

Toetsen op School Primair onderwijs

Judith Hollenberg en Marleen van der Lubbe
Piet Sanders (redactie)



Toetsen op School Primair onderwijs

Judith Hollenberg en Marleen van der Lubbe

Piet Sanders (redactie)

Toetsen op School PO is een uitgave van Cito

- Redactie: Piet Sanders
- Auteurs: Judith Hollenberg, Marleen van der Lubbe en Piet Sanders
- Opmaak: Service unit, MMS
- Foto omslag: Ron Steemers

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2012) | Herziene versie (2017)

Alle rechten voorbehouden. Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook. Stichting Cito Instituut voor Toetsontwikkeling Arnhem heeft getracht alle rechthebbenden te achterhalen. Indien iemand meent als rechthebbende in aanmerking te komen, kan hij of zij zich tot Cito wenden.

Inhoud

	Inleiding	5
1	Het doel van toetsen	9
1.1	Toetsen op leerlingniveau	10
1.1.1	Het selecteren van leerlingen	10
1.1.2	Het classificeren van leerlingen	12
1.1.3	Het plaatsen van leerlingen	18
1.2	Toetsen voor het beoordelen van het onderwijsleerproces	19
1.3	Toetsen voor het beoordelen van groepen leerlingen en scholen	26
1.4	Toetsen voor het beoordelen van de kwaliteit van het onderwijs	35
2	De inhoud en constructie van toetsen	41
2.1	Visie op leren en toetsen	42
2.2	Aansluiting toetsen bij kerndoelen en referentieniveaus	45
2.3	Van kerndoelen en referentieniveaus naar leerlijnen	48
2.4	De constructie van LVS-toetsen	54
3	Interpretatie en gebruik van toetsresultaten	59
3.1	Soorten toetsscores en normeringen	60
3.1.1	Aantal goed en percentage goed	61
3.1.2	Werken met vaardigheidsscores	63
3.1.3	Andere soorten toetsscores	65
3.2	Van toetsscore naar handelen in de groep	68
3.2.1	Kijken naar niveau	68
3.2.2	Kijken naar groei	72
3.2.3	Kijken naar specifieke aandachtsgebieden	74
3.3	Toetsscores als basis voor ontwikkelingsperspectief	78
4	De kwaliteit van toetsen in het primair onderwijs	83
4.1	COTAN Beoordelingssysteem voor de kwaliteit van tests	84
4.1.1	Voorbeeld van een COTAN-beoordeling	92
4.2	Beoordelingskader Eindtoets	93
4.2.1	Voorbeelden van beoordelingen andere eindtoetsen	95
4.3	Beoordelingskader LOVS-toetsen	97
4.3.1	Voorbeeld van beoordeling LOVS-toets	99
4.4	Conclusies en discussie	101

5	Gebruikers van toetsresultaten	103
5.1	Ouders en leerlingen	104
5.2	Schoolteam en schooldirectie	111
5.3	Schoolbestuur	114
5.4	Onderwijsinspectie	116
	Geraadpleegde literatuur	119
	Index	123

Inleiding

Inleiding

Een leerkracht in het primair onderwijs (PO) is, veelal ongemerkt, dagelijks bezig met toetsen. Het stellen van een vraag of observeren hoe een leerling een opgave oplost, geeft een leerkracht informatie over wat een leerling al beheerst en wat de leerling nog moet leren. Ook het afnemen van toetsen als onderdeel van de methode (methodegebonden toetsen) maakt duidelijk of de leerstof die in de afgelopen lessen behandeld is, door een leerling wordt beheerst. Deze dagelijkse vormen van evaluatie hebben echter beperkingen omdat ze subjectief en persoonsgebonden zijn. Twee leerkrachten kunnen dezelfde prestatie of hetzelfde gedrag verschillend beoordelen. Zo kan de ene leerkracht vinden dat een leerling een rekenopgave vlot oplost, terwijl een andere leerkracht van mening is dat de leerling er behoorlijk lang over doet en dat de stof dus nog niet voldoende geautomatiseerd is. Het is ook mogelijk dat de beoordeling van de leerkracht niet consistent is over de tijd of over leerlingen heen. Wanneer een leerkracht bijvoorbeeld de werkstukken van zijn leerlingen beoordeelt, is het zeer waarschijnlijk dat de criteria die hij voor de beoordeling gebruikt niet voor elke leerling op dezelfde manier worden geïnterpreteerd. Zijn beoordeling op maandag kan afwijken van zijn beoordeling op woensdag. Daarnaast ontbreekt bij dagelijkse evaluatie vaak een objectieve standaard voor beoordeling: Wanneer spreken we bijvoorbeeld over actieve betrokkenheid van een leerling bij het kringgesprek? Op welk moment beschouwen we deze betrokkenheid als (on)voldoende? Wanneer is de prestatie op een methodegebonden toets voldoende: Is dat als zes van de tien of als acht van de tien opgaven goed wordt beantwoord? Een ander vraagstuk bij dagelijkse evaluatie is of we wel systematisch even veel tijd aan de dagelijkse evaluatie van alle leerlingen in de groep besteden of dat we misschien voor sommige leerlingen over een beperktere hoeveelheid informatie beschikken dan voor anderen. Tot slot worden de via dagelijkse evaluatie verzamelde gegevens veelal niet vastgelegd in bijvoorbeeld het leerlingdossier. Door het ontbreken van eenzelfde en door alle leerkrachten gedeeld referentiekader voor het beoordelen en vastleggen van prestaties of gedrag, is het moeilijk om zicht te krijgen op ontwikkeling over een langere periode in de tijd en een leerling in zijn ontwikkeling te volgen. Bovendien is op deze manier niet goed te bepalen of de ontwikkeling van een leerling voldoende is.

Naast vormen van dagelijkse evaluatie, zijn er in het PO daarom ook meer systematische manieren om leerlingen te toetsen met behulp van methodeonafhankelijke toetsen of observatie-instrumenten. Met behulp van deze toetsen en instrumenten kunnen op vaste momenten en aan de hand van objectieve standaarden voor alle leerlingen betrouwbare gegevens over hun ontwikkeling worden verzameld en geregistreerd. De resultaten kunnen vervolgens worden gebruikt om waar nodig het onderwijsleerproces bij te sturen en het handelen van de leerkracht beter op de leerling af te stemmen.

Toetsen in het PO gebeurt op verschillende manieren. De meest bekende manier is een toets op papier. In veel gevallen leest de leerling zelf de opgave en noteert het antwoord. In sommige gevallen leest de leerkracht de opgave voor en noteert de leerling het antwoord, bijvoorbeeld bij een spellingdictee. Bij andere toetsen wijst de leerling het antwoord aan zoals vaak gebeurt als er bij peuters getoetst wordt. Ook in de lagere groepen (t/m medio groep 4) zal de leerkracht afhankelijk van het leergebied de opgaven doorlezen.

Naast papieren toetsen zijn er binnen het PO andere toetsvormen in gebruik zoals digitale toetsen (afname en beoordeling via de computer) en observatiesystemen. Een toets kan bestaan uit open vragen of opdrachten, waarbij de leerling zelf een antwoord

moet opschrijven, of gesloten vragen, waarbij de leerling kiest tussen enkele antwoord-alternatieven. Ook bij observatiesystemen worden zowel open als gesloten vragen gebruikt. In een dergelijk geval noteert de observator zijn antwoord aan de hand van een gerichte vraag, zoals 'Hoe vaak is de leerling per les afgeleid van zijn werk?' Een andere optie is dat de observator een keuze maakt uit gegeven mogelijkheden, zoals bij de stelling 'de leerling vertelt iets in een groepsgesprek', met daarbij de keuze uit nooit, zelden, soms, geregeld en vaak.

Grofweg zijn voor het PO de volgende toetsvormen te onderscheiden.

1 Schriftelijke toetsen

Schriftelijke toetsen zijn nu nog de meest voorkomende vorm van toetsen in het PO.

De leerling krijgt een opgavenboekje (in de hogere groepen met antwoordblad) en noteert de antwoorden in het opgavenboekje (of op het antwoordblad). De leerling leest de opgaven zelf of er is sprake van een door de leerkracht geleide afname, dat wil zeggen dat de leerkracht de opgaven voorleest en de leerling de antwoorden noteert of aanwijst.

De meeste toetsen voor het PO die opgenomen zijn in Toetsgids (onderdeel van www.toetswijzer.nl met informatie over schooltoetsen; zie hoofdstuk 5 van het algemene deel van Toetsen op School) zijn toetsen in deze categorie. Een goed voorbeeld zijn de LVS-toetsen als onderdeel van het Cito Volgsysteem primair en speciaal onderwijs (LOVS).

2 Digitale toetsen

De meeste digitale toetsen die in het PO worden gebruikt, verschillen nauwelijks van de hierboven genoemde schriftelijke toetsen (ook wel eens als pen-en-papier toetsen aangeduid). De leerling leest de opgave en typt (bij open vragen) of selecteert (bij gesloten vragen) het antwoord. Digitaal toetsen biedt veel mogelijkheden zoals het laten voorlezen van opgaven en het geven van feedback (wel/niet goed geantwoord). Voordeel van digitaal toetsen is vooral de tijdsbesparing voor de leerkracht: de computer verzorgt de afname, neemt het nakijken en scoren over (bij gesloten vragen) en rapporteert direct na de afname. Een speciale vorm van digitaal toetsen is adaptief toetsen. Bij adaptieve toetsen bepaalt de computer op basis van het antwoord op een voorafgaande opgave welke opgave vervolgens aan de leerling zal worden aangeboden. Beantwoordt een leerling een opgave fout, dan krijgt hij een wat gemakkelijkere opgave aangeboden. Geeft hij een goed antwoord, dan volgt een opgave die iets moeilijker is. De keuze van de opgave wordt dus afgestemd op het niveau van de leerling. Zo krijgt iedere leerling een toets op maat en wordt hij uitgedaagd op zijn eigen niveau. Bovendien zijn in vergelijking met een lineaire digitale toets, die eenzelfde set opgaven bevat voor alle leerlingen, bij een adaptieve digitale toets minder opgaven nodig om nauwkeurig te meten hoe vaardig een leerling is. Dit komt omdat de opgaven optimaal aansluiten bij de vaardigheid van een leerling.

Van bijvoorbeeld de meeste LVS-toetsen van Cito is inmiddels naast een papieren versie ook een digitale variant beschikbaar, zie www.cito.nl > Primair en speciaal onderwijs > LVS-toetsen > Digitaal toetsen.

3 Mondelinge toetsen

Deze toetsvorm wordt in het PO bijvoorbeeld gebruikt voor het toetsen van de technische leesvaardigheid. De leerling leest letters (productieve letterkennis), woorden of een tekst voor en de leerkracht registreert de fouten. Bij deze toetsen zijn er over het algemeen duidelijke voorschriften over wat als juist of onjuist dient te worden beoordeeld (woordjes die bijvoorbeeld letter voor letter worden verklankt en die niet worden samengevoegd worden als onjuist beoordeeld). Ook spreekvaardigheid kan bijvoorbeeld alleen mondeling getoetst worden. Om de prestaties bij dergelijke toetsing zo objectief mogelijk te kunnen beoordelen, is het belangrijk dat er een beoordelingsmodel bij de toets beschikbaar is om te kunnen bepalen welke antwoorden wel/niet goed gerekend worden.

4 Observatiesystemen

Observatiesystemen bieden houvast om iemand volgens vooraf gedefinieerde punten in een bepaalde situatie te beoordelen. Het kan gaan om gedrag waarbij wordt geobserveerd of een leerling dit wel of niet vertoont, maar de beoordeling kan ook plaatsvinden op basis van een beoordelingsschaal (bijvoorbeeld nooit, zelden, soms, geregeld en vaak). Belangrijk is dat de observatiepunten zo objectief mogelijk zijn geformuleerd zodat verschillende observatoren tot eenzelfde oordeel kunnen komen (intersubjectiviteit van de beoordeling). Sommige observatiesystemen zijn bedoeld om in te vullen op basis van het gedrag in een natuurlijke situatie, zoals de TRF (Teacher Report Form) waarbij leerkrachten vragen beantwoorden over het schoolwerk van leerlingen, hun functioneren en emotionele en gedragsproblemen. Andere observatiesystemen zijn gekoppeld aan specifiek 'uit te lokken' situaties. Zo zijn er beoordelingsmodellen voor bijvoorbeeld het beheersen van praktische vaardigheden. De leerling voert een bepaalde opdracht uit waarbij de beoordelaar volgens het observatiesysteem scoort in welke mate de leerling de vaardigheid beheerst. Denk aan een opdracht rondom techniek (Onderzoekend en Ontwerpend Leren bij Natuur en Techniek, zie www.slo.nl > PO > leergebieden > oriëntatie op jezelf en de wereld > natuur en techniek > verbreding techniek basisonderwijs > downloads).

Welke toetsvorm geschikt is, is afhankelijk van wat we willen meten (de vaardigheid), het doel waarom we toetsen (wordt de toets bijvoorbeeld gebruikt om te selecteren) en de kenmerken van de toets (uit hoeveel opgaven mag de toets bijvoorbeeld bestaan, moet deze groepsgewijs afgenomen kunnen worden etc.). Bij het zoeken naar een passende toetsvorm wordt gezocht naar een balans tussen de validiteit (meet de toets wat we willen meten), betrouwbaarheid (mate waarin we de resultaten kunnen vertrouwen) en de wensen van de gebruiker. Voor belangrijke beslissingen, bijvoorbeeld zakken of slagen, moeten we kunnen vertrouwen op de toetsresultaten en zijn de eisen aan de betrouwbaarheid en validiteit van een toets hoog. Dat is anders als het gaat om het gebruik van bijvoorbeeld methodegebonden toetsen voor de dagelijkse bijsturing van het onderwijsleerproces. Een hoge betrouwbaarheid is bij dergelijke toetsen minder belangrijk.

Toetsen op School op internet

Dit boek wordt ondersteund door een themasite op de website Toetswijzer, www.toetsenopschool.nl. Deze site bevat de elektronische versies van de verschillende uitgaven van Toetsen op School en aanvullende informatie bij dit boek.

1 Het doel van toetsen

1 Het doel van toetsen

Marleen van der Lubbe en Judith Hollenberg

Zoals in hoofdstuk 1 van het algemene deel van Toetsen op School besproken is, kunnen ook in het PO toetsen voor verschillende doelen worden ingezet. In dit hoofdstuk komen in paragraaf 1.1 toetsen aan de orde die tot doel hebben leerlingen te beoordelen, dat wil zeggen leerlingen te selecteren, te classificeren, te plaatsