2015-'16 ook een specifiek project opgestart waarbij scholen zich konden inschrijven voor twee schoolsessies, waarbij specifiek met de eigen schoolresultaten aan de slag gegaan wordt. Daarnaast wordt ook jaarlijks een

introductiesessie georganiseerd voor (grote groepen) leerkrachten en directeurs die voor het eerst deelnemen aan IDP.

Reflecties vanuit de koepels

In het najaar verschijnen er artikels over de resultaten van IDP in *School+visie*. Elk jaar wordt een deelgebied in de kijker gezet waarbij er dan commentaren en suggesties worden gegeven.

Ook vanuit OVSG werd in het najaar een rapport uitgegeven waarin werd teruggeblikt op de resultaten van de OVSG-toetsen. Daarbij werd ook een didactische reader opgesteld voor getallenkennis, dat dit jaar het leergebied in de kijker was. Voor dit domein waren er daarom meer vragen opgenomen in de schriftelijke toets. Ook verzamelden de pedagogisch begeleiders de itemantwoorden van 673 leerlingen en werden een zestigtal leerlingen door hen geïnterviewd over hun oplossingsstrategieën. Bij de didactische reader horen ook zes mini-wijzers die betrekking hebben op verschillende aspecten van getallenkennis en die over de gehele basisschool kunnen worden gehanteerd.

Evaluatie volgens het RCEC-beoordelingssysteem

Overzicht van de oordelen

Zowel voor IDP als voor de OVSG-toets werd in het aparte deelrapport over het algemene luik een gedetailleerd overzicht gegeven van de evaluatie van de schriftelijke toetsen volgens elk criterium uit het RCEC-beoordelingssysteem aan de hand van de bijhorende deelvragen. In deze samenvatting worden enkel de belangrijkste punten opgenomen van de evaluatie van beide toetsinstrumentaria.

Criterium 1: Doel en gebruik

1.1. Doelgroep

De scholen zijn voor zowel IDP als voor de OVSG-toets de directe doelgroep. De zesdeklassers die de toetsen afleggen zijn voor IDP slechts een indirecte doelgroep. Voor de OVSG-toets zijn die leerlingen de tweede directe doelgroep van de toetsen.

Zowel bij IDP als bij de OVSG-toets worden de scholen aangemoedigd om de leerlingen die (a) gebruik maken van RA of compenserende maatregelen of (b) een IAC volgen of van dispenserende maatregelen genieten ook te laten deelnemen aan de toetsen. Bij IDP worden de scores van de leerlingen met RA of een IAC apart geteld bij het opstellen van de schoolfeedbackrapporten. OVSG adviseert om de leerlingen die compenserende maatregelen krijgen wel op te nemen bij het doorgeven van de toetsresultaten. De scores van de leerlingen met dispenserende maatregelen worden niet meegenomen bij het opstellen van de schoolfeedbackrapporten.

1.2. Meetdoel

Inhoudelijk is het meetdoel van IDP om zicht te krijgen op de beheersing van een relevante selectie leerplandoelen uit de leergebieden wiskunde, Nederlands en wereldoriëntatie (zie Tabel 3). Inhoudelijk wil de OVSG-toets een alle-talenten-toets zijn waarin alle leergebieden en domeinen uit het basisonderwijs waarvoor er eindtermen zijn geformuleerd met schriftelijke toetsen en praktische proeven worden getoetst (zie Tabellen 4 en 5). De manier waarop de leerplandoelen in beide toetsen worden geselecteerd, wordt besproken bij Criterium 3.

Op schoolniveau loopt het meetdoel van IDP en de OVSG-toets gelijk. Het meetdoel is om per toets en per leergebied de gemiddelde leerlingprestaties vast te stellen met het oog op het geven van informatie over waar de school staat t.o.v. (vergelijkbare) Vlaamse scholen.

Het leerlingniveau wordt niet als meetdoel beschouwd bij IDP. Bij de OVSG-toets is het meetdoel op leerlingniveau om de prestatie van de leerling vast te stellen m.b.t. de verschillende leergebieden uit het basisonderwijs op basis van een relevante selectie van leerplandoelen per toets of per praktische proef.

Voor beide toetsinstrumentaria ligt het meetdoel op het niveau van de toets of het leergebied. Daarnaast worden scholen in het kader van een diepgaande zelfevaluatie voor een bepaald leergebied bij beide toetsinstrumentaria aangemoedigd om ook te kijken naar de resultaten op itemniveau. Daar zijn enkele kanttekeningen bij te plaatsen:

- Bij de OVSG-toets heeft de school geen informatie over de itemmoeilijkheidsgraad voor de gehele groep van deelnemende scholen. Als een item erg moeilijk is voor alle scholen, dan is de diagnostische waarde van een lage itemscore voor een school anders dan wanneer het zou gaan om een item waar de school specifiek op uitvalt. Enkel voor het leergebied in de kijker krijgen de scholen informatie over de moeilijkheidsgraad.
- Bovendien kan de diagnostische waarde van de leerlingprestaties op individuele items principieel in vraag gesteld worden. Elk concreet item heeft immers per definitie een grote mate van specificiteit, waardoor de veralgemeenbaarheid beperkt is. De proportie leerlingen die een item juist maken is ook niet altijd betrouwbaar te meten gegeven het aantal leerlingen binnen een klasgroep. Ten slotte, als een toets intern consistent blijkt (zie het specifieke luik), dan kan verwacht worden dat slechts weinig scholen van het algemene patroon van de itemmoeilijkheidsgraden zullen afwijken.

1.3. Gebruiksdoel

Op schoolniveau hebben IDP en de OVSG-toets hetzelfde gebruiksdoel voor ogen: een instrument bieden voor zelfevaluatie in het kader van de interne kwaliteitsontwikkeling van een school.

IDP neemt het leerlingniveau niet op als gebruiksdoel⁶. Enkel de OVSG-toets heeft een gebruiksdoel op leerlingniveau. Dit wordt in de visietekst over de OVSG-toets⁷ als volgt geformuleerd:

De OVSG-toets is tevens een relevante bevraging van een selectie van leerplandoelen op leerlingniveau. Het is een bron van informatie bij de evaluatie van leerlingen. Dat bij het evalueren van leerlingen de gegevens van de OVSG-toets steeds naast de schooleigen gegevens moeten gelegd worden, is een evidentie. De OVSG-toets is dus geen alleenstaande norm voor het al dan niet toekennen van het getuigschrift basisonderwijs. De OVSG-toets is één element in de globale beoordeling van de leerling op het einde van de basisschool in functie van de oriëntering naar het secundair onderwijs.

Tot slot hebben beide toetsinstrumentaria ook gelijklopende gebruikersdoelen op het niveau van de koepel zelf. Enerzijds wil men met behulp van de eigen toetsen aan de scholen tonen hoe leerplandoelen geconcretiseerd en geëvalueerd kunnen worden. Anderzijds wil men via de toetsen ook zaken onder de aandacht brengen van de scholen via de publicatie van algemene reflecties op de resultaten en voor de OVSG-toets ook door jaarlijks een leergebied speciaal in de kijker te zetten.

Criterion 2: Toets- en examenmateriaal

Bij IDP en de OVSG-toets scoort de kwaliteit van het schriftelijke toetsmateriaal voldoende op de verschillende deelaspecten uit het RCEC-beoordelingsinstrument: (1) gestandaardiseerde vragen; (2)

⁶ Al krijgen de scholen de toetsresultaten op leerlingniveau wel toegestuurd.

⁷ zie hiervoor <http://www.ovsg.be/visieteksten/visietekst-situering-van-de-ovsg-toets>

In het citaat is de cursivering uit het oorspronkelijk document weggelaten.

objectief scoringssysteem en beoordelingsvoorschrift; (3) ontwerp toetsboekjes; (4) ontwerp scoringssysteem; (5) instructies; (6) formulering vragen en (7) lay-out en vormgeving.

Er werden wel enkele aandachtspunten geformuleerd. Bij beide toetsen werd de vraag gesteld naar de kosten-batenanalyse van het gebruik van items met open antwoordvorm in schriftelijke toetsen voor grootschalige toetsafnames. Bij IDP lijkt het raadzaam om systematisch te onderzoeken of er verschillen in moeilijkheidsgraad zijn tussen de vragen uit de papieren versie en de corresponderende vragen uit de digitale afname en waarmee eventuele verschillen samenhangen. Bij de OVSG-toets kon het onderzoeksteam de scoring van de antwoorden van de leerkrachten bekijken bij het intikken van de scores voor de toetsen wereldoriëntatie uit het specifieke luik. Op basis van toevallige observaties werd vastgesteld dat de beoordelingsvoorschriften niet door alle leerkrachten uniform werden geïnterpreteerd. Ook al werd dit eventuele probleem niet systematisch onderzocht binnen dit onderzoek, toch is het in het algemeen een aandachtspunt bij de formulering van de specifieke vraagstelling van items en van hun beoordelingsvoorschriften.

criterium 3: Representativiteit

3.1. Toetsmatrijs afgestemd op meetdoel

Leerplandoelen

Zowel bij IDP als bij de OVSG-toets is elke toets inhoudelijk afgestemd op haar meetdoel. Een positief punt daarbij is de uitgewerkte concordantie tussen de items en de leerplandoelen en eindtermen⁸ bij beide toetsinstrumenten. Deze informatie is telkens ook beschikbaar voor de leerkrachten.

Opvallend is dat de OVSG-toets een aangepaste toetsing voorziet voor alle leergebieden en domeinen waarvoor de overheid eindtermen voorziet.

Per toets wordt een relevante selectie gemaakt van de leerplandoelen die getoetst worden. Bij de selectie van leerplandoelen worden bij IDP drie criteria gehanteerd: (a) de belangrijkheid van het leerplandoel; (b) de keuze om een bepaald leerplandoel in de kijker te zetten op basis van input uit het veld en (c) de mate waarin het leerplandoel inpasbaar is binnen het gekozen thema. Bij de selectie van leerplandoelen voor de OVSG-toets wordt rekening gehouden met twee criteria: variatie en breedte. Over toetsjaargangen heen wordt gestreefd om te variëren in de getoetste leerplandoelen en binnen een toets wordt elk jaar gestreefd om voor elk groot onderdeel van een leergebied minstens een leerplandoel te selecteren. Bij de OVSG-toets worden geen uitbreidingsdoelen getoetst.

Zowel voor IDP als voor de OVSG-toets staat de inhoudsvaliditeit⁹ van elke toets noodgedwongen onder druk omdat elke toets slechts een beperkt aantal items¹⁰ kan bevatten o.w.v. de beperkte afnametijd. Doorgaans kunnen niet alle betreffende leerplandoelen in een toets worden opgenomen en de opgenomen leerplandoelen worden met een beperkt aantal items getoetst. De inhoudsvaliditeit betreft dan niet de volledige dekking van het domein, maar wel of er een representatieve steekproef is van leerplandoelen per toets. In deze laatste betekenis kunnen de toetsen uit IDP en de OVSG-toets als voldoende inhoudsvalide worden beschouwd t.a.v. de leerplannen.

De toetsontwikkelaars maken elk jaar een relevante selectie van leerplandoelen per toets. Het is aangewezen om de keuze van de bevroegde leerplandoelen en hun aandeel ervan in de toets nog meer

⁸ Voor de GO!-scholen wordt de specifieke concordantie van de OVSG-toets met de leerplandoelen door de pedagogische begeleidingsdienst van GO! voorzien.

⁹ De inhoudsvaliditeit is de mate waarin een toets het getoetste domein afdekt. Een toets met een hoge inhoudsvaliditeit bevraagt een leergebied met een representatieve steekproef van leerdoelstellingen met een ruim aantal items.

¹⁰ Dit heeft ook zijn weerslag op de betrouwbaarheid van de toets.

systematisch aan te pakken via het (op voorhand) opmaken van een toetsmatrijs. Het is belangrijk om deze informatie over schooljaren heen te monitoren.

Eindtermen

Hoewel de inhoudsvaliditeit van beide toetsinstrumentaria primair gericht is op de leerplandoelen, werd in dit onderzoeksproject ook nagegaan in welke mate de toetsinstrumentaria inhoudsvaliditeit hebben t.a.v. de eindtermen. In principe zou dit kunnen worden afgeleid op basis van een transitiviteitsargument, namelijk dat men de inhoudsvaliditeit t.a.v. de eindtermen kan veronderstellen omdat de toetsen inhoudsvalide zijn t.a.v. de leerplandoelen en de leerplandoelen goedgekeurd zijn wat betreft hun dekking van de eindtermen.

Daarnaast kan de inhoudsvaliditeit ook empirisch ondersteund worden omdat is onderzocht in welke mate de toetsen van juni 2016 de eindtermen in kaart brengen. Daarbij werd in een Excel-tabel alle eindtermen van de getoetste leergebieden opgelijst en werd aangeduid welke eindtermen in elke toets aan bod kwamen op basis van de gegevens uit de concordantietabellen die door de koepels worden voorzien. Deze Excel-tabellen voor IDP en de OVSG-toets zijn toegevoegd als bijlage. Net als bij de beoordeling van de inhoudsvaliditeit t.a.v. de leerplandoelen moet worden opgemerkt dat het onmogelijk is om te komen tot een volledige dekking van de eindtermen. Er moet noodzakelijkerwijs worden gewerkt met een steekproef van eindtermen binnen elke toets. Op basis van de gemaakte analyses kon alvast geen systematische vertekening worden vastgesteld in de feitelijke selectie van eindtermen per toets, noch voor IDP, noch voor de OVSG-toets. Beide toetsen kunnen dus beschouwd worden als een relevante selectie van de eindtermen. Ook hier is het aangewezen om ook deze dekking t.a.v. de eindtermen over toetsjaargangen heen te monitoren en deze bijkomende informatie te blijven delen met de toetsgebruikers, zodat die bij de interpretatie van de resultaten op een bepaalde toets precies weten welke eindtermen wel en niet betrokken zijn.

3.2. Moeilijkheidsgraad afgestemd op doelgroep

Bij zowel IDP als bij de OVSG-toets wordt de moeilijkheidsgraad van de nieuw ontwikkelde items ingeschat op basis van pilootafnames, al gebeurt dit bij IDP meer systematisch en op duidelijk grotere schaal.

Bij IDP heeft men op basis van de digitale toetsafname en de doorgegeven leerlingantwoorden voor de papieren toetsen een uitgebreid zicht op de itemmoeilijkheidsgraden en op de keuze van de afleiders bij meerkeuzevragen.

Voor de OVSG-toets wordt de moeilijkheidsgraad op itemniveau van de uiteindelijke toetsen niet bepaald, behalve voor het leergebied dat in de kijker staat. Het is nodig om deze informatie voor alle toetsen te bepalen. Dit is ook zinvol voor de evaluatiecriteria van de praktische proeven.

Bij zowel IDP als bij de OVSG-toets wordt impliciet gewerkt met een richtnorm van .70 voor de gemiddelde moeilijkheidsgraad van de toetsitems. Deze keuze ligt in lijn met de keuze voor een beheersingstoets¹¹ i.p.v. voor een vaardigheidstoets. Deze laatste keuze sluit aan bij de vooropgestelde meetdoelen.

¹¹ Een beheersingstoets gaat na of leerlingen bepaalde leerstof onder de knie hebben. Een vaardigheidstoets bevat daarentegen ook opgaven die makkelijker of moeilijker zijn dan de eigenlijke leerstof om zo verschillen tussen leerlingen duidelijker in beeld te brengen.

Criterion 4: Betrouwbaarheid

Er worden tot nu toe noch door het IDP-team, noch door het OVSG-team analyses uitgevoerd voor het bepalen van de toetsbetrouwbaarheid en de onderlinge samenhang van de items. In tegenstelling tot de situatie bij IDP, heeft OVSG de data hiervoor ook niet ter beschikking.

Het is nodig om dit soort analyses in te bouwen bij de toekomstige toetsontwikkelingen voor IDP en de OVSG-toets. Zoals in het specifieke luik werd geïllustreerd geeft de betrouwbaarheid de contouren aan waarbinnen de toetsen als meetinstrumenten ingezet kunnen worden. Een gedetailleerde itemanalyse geeft interessante feedback aan de toetsontwikkelaars, die kan helpen om de itemkwaliteit van de toetsen verder te optimaliseren. Deze itemanalyses kunnen ook al worden ingezet na pilootafnames bij de selectie van de uiteindelijke toetsopgaven.

Criterion 5: Standaardbepaling en normhandhaving

Relatieve meting

Noch bij IDP, noch bij de OVSG-toets wordt er een cesuur vastgelegd om te bepalen wat de minimale toetsprestatie is om van een voldoende beheersing van de getoetste leerplandoelen te spreken. Er is dus geen sprake van een absolute meting.

Op schoolniveau beogen beide toetsinstrumentaria een relatieve meting van hoe goed de school het doet in vergelijking met het Vlaamse gemiddelde en in vergelijking met scholen met een vergelijkbaar leerlingpubliek. Bij deze relatieve meting kunnen enkele kanttekeningen worden geplaatst:

- Er wordt aan de scholen enkel het gemiddelde van beide referentiegroepen gegeven. Er zou over gedacht kunnen worden om bijkomende informatie te geven over de scoreverdeling in de normgroepen, bijvoorbeeld in een rupsgrafiek, waarbij scholen ook kunnen afleiden of hun resultaten statistisch verschillen van het gemiddelde.
- De vergelijking met vergelijkbare scholen gebeurt door de scholen eerst in te delen in referentiegroepen (cf. supra). Daarna wordt per groep het gemiddelde toetsresultaat gerapporteerd. Scholen kunnen zich ook vergelijken met het gemiddelde van andere groepen. De werkwijze dat scholen zich kunnen vergelijken met andere scholen uit hun groep is een werkwijze die voor scholen waarschijnlijk goed begrijpbaar is. Toch is in het algemeen zo'n groepsindeling in zekere zin arbitrair omdat een continue variabele wordt opgesplitst in afzonderlijke categorieën. Dit houdt bijvoorbeeld in dat scholen die binnen de ene categorie vallen, niet zoveel verschillen van scholen die net in de andere categorie passen.
- Er zijn ook betere statistische methoden beschikbaar om scholen te vergelijken na controle voor verschillen in hun leerlingpubliek. Toegegeven, deze multiniveau-analyses vragen een stevige statistische expertise, die momenteel niet voorhanden is binnen de teams van toetsontwikkelaars. De koepels zouden hiervoor kunnen samenwerken met andere instanties.
- De gebruikte classificatiemethode om scholen in te delen in groepen bij de OVSG-toets is statistisch niet helemaal goed onderbouwd.

Op leerlingniveau wordt bij de OVSG-toets geen absolute, maar ook geen relatieve meting beoogd, al zou de leerkracht de resultaten van een leerling kunnen vergelijken met het schoolgemiddelde of het Vlaamse gemiddelde. Dit laatste is trouwens de vergelijkende informatie die IDP toevoegt aan de leerlinglijsten met resultaten, ook al is het leerlingniveau daar geen meet- of gebruiksdoel. Als het leerlingniveau toch een meetdoel is of wordt, dan valt het te overwegen om de informatiewaarde op leerlingniveau te verhogen door meer informatie te geven over de scoreverdeling en de positie van individuele leerlingen daarin. Zoals in leerlingvolgsystemen zou men bij zo'n relatieve meting kunnen werken met vijf niveaugroepen (A, B, C, D, E) die onderling afgegrensd zijn op basis van respectievelijk percentielen 75, 50, 25, en 15.

Normhandhaving

Behalve door het volgen van een gelijkaardige procedure van itemselectie wordt in het huidige design van beide toetsinstrumentaria geen waarborgen ingebouwd opdat de toetsen jaar na jaar vergelijkbaar zijn qua moeilijkheidsgraad. Op zich is dat niet strikt noodzakelijk voor een relatieve meting. Om te komen tot een normhandhaving zou er een gedeeltelijke itemoverlap tussen de toetsen uit verschillende jaren nodig zijn.

Bij de OVSG-toets kunnen de scholen wel zelf hun schoolresultaten en de resultaten van de normgroepen opvragen voor de voorbije zeven schooljaren. Bij IDP krijgen de scholen de resultaten van de voorbije vier schooljaren van hun school. Vanaf volgend schooljaar zouden daarbij ook de gegevens van het Vlaamse gemiddelde worden toegevoegd. De relatieve informatie over de trends in Vlaanderen is belangrijke informatie opdat scholen hun eigen evolutie over de tijd heen correct kunnen interpreteren.

Criterium 6: Afname en beveiliging

De voorschriften voor de afname en de beveiliging van IDP en de OVSG-toets zijn goed. Er zijn daarbij geen op- of aanmerkingen. In dit onderzoek is de feitelijke afnameprocedure binnen scholen niet onderzocht.

Conclusies

Algemeen

Binnen het algemene luik van dit valideringsonderzoek heeft het onderzoeksteam met veel waardering kennisgemaakt met IDP en de OVSG-toets. Beide teams van toetsontwikkelaars doorlopen een gericht proces van toetsontwikkeling en hebben een duidelijke visie op het bedoelde gebruik van hun toetsen. Ze slagen er beiden in om een heel groot deel van hun scholen hiervoor te engageren. De scholen krijgen daarbij niet alleen inhoudelijk up-to-date toetsen, maar worden ook snel na de toetsafnames van schoolfeedback te voorzien. Dit alles gebeurt met relatief beperkte middelen. Beide koepels hebben ook wel elk hun eigen traditie en elk toetsinstrumentarium kent zijn sterke en minder sterke kanten.

Opvallend bij IDP zijn (a) de organisatie van digitale toetsafnames en de mogelijkheden die dat biedt voor het verzamelen van informatie op itemniveau, (b) het systematische proces van toetsontwikkeling met een grootschalige pilootafname die knap is ingeschakeld in het proces van toetsontwikkeling én in de digitale proefafname van IDP bij de scholen, (c) de uitgebreide schoolfeedback en (d) sinds kort de inschakeling van praktische proeven.

Opvallend bij de OVSG-toets zijn (a) de breedte van de OVSG-toets m.b.t. de getoetste leergebieden, (b) het grote aantal praktische proeven om competentiegerichte leerplandoelstellingen te evalueren, (c) de overzichtelijke schoolfeedback en (d) het bijkomende inhoudelijke onderzoek van de toets getallenkennis.

Aandachtspunten vanuit statistische hoek

Beide toetsinstrumentaria kregen ook een goede beoordeling voor verschillende kwaliteitseisen die voor examens en toetsen in het onderwijs worden verwacht. Niet onverwacht kwamen de aandachtspunten die op basis van het RCEC-beoordelingsinstrument werden geformuleerd vooral vanuit psychometrische en statistische hoek, zowel op het vlak van het opvolgen van de toetskwaliteit (waaronder betrouwbaarheid) als op het vlak van het geven van schoolfeedback. Het lijkt dat de koepels daar op dit moment ook niet de nodige expertise voor in huis hebben, wat niet verwonderlijk is gegeven dat er hier geen specifieke financiële middelen tegenover staan.

Het opvolgen van de psychometrische kwaliteiten van hun toetsen zou op korte termijn mogelijk wel bijkomend opgenomen kunnen worden door de koepels zelf. Ze zouden daarvoor een beroep kunnen doen op de standaardanalyses binnen statistische pakketten zoals SPSS. Daarnaast kan overwogen

worden om de Excel-tool voor toets- en itemanalyses die door het onderzoeksteam ontwikkeld wordt aan de koepels ter beschikking te stellen. Deze tool biedt immers een visuele itemanalyse op basis van een scoregroepanalyse die, afgaand op hun reacties tijdens de voorstelling van de resultaten tijdens het tweede bezoek ter plaatse, erg toegankelijk is voor de toetsontwikkelaars. Deze tool is echter niet geheel gebruiksklaar omdat die in principe alleen ontworpen is voor toetsen met louter meerkeuzevragen. Om deze toets- en itemanalyses te realiseren voor de OVSG-toets moet er dan wel een manier gevonden worden om ten minste voor een deelsteekproef van leerlingen de resultaten op itemniveau te verzamelen.

Voor het optimaliseren van de schoolfeedback zijn meer complexe statistische modellen nodig. Deze multiniveaumodellen zouden enerzijds een ondersteuning kunnen bieden voor de huidige, rudimentaire, maar op zich correcte schoolfeedback. Nu wordt aan de scholen een vergelijking geboden met het Vlaamse gemiddelde en met het gemiddelde van een referentiegroep. Met behulp van multiniveaumodellen zou bepaald kunnen worden hoe groot de schoolverschillen zijn en hoeveel van de variantie op schoolniveau gevat kan worden met de indeling van de scholen in referentiegroepen. Multiniveaumodellen zouden ook rupsgrafieken kunnen bieden waarbij de scholen een idee krijgen over het scorebereik en of hun schoolresultaten statistisch significant verschillen van het gemiddelde.

De multiniveaumodellen bieden anderzijds ook een verfijning van de huidige schoolfeedback. Deze modellen kunnen immers scholen vergelijken na controle voor verschillen in hun leerlingenpubliek op het vlak van verschillende criteria en specifiek voor de concrete leerlingen die deelnamen aan de toetsen i.p.v. op basis van een samenvattende statistiek van de leerlingenpopulatie van de school. Deze modellen geven dus meer correcte informatie aan scholen en zijn daarom in principe te verkiezen boven de groepsvergelijkende vormen van schoolfeedback. Er kan echter niet verwacht worden dat deze expertise op korte termijn aanwezig is binnen de teams van toetsontwikkelaars. Misschien kan overwogen worden dat er hiervoor wordt samengewerkt met universiteiten of met het Centrum voor Schoolfeedback in het bijzonder.

[Beperkingen van dit onderzoeksluik](#)

In dit onderzoek lag de focus van de RCEC-evaluatie op de schriftelijke toetsen. De praktische proeven kwamen aan bod op de workshop, maar hun kwaliteit naar bijvoorbeeld constructie en evaluatiecriteria werd niet nader geanalyseerd. Momenteel zijn er ook geen pasklare wetenschappelijke criteria beschikbaar voor de beoordeling van de psychometrische kwaliteit van praktische proeven.

In dit onderzoek werd niet onderzocht hoe de feitelijke afnameprocedure is binnen de scholen en hoe de toetsen in de realiteit door de scholen worden gebruikt. De toetsontwikkelaars geven hierover duidelijke communicatie over naar de scholen. Het is een aandachtspunt om hier blijvend op in te zetten.

Gebruikersonderzoek

Inleiding

Doelstelling en werkwijze

Om na te gaan wat het feitelijk gebruik is van IDP en de OVSG-toets door de scholen werd voor zowel IDP, OVSG als GO! een telefonisch interview afgenomen bij de directie van twintig scholen. Alle scholen die data aanleverden voor het specifieke luik werden per e-mail uitgenodigd en daarna telefonisch gecontacteerd door de projectmedewerker tot er 20 scholen bereid waren om deel te nemen. In het telefonisch interview werd gepeild naar hoe de deelname aan de toets binnen de school georganiseerd is en wat er doorgaans met de resultaten gebeurt. De interviewleidraad is weergegeven in Tabel 6. De gemiddelde interviewduur bedroeg iets minder dan een kwartier. De interviews vonden plaats tussen 25 september en 17 oktober voor IDP en OVSG en tussen 14 oktober en 17 november voor de GO!-scholen.

Tabel 6

Interviewleidraad voor de telefonische interviews met directies in het kader van het gebruikersonderzoek

Onderwerp	Richtvragen
Deelname	<ul style="list-style-type: none">- Sinds wanneer neemt uw school deel aan IDP/de OVSG-toets?- Neemt uw school jaarlijks deel aan IDP/de OVSG-toets?
Proces	<ul style="list-style-type: none">- Wat is uw motivatie om aan IDP/de OVSG-toets deel te nemen?- Wie is verantwoordelijk voor IDP/de OVSG-toets in uw school?- Wie verbetert de IDP/OVSG-toets?- Krijgt u na de afname van de proeven/toetsen feedback van de leerkracht?- Wanneer bespreekt u de resultaten, in juni en/of na de vakantie?- Met wie bespreekt u de resultaten?
Leerlingniveau	<ul style="list-style-type: none">- Nemen alle leerlingen deel?- Worden de resultaten van IDP/de OVSG-toets vermeld op het rapport van de leerlingen?- Wordt over het resultaat dat een leerling op IDP/de OVSG-toets behaalt, gecommuniceerd met de ouders?- Speelt het resultaat dat een leerling op IDP/de OVSG-toets behaalt een rol bij het al dan niet toekennen van het getuigschrift basisonderwijs?- Vindt u dat de resultaten die de leerlingen op IDP/de OVSG-toets behalen een weerspiegeling vormen van de mogelijkheden van de leerlingen?
Schoolniveau	<ul style="list-style-type: none">- Vindt u dat de resultaten en de feedback die u van IDP/OVSG ontvangt een weerspiegeling van uw school vormen?- Op welke manier gebruikt uw school de resultaten van IDP/de OVSG-toets?- Op welke manier gebruikt uw school de feedback die ze van IDP/OVSG ontvangt?- Vindt u de feedback die u ontvangt begrijpelijk?
Suggesties	<ul style="list-style-type: none">- Heeft u nog suggesties ter verbetering van IDP/de OVSG-toets?- Heeft u nog suggesties ter verbetering van de feedback die u ontvangt?

Rapportering

Over deze interviews werd een eerste keer mondeling gerapporteerd tijdens het tweede bezoek aan de twee organiserende koepels: op 24 oktober 2016 voor OVSG en op 27 oktober 2016 voor IDP. Voor het GO! werd mondeling gerapporteerd op 21 november 2016. Er werd daarna ook een apart rapport geschreven voor de drie gebruikersgroepen. In een aparte bijlage bij elk rapport zijn in een Exceltabel geanonimiseerd de antwoorden op alle vragen van de twintig scholen samengevat weergegeven. In dit rapport worden de belangrijkste bevindingen uit de drie gebruikersstudies samengevat.

Disclaimer

Deze bevraging geeft slechts een ruwe indicatie van het feitelijk gebruik van IDP en de OVSG-toets door de scholen. Er werd immers maar een beperkte groep van scholen gecontacteerd, er werd doorgaans alleen met de directie gepraat (die misschien niet voor alle vragen de beste informant is) en er werd meestal niet verder gevraagd naar bijkomende verduidelijking van de gegeven antwoorden. De resultaten

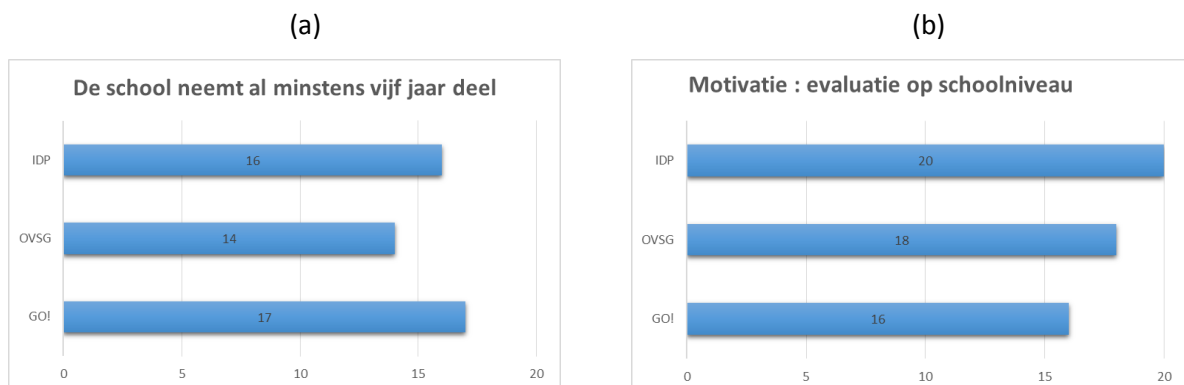
moeten dus absoluut met de nodige voorzichtigheid worden geïnterpreteerd. Omwille van deze beperkingen van het onderzoek en de bijhorende grote onzekerheidsmarge worden de resultaten op de vragen bij de drie gebruikersgroepen hieronder enkel descriptief weergegeven, zonder dat uitspraken worden gedaan of gerapporteerde verschillen statistisch significant zijn.

Bevindingen

Deelname

Zowel IDP als de OVSG-toets heeft een trouw publiek. Ruim 3 op de 4 scholen geeft in de bevraging aan dat ze al minstens vijf jaar deelnemen aan deze toetsafnames (zie Figuur 1a). Dit zou betekenen dat zowat 1 op de 4 scholen pas relatief recent heeft besloten tot deelname. Dat zou dan tot uiting moeten komen in een stijgende trend in de deelnamecijfers door de koepels. Een andere interpretatie is dat sommige scholen officieel nog niet lang deelnemen omdat ze recent gereorganiseerd waren. Dat was toch wat een enkele directeur in de marge van deze vraag opmerkte.

De gerapporteerde motivatie van scholen om deel te nemen ligt nagenoeg in alle scholen in lijn met het primaire gebruiksdoel van zowel IDP als de OVSG-toets: evaluatie op schoolniveau (zie Figuur 1b). Enkele OVSG- en GOI-scholen verwijzen in hun motivatie tot deelname exclusief naar een evaluatie op leerlingniveau.



Figuur 1. Verdeling van het aantal scholen (a) dat al minstens vijf jaar deelneemt en (b) waar verwezen wordt naar een evaluatie op schoolniveau als motivatie voor de deelname.

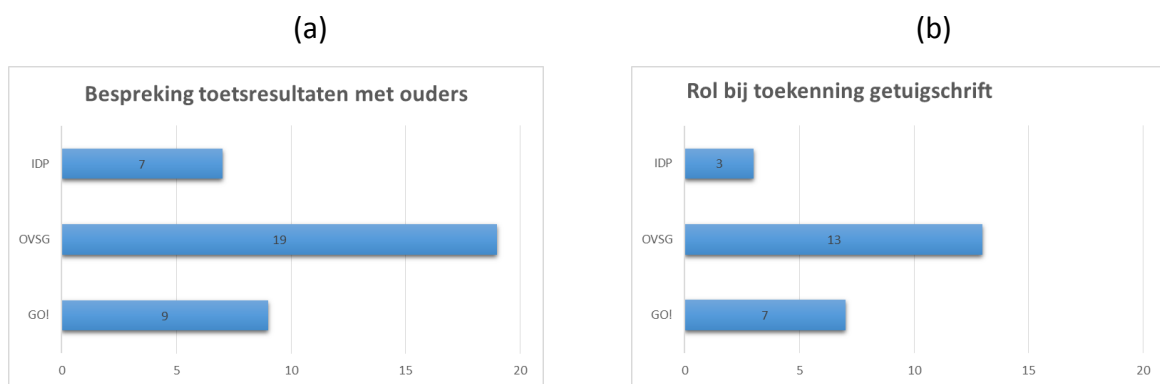
Organisatie

De organisatie van de toetsafnames berust voornamelijk op de leerkrachten van het zesde leerjaar, al neemt ook de directie of de zorgcoördinator een (ondersteunende) rol hierin op. 11 van de 20 scholen die deelnamen aan IDP kozen voor de digitale toetsafname, wat in lijn ligt met de 44 procent digitale afnames in de gehele groep van deelnemende scholen. Alle schoolhoofden bevestigen dat al hun zesdeklassers deelnemen, al verwijzen in alle gebruikersgroepen enkele schoolhoofden naar leerlingen die RA genieten of een IAC volgen als nuancering van dit antwoord.

Gebruik van de toetsen op leerlingniveau

Er zijn grote verschillen tussen de drie gebruikergroepen m.b.t. het gebruik van de toetsen op leerlingniveau, wat niet verwonderlijk is gegeven de verschillende visies die daarover gelden bij IDP en de OVSG-toets. Als indicatie van het gebruik van de toetsen op leerlingniveau wordt in Figuur 2a weergegeven in hoeveel scholen de leerlingresultaten besproken worden met de ouders. Bij de OVSG-

scholen gebeurt dit in nagenoeg alle bevroagde scholen. Voor de bevroagde scholen uit het katholieke onderwijsnet is dat een minderheid van 7 van de 20 bevroagde scholen. Ook bij de scholen uit het GO! geldt dat voor net minder dan de helft van de bevroagde scholen, ook al nemen die ook de OVSG-toets af. Zoals uit Figuur 2b blijkt, geeft enkel bij OVSG een meerderheid van de bevroagde schoolhoofden aan dat de resultaten op de OVSG-toets een rol spelen bij de toekenning van het getuigschrift basisonderwijs. Voor het merendeel van hen is dat gebruik in lijn met de visie die OVSG daarbij vooropstelt, namelijk als objectieve, aanvullende informatie naast de schooleigen informatie. Toch waren er drie scholen in de steekproef van OVSG-scholen waar de OVSG-toets sterk het toekennen van het getuigschrift bepaalt. In één school moeten de leerlingen bijvoorbeeld minstens 50 procent op de OVSG-toets behalen. In een andere school tellen de resultaten op de OVSG-toets mee voor de helft van het eindtotaal waarbij de andere 50 procent uit de resultaten op schooleigen toetsen bestaat. Een dergelijk 'high-stakes' gebruik van de toetsen op leerlingniveau werd niet gerapporteerd binnen de huidige steekproef van IDP- en GO!-scholen. Bij deze gebruikersgroepen werden de toetsen in een aantal scholen wel ingeroepen bij twijfelgevallen of in combinatie met de gehele schoolloopbaan.



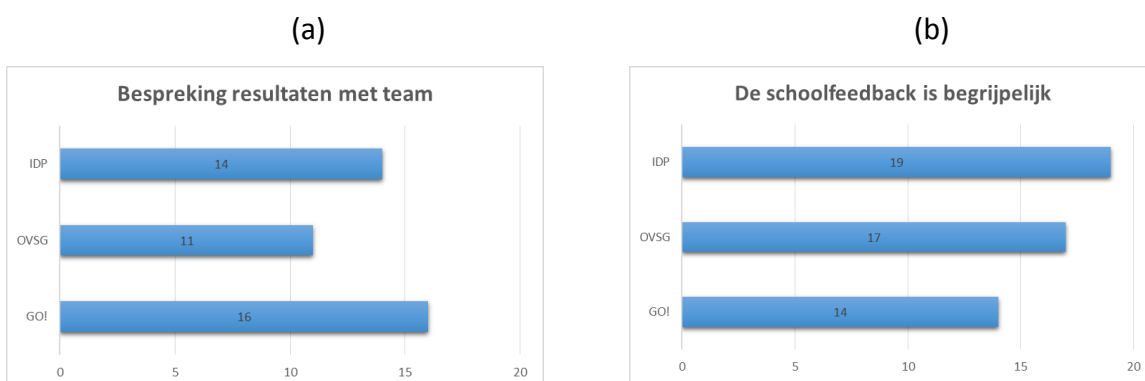
Figuur 2. Verdeling van het aantal scholen (a) waar de toetsresultaten van leerlingen besproken worden met de ouders en (b) waar de toetsen een rol spelen bij het toekennen van het getuigschrift basisonderwijs.

Gebruik van de toetsen op schoolniveau

Het gerapporteerde primaire gebruik van de toetsen ligt in lijn met de door de koepels vooropgestelde doelstellingen, namelijk interne kwaliteitszorg op schoolniveau. Figuur 3a geeft aan dat de resultaten evenwel niet in alle scholen met het hele schoolteam besproken worden.

Figuur 3b geeft aan dat de bevroagde scholen de gegeven schoolfeedback begrijpelijk vinden. Dat lijkt mogelijk wel iets minder het geval voor de GO!-scholen. Merk op dat de vaststelling dat scholen de schoolfeedback begrijpelijk vinden, in principe nog niet impliceert dat de scholen de schoolfeedback ook echt begrepen hebben.

Uit de bevraging van de schoolhoofden bleek ook dat de scholen niet altijd gebruik maken van alle feedback die hen geboden worden. In de aparte rapporten over het gebruikersonderzoek wordt daarover in detail gerapporteerd.



Figuur 3. Verdeling van het aantal scholen (a) waar de schoolresultaten worden besproken met het schoolteam en (b) die de schoolfeedback begrijpelijk vinden.

Op vraag van de stuurgroep werd aan de directies tijdens het interview ook gevraagd of ze vinden dat de toetsen een goede weerspiegeling vinden van hun school en van de mogelijkheden van hun leerlingen (zie Figuren 4a en 4b). Opvallend is dat voor het OVSG en GO! de directies hier telkens unaniem ja op antwoorden¹². Ook bijna vier op de vijf scholen die deelnamen aan IDP stemmen in met beide vragen¹³. Bij de vier scholen die niet akkoord waren met de vraag of de toetsen een goede weerspiegeling op schoolniveau zijn, gaven drie als reden aan dat ze de vergelijkende informatie niet eerlijk vonden omdat volgens hen bepaalde scholen niet alle leerlingen laten deelnemen. We kunnen deze toelichting niet verder checken of duiden. Op de vraag of de toetsen een goede weerspiegeling zijn op leerlingniveau



Figuur 4. Verdeling van het aantal scholen die de toetsen een weerspiegeling vinden (a) van hun school en (b) van de mogelijkheden van hun leerlingen.

gaven 10 scholen die deelnamen aan IDP aan dat er soms of vaak verrassingen voorkomen, zowel in positieve zin (leerlingen scoren hoger dan verwacht) als in negatieve zin (leerlingen scoren lager dan verwacht). Er waren eigenlijk ook 10 OVSG-scholen die ditzelfde opmerkten voor de OVSG-toets. Bij GO! waren dat vier scholen. Merk op dat de opmerking over de soms onverwachte resultaten op leerlingniveau misschien niet geheel onverwacht is, gegeven dat in het specifieke luik werd vastgesteld dat voor de daar onderzochte leergebieden de toetsen op individueel niveau onvoldoende betrouwbaar

¹² Bij GO! kon een schoolhoofd niet op deze vraag antwoorden omdat ze nog maar pas aangesteld was.

¹³ Het gaat immers om 15 op de 19 scholen, omdat één schoolhoofd pas in functie was en aangaf daarom niet te kunnen antwoorden op deze vragen.

zijn. Op zich kan het trouwens ook een kwaliteit zijn van een toetsinstrumentarium als de gebruikers vinden dat ze daarbij informatie krijgen die niet altijd in lijn ligt met de eigen perceptie. Een alternatieve verklaring dat scholen de resultaten soms verrassend vinden is dat de perceptie van de leerkrachten over individuele leerlingen onterecht was of dat de scholen er zelf geen goede evaluatiepraktijk op nahouden.

Conclusies

- Het gebruik van IDP en de OVSG-toets ligt grotendeels in lijn met de door de koepels vooropgestelde doelstellingen, namelijk interne kwaliteitszorg op schoolniveau. Dit blijkt uit de redenen die scholen geven waarom ze deelnemen, maar ook uit het gebruik van de toetsen op schoolniveau. Scholen vinden de resultaten op de toetsen doorgaans ook een goede weerspiegeling van hun school.
- De toetsen worden volgens de directies in mindere mate ingezet op leerlingniveau en ouders krijgen niet altijd informatie over de toetsprestaties van hun kinderen. Dit ligt anders bij de OVSG-scholen waar de leerlingen ook een expliciete doelgroep zijn van de OVSG-toets. Daarbij wordt door OVSG echter vooropgesteld dat de OVSG-toets enkel aanvullende informatie biedt op de schooleigen informatie over de leerlingen (cf. algemeen luik). Slechts in een beperkt aantal bevraagde OVSG-scholen spelen de resultaten op leerlingniveau een crucialere rol bij het toekennen van het getuigschrift basisonderwijs. De onderwijsverstrekkers moeten blijven inzetten op hun communicatie om dergelijk oneigenlijk gebruik van de toetsen tegen te gaan. Voor scholen zijn de resultaten op leerlingniveau ook een goede weerspiegeling van de kwaliteiten van de leerlingen, maar geregeld wordt aangegeven dat de toetsresultaten van individuele leerlingen verrassend kunnen zijn, zowel in positieve zin (leerlingen doen het beter dan verwacht) als in negatieve zin (leerling doet het minder goed dan verwacht).
- Het gebruik van de toetsen als schoolfeedback kan in alle netten nog verder geoptimaliseerd worden. Bijvoorbeeld, niet alle scholen gebruiken alle informatie uit de schoolfeedbackrapporten en niet in alle scholen worden de schoolresultaten teruggekoppeld naar het team. Een beperkt aantal scholen geeft aan dat voor hen de feedback niet begrijpelijk is.

Workshop praktische proeven

Beschrijving en rapportering

Op 4 oktober werd een workshop georganiseerd over principes en problemen bij praktische proeven. Na een korte inleiding presenteerden Netje De Groof en Alexia Deneire (UA) in de voormiddag de resultaten van OBPWO-project 13.02 waarin een evaluatiematrix voor grootschalige competentietoetsen werd ontwikkeld. In de namiddag werd verder stilgestaan bij de uitdagingen van praktische proeven en werd de werkwijze voorgesteld voor de ontwikkeling van praktische proeven bij het peilingsonderzoek. Daarbij was er gelegenheid tot vragen en discussie vanuit de deelnemers.

Aan deze workshop namen zes personen deel van Katholiek Onderwijs Vlaanderen en ook zes van het OVSG. Het GO! nam deel met een persoon. De overheid en inspectie waren elk vertegenwoordigd met een persoon.

Als verslag van de workshop praktische proeven zijn de presentaties gebundeld (zie bijlagen) en is er ook een korte samenvatting van de discussiepunten die naar voor werden gebracht in de namiddag (eveneens in de bijlagen).

Conclusies

Praktische proeven zijn noodzakelijk voor het meten van bepaalde leerplandoelen, maar zijn noodgedwongen beperkt in hun meetdoel en gebruiksdoel. Deze beperkingen moeten niet wegnemen dat de koepels op de ingeslagen weg verder gaan om ook praktische proeven te ontwerpen voor hun toetsinstrumentarium. De focus kan daarbij wellicht iets meer verschuiven van het gebruik van de praktische proeven als didactisch instrument om leerplandoelen onder de aandacht te brengen naar het verder uitwerken van de praktische proeven als evaluatie-instrument. Er wordt gehoopt dat ook de overheid de praktische proeven naar waarde kan schatten voor het toetsinstrumentarium voor zesdeklassers op het einde van de basisschool.

De workshop was een eerste aanzet om gezamenlijk met de verschillende teams van toetsontwikkelaars na te denken over de principes en problemen van praktische proeven. Tijdens de vraag- en discussierondes is alleen in het algemeen ingegaan op de praktische proeven uit IDP en de OVSG-toets. Hun constructie en evaluatiecriteria werden niet in detail onder de loep genomen. Dat is misschien een hiaat van dit onderzoeksproject, al viel dat niet onder de vooropgestelde doelstellingen. Ook biedt de psychometrie momenteel geen pasklare criteria hiervoor.

Specifieke luik

Inleiding

Doelstelling

In aanvulling op het algemene luik van het valideringsonderzoek werden voor één leergebied empirische analyses uitgevoerd om de psychometrische kwaliteiten van de toetsen te onderzoeken. Voor IDP werd het leergebied Nederlands onderzocht. Voor de OVSG-toets was dat het leergebied wereldoriëntatie. Daarbij werden niet alleen analyses uitgevoerd voor leerlingen uit OVSG-scholen, maar ook voor leerlingen uit het GO!, zodat kon nagegaan worden of de OVSG-toets voor deze gebruikersgroep dezelfde meetkwaliteiten heeft. De schriftelijke toetsen uit beide leergebieden werden geanalyseerd, zowel op het niveau van de scores als op het niveau van items. Daarbij werd onderzocht in welke mate de scores op de toetsen kunnen gebruikt worden om uitspraken te doen op leerling- en schoolniveau. Tot slot werd nagegaan in welke mate de toetsen van dit leergebied samenhangen. De doelstelling van deze analyses was tweeledig: enerzijds werd een oordeel uitgesproken over de kwaliteit van de bestudeerde toetsen en anderzijds werden suggesties geformuleerd om de kwaliteit verder te verbeteren.

Werkwijze

Er werd vooropgesteld om te werken met een deelsteekproef van 1000 leerlingen die in juni 2016 deelnamen aan IDP of aan de OVSG-toets binnen OVSG en binnen het GO!. Hiervoor werd door de onderzoekers een representatieve steekproef van 40 scholen getrokken uit de populatie van scholen van de betreffende koepel of het betreffende net, rekening houdend met de stratificatievariabelen schoolgrootte en provincie. Voor elke steekproefschool werden ook twee reservescholen getrokken. Deze selectie van scholen werd in mei aan de drie teams bezorgd. Zij zorgden voor het contacteren van de scholen en het verzamelen van de gegevens.

Tabel 7 geeft een overzicht van de drie steekproeven. Door het IDP-team werden de data elektronisch aangebracht. De OVSG- en de GO!-scholen bezorgden het onderzoeksteam de toetsboekjes van deze leerlingen. Tijdens de zomermaanden werden de antwoorden van de leerlingen op de toetsvragen over wereldoriëntatie ingetikt door jobstudenten. Op dit databestand werden de analyses uitgevoerd.

Tabel 7

Overzicht van de steekproeven voor het specifieke luik

Toets	Leergebied	Aantal scholen	Aantal leerlingen
IDP	Nederlands	39	1063
OVSG-toets: OVSG	wereldoriëntatie	38	1032
GO!	wereldoriëntatie	38	1075

De psychometrische analyses werden uitgevoerd met het statistische pakket SPSS en met een zelfontworpen Excel-programma voor toets- en itemanalyse (Vermeulen & Janssen, 2016). Binnen dit programma wordt per item een grafische scoregroepanalyse¹⁴ weergegeven. Deze grafiek biedt een visuele weergave van de moeilijkheidsgraad van het item en van hoe het item functioneert binnen de toets.

Rapportering

Voor elke toets is al de informatie uit de analyse samengevat binnen een afzonderlijk Excel-bestand dat in november aan de koepels werd bezorgd. Tijdens het tweede bezoek ter plaatse werden de globale

¹⁴ Deze analysetechniek werd in de aparte rapporten over het specifieke luik verder toegelicht.

resultaten van deze analyses besproken en nader toegelicht. Exemplarisch werd de itemanalyse voor één toets in detail besproken. Het schriftelijke rapport voor elk van beide toetsen volgde dezelfde logica als de mondelinge rapportering¹⁵. In dit rapport worden de belangrijkste bevindingen samengevat.

Bevindingen

Afnamevorm

Bij IDP werd als aandachtspunt geformuleerd om te onderzoeken of er verschillen zijn in de psychometrische kwaliteit van de papieren en digitale afnamevorm. Dit kon in het huidige project niet worden nagegaan omdat in de dataset de afnamevorm niet stond aangegeven. De koepel neemt dit momenteel zelf verder op.

Vraagvormen

Bij beide toetsinstrumentaria werden overwegingen aangereikt m.b.t. de vraagvorm binnen meerkeuzevragen (IDP) of de variëteit aan vraagvormen (OVSG-toets).

Itemstatistieken

Het is nodig dat het team van toetsontwikkelaars zelf systematisch informatie verzamelt over de moeilijkheidsgraad van de itemanalyse en bij meerkeuzevragen ook zelf een afleidersanalyse kan uitvoeren. Dit gebeurt nu al bij IDP.

Beheersingstoetsen

De gemiddelde moeilijkheidsgraad van de toetsen Nederlands uit IDP varieerde tussen .71 en .77. Voor de toetsen wereldoriëntatie van de OVSG-toets lag de gemiddelde moeilijkheidsgraad tussen .67 en .86. Enkel voor een toets bij wereldoriëntatie werd een plafondeffect¹⁶ vastgesteld. Het is beter om extreem makkelijke items te vermijden (OVSG-toets). De moeilijkheidsgraden zijn van acceptabel niveau voor beheersingstoetsen¹⁷.

Scoregroepanalyse

Uit de scoregroepanalyses kwam bij IDP en bij de OVSG-toets geen enkel item naar voor dat manifest van onvoldoende toetskwaliteit was. Dit bevestigt de expertise van de toetsontwikkelaars en geeft aan dat er in het toetsontwikkelingsproces voldoende screening is van de itemkwaliteit voor de grootschalige toetsafname.

De scoregroepanalyses geven verfijnde informatie over de itemkwaliteit en geven ook zinvolle feedback naar de toetsontwikkelaars. Het is zinvol om dit soort van analyses in te bouwen in het toetsontwikkelingsproces.

Gebruik van de toetsen op leerlingenniveau

Voor beide toetsinstrumentaria bleek dat de afzonderlijke schriftelijke toetsen uit de twee leergebieden onvoldoende betrouwbaar zijn om een oordeel te vellen over individuele leerlingen. Dit is op zich niet zo verwonderlijk gegeven het relatief beperkte aantal items per toets. Enkel een geaggregeerde score voor het hele leergebied wereldoriëntatie (OVSG-toets) of taal (IDP) heeft een acceptabele meetfout om op individueel niveau te worden gebruikt.

¹⁵ In overleg met GO! werd geen apart rapport opgemaakt voor de analyses van de OVSG-toets voor leerlingen uit GO!-scholen omdat het GO! slechts gebruikers zijn van de toets.

¹⁶ Een toets met een plafondeffect is een toets die zo gemakkelijk is voor de meeste leerlingen dat de toets niet meer kan differentiëren tussen de sterkere leerlingen.

¹⁷ Beheersingstoetsen gaan na of leerlingen bepaalde leerstof onder de knie hebben. Vaardigheidstoetsen bevatten daarentegen ook opgaven die makkelijker of moeilijker zijn dan de eigenlijke leerstof.

De lage betrouwbaarheid op leerlingniveau zegt op zich niets over de itemkwaliteit. De itemkwaliteit kan beter beoordeeld worden in een itemanalyse en scoregroepanalyse. Voor beide leergebieden werden inhoudelijke interpretaties aangereikt voor de verschillen in betrouwbaarheid tussen de toetsen. Het relatief beperkt aantal items per toets lijkt alle toetsen parten te spelen.

Voor IDP vormen de leerlingen geen doelgroep van de toetsen (cf. algemeen luik). Voor OVSG kunnen de resultaten op de OVSG-toets enkel ingezet kunnen worden als een bijkomende bron van informatie naast de schooleigen gegevens. Ze kunnen dus geen alleenstaand element zijn bij het al dan niet toekennen van het getuigschrift basisonderwijs. De bevindingen over de betrouwbaarheid bieden een ondersteuning voor dit gebruik, maar dan enkel voor de score op het leergebied.

Gebruik van de toetsen op schoolniveau

De interne consistentie van de onderzochte toetsen is van net voldoende hoog niveau opdat per toets het gemiddelde op schoolniveau betrouwbaar gemeten kan worden. Deze voorzichtige vaststelling bevestigt het doel van IDP en het primaire doel van de OVSG-toets als instrument voor de zelfevaluatie van de scholen.

Gebruik van de OVSG-toets binnen GO!

Er werden ook aparte analyses uitgevoerd om voor wereldoriëntatie de kwaliteit van de OVSG-toets na te gaan voor de leerlingen uit het GO! Er waren slechts weinig verschillen met de resultaten voor OVSG. De OVSG-toets kan alvast voor wereldoriëntatie zonder problemen worden gebruikt in scholen uit het GO!.

Conclusies

Uit het algemeen luik kwam uit de evaluatie van IDP en de OVSG-toets o.a. naar voren dat voor beide toetsinstrumentaria weinig empirische gegevens beschikbaar waren over de kwaliteit van hun toetsen. Deze informatie werd in het specifieke luik aangevuld voor één leergebied: Nederlands voor IDP en wereldoriëntatie voor de OVSG-toets. Het is aangewezen om deze analyses in de toekomst in te bouwen in het proces van toetsontwikkeling.

Reflecties en aanbevelingen i.v.m. de ‘toolkit met gevalideerde toetsen’

In dit hoofdstuk wordt op basis van de bevindingen in dit project gereflecteerd over verschillende aspecten van het beleid rond de toolkit met gevalideerde toetsen. Als beknopte toelichting daarop wordt eerst geciteerd uit enkele recente beleidsdocumenten.

Beknopte toelichting van de toolkit

In haar beleidsnota 2014-2019 refereerde Crevits (2014) naar de toetsing van zesdeklassers en naar dit project:

Zoals voorzien in het masterplan secundair onderwijs worden alle leerlingen op het einde van het basisonderwijs getoetst. Met de onderwijsverstrekkers ga ik na hoe we de huidige toetsen van het Onderwijssecretariaat van de Steden en Gemeenten van de Vlaamse Gemeenschap (OVSG) en de interdiocesane proeven kunnen laten valideren door het Steunpunt Toetsontwikkeling en Peilingen. Vertrekkend van deze toetsen en de paralleltoetsen die voortkomen uit de peilingsproeven, maak ik werk van een ‘toolkit met gevalideerde toetsen’ waar scholen naar eigen keuze gebruik van kunnen maken in het kader van interne kwaliteitszorg en heeft ook tot doel gegevens te verzamelen op systeemniveau. Met de onderwijsverstrekkers van het secundair onderwijs bekijk ik of deze werkwijze ook voor hen kan worden opgezet. Wat betreft toetsen die uitspraak doen over prestaties van individuele leerlingen, vertrekken we van het principe dat het evaluatiebeleid blijft behoren tot de volledige autonomie van de school.

Deze beleidsnota werd verder geconcretiseerd in Crevits’ (2016) beleidsbrief 2016-2017:

Eind 2016 worden de resultaten van het onderzoek naar de validering OVSG en IDP-toetsen opgeleverd. Aan de hand daarvan zal ik in overleg met de onderwijsverstrekkers werk maken van een toolkit met gevalideerde toetsen dat meerdere leergebieden omvat en waar ook de paralleltoetsen van de peilingen deel van zullen uitmaken. Het is de bedoeling dat scholen vanaf het schooljaar 2017-2018 alle leerlingen voor minstens twee leergebieden verplicht laten deelnemen aan een gevalideerde toets uit deze toolkit. Ik zal voldoende aandacht besteden aan de communicatie naar scholen om aan te geven dat de toetsen kaderen in hun interne kwaliteitszorg, met het oog op het verkrijgen van informatie op schoolniveau.

In het voorontwerp van Onderwijsdecreet XXVII van december 2016 wordt ook verwezen naar de toolkit. Specifiek wordt aangegeven dat verwacht wordt dat vanaf het schooljaar 2018-2019 scholen op het einde van het gewoon lager onderwijs van elke leerling een gevalideerde toets afneemt voor ten minste drie leergebieden. Daarbij wordt ook het gebruiksdoel verder verduidelijkt:

De resultaten van deze toetsafnames zijn gericht op het verkrijgen van informatie op schoolniveau over de mate waarin de leerlingenpopulatie de eindtermen bereikt en ze worden aangewend in het kader van interne kwaliteitszorg. De resultaten kunnen één van de elementen zijn waar de klassenraad rekening mee houdt bij de toekenning van een getuigschrift zoals bepaald in artikel 53.

Het toetsen van alle leerlingen

Welke scholen nog niet?

Op dit moment kent zowel IDP als de OVSG-toets een groot bereik in het onderwijsveld. Naar schatting gebruiken nu al bijna 9 op de 10 Vlaamse basisscholen een van beide toetsinstrumenten op het einde van de basisschool. Specifiek bereikte IDP in juni 2016 85 procent van de scholen uit het katholieke onderwijsnet. De OVSG-toets werd gebruikt door 98 procent van de scholen uit het stedelijke en gemeentelijke net en door iets meer dan negentig procent van de scholen uit het GO!

Hoewel deze drie onderwijsnetten niet alle Vlaamse basisscholen vertegenwoordigen, is er dus al in het afgelopen schooljaar een grote opvolging van de nakende verplichting aan basisscholen om voor twee (of drie) leergebieden toetsen af te nemen bij hun zesdeklassers. Door deel te nemen aan IDP of aan de OVSG-toets kunnen scholen immers aan deze verplichting voldoen.

Het zou interessant zijn om meer te weten over de scholen die nu nog geen gebruik zouden maken van toetsen op het einde van de basisschool. Misschien dat de onderwijsverstrekkers hier zelf ook een initiatief in kunnen nemen om de beweegredenen van deze beperkte groep van hun scholen te achterhalen.

Alle leerlingen binnen een school

Zowel bij IDP als bij de OVSG-toets worden scholen aangemoedigd om leerlingen die RA genieten of die een IAC volgen ook te laten deelnemen aan (een deel van) hun toetsen. Op basis van dit onderzoek weten we niet of de scholen op uniforme wijze met dit advies omgaan. Weliswaar bleek in het gebruikersonderzoek dat een aantal schoolhoofden aanhalen dat er volgens hen hierin verschillen tussen scholen zijn. Het is moeilijk om de gegrondheid van hun uitspraak te achterhalen. Toch verdient het de aanbeveling dat de toetsmakers hierover in de toekomst duidelijke richtlijnen naar scholen blijven geven¹⁸. Een positief punt dat in dit rapport nog niet eerder werd vermeld, is dat zowel bij IDP als bij de OVSG-toets de scholen de mogelijkheid krijgen om de toetsen op te vragen in een aangepast format, zodat de toetsen ook met courante voorleessoftware kunnen worden afgelegd.

Welk soort toetsen?

Het beleid spreekt zich niet uit welke leergebieden of domeinen getoetst zouden moeten worden op het einde van het basisonderwijs. Ook wordt niet aangegeven in welke mate dat de gebruikte toetsen het gehele scala van leerplandoelen en eindtermen uit het leergebied moeten omvatten. In dit project lag de focus vooral op schriftelijke toetsen. Een aandachtspunt bij de implementatie van de beleidsnota is dat ook praktische proeven relevant kunnen zijn voor het meten van competentiegerichte vaardigheden van leerlingen. In dit onderzoek is vooral gewezen op het feit dat praktische proeven nodig zijn, zeker in het kader van breed evalueren, maar ook noodgedwongen beperkt. Hun gebruik op leerling- en schoolniveau is nog niet grondig onderzocht. Momenteel zijn er in de wetenschappelijke literatuur ook nog geen pasklare evaluatiecriteria voor praktische proeven voorhanden.

Het gebruik van de toetsen op schoolniveau

Bij IDP ligt het enige gebruiksdoel op schoolniveau. Bij de OVSG-toets is dat het belangrijkste gebruiksdoel. Beide toetsinstrumenten willen scholen een middel aanreiken voor hun zelfevaluatie en hun interne kwaliteitsontwikkeling. Daarbij wordt voorzien in relatieve meting. Scholen kunnen de gemiddelde toetsresultaten van hun leerlingen vergelijken met het Vlaamse gemiddelde en met het gemiddelde van scholen die een vergelijkbaar leerlingenpubliek hebben op het vlak van gelijkheidsindicatoren. In het algemene luik werd al aangegeven hoe deze relatieve metingen nog verder geoptimaliseerd kunnen worden vanuit statistisch perspectief.

Het gebruik van de toetsen op leerlingniveau

IDP en de OVSG-toets verschillen m.b.t. het toevoegen van het leerlingniveau als gebruiksdoel. Momenteel ziet Katholiek Onderwijs Vlaanderen de leerlingen niet als doelgroep, terwijl dit voor OVSG genuanceerder ligt. Zij vinden het belangrijk om met de OVSG-toets aan scholen bijkomende, onafhankelijke informatie te bieden die naast de schooleigen informatie kan helpen in het nemen van

¹⁸ Dit is trouwens ook een aandachtspunt voor de peilingen.

beslissingen over individuele leerlingen. Dit gebruiksdoel wordt ook in het nieuwe onderwijsdecreet als mogelijkheid vermeld.

In het licht van de (hernieuwde) keuzes die de onderwijsverstrekkers in de toekomst zullen maken over het gebruik van IDP en de OVSG-toets op leerlingniveau is het vermeldenswaard dat sommige onderzoekers duidelijk pleiten voor het gebruik van gestandaardiseerde toetsen in het onderwijs bij het nemen van belangrijke voortgangsbepalingen over individuele leerlingen. Agirdag (2014) verwijst naar de emancipatorische kracht van “low-stakes”, gestandaardiseerde toetsen. Volgens hem kunnen goede toetsen immers de onderwijsongelijkheid tegengaan. De internationale studie van Bol, Witschge, Van de Werfhorst en Dronkers (2014) in 38 landen toont aan dat centrale examens ervoor zorgen dat scholen bij de studiekeuze van hun leerlingen meer objectieve indicatoren laten doorwegen. Deze bevinding kan misschien gegeneraliseerd worden naar het gebruik van low-stakes gestandaardiseerde toetsen. Tot slot zijn de argumenten die Verschuere (2014, p. 35) aangeeft voor het gebruik van toetsen bij de overgang van kleuter- naar lagere school ook relevant in de discussie over het gebruik van toetsen op leerlingniveau op het einde van het basisonderwijs:

Tests hebben, naast observaties en ervaringen van leerkrachten, hierin absoluut een meerwaarde. Ze helpen de leerkracht om indrukken en ervaringen te onderbouwen en verbreden de vergelijkingsbasis. Ook en misschien zéker voor kinderen uit kansengroepen hebben zij een meerwaarde. Dagelijkse oordelen zijn immers mogelijk onderhevig aan vertekening en stereotypering. Aanvulling met objectieve testresultaten kan ons hiervoor behoeden en kan de sterktes van een kind soms zelfs beter tot hun recht laten komen.

Toolkit gevalideerde toetsen

Op basis van dit onderzoek bevelen we aan om IDP en de OVSG-toets op te nemen in de toolkit van gevalideerde toetsen voor het basisonderwijs. De schriftelijke toetsen die gebruikt werden in juni 2016 worden van voldoende kwaliteit geacht voor het realiseren van de basisdoelstelling voor het geven van schoolfeedback aan scholen uit het basisonderwijs, ook al werden daar in de verschillende onderdelen van dit project aandachtspunten bij geformuleerd. Ook wijzen de resultaten van dit onderzoek in de richting dat de schriftelijke toetsen een voldoende betrouwbaarheid hebben om voor een leergebied op leerlingniveau te rapporteren. De inzet van praktische proeven is relevant met het oog op een brede evaluatie.

De vraag zou gesteld kunnen worden of dit advies wel gegeven kan worden op basis van de beoordeling van één toetsjaargang. Als antwoord op deze vraag kan verwezen worden naar het RCEC-beoordelingsinstrument. Daar wordt een beoordeling uitgesproken die geldt voor vijf jaar, ook al worden in die vijf jaar telkens andere toetsen of examens opgesteld. De idee hierachter is dat voor het behalen van een goede beoordeling voor één editie van toetsen of examens kwaliteitsvolle procedures en degelijke toetsexpertise nodig zijn, waarvan verondersteld kan worden dat die ook bij toekomstige toetsontwikkelingen ingezet zullen worden. Het is daarbij evenwel ook nodig dat de kwaliteit van toekomstige toetsen systematisch wordt opgevolgd.

Referenties

- Agirdag, O. (2014). Niet toevallig, die standaardisering. *Opiniebijdrage. De Standaard*, 27 augustus.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- Bol, T., Witschge, H. G., Van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces*, 92, 1545-1572.
- Crevits, H. (2014). *Beleidsnota Onderwijs 2014-2019*. Brussel: Vlaams Parlement.
- Crevits, H. (2016). *Beleidsbrief Onderwijs 2016-2017*. Brussel: Vlaams Parlement
- Eggen T.J.H.M., & Sanders, P.F. (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Sanders, P., van Dijk, P., Eggen, T., den Otter, D., & Veldkamp, B. (2016). *RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens*. Enschede: RCEC.
- Verschueren, K. (2014). Veel getoeter; weinig nuance. *Caleidoscoop*, 26 (2), 35.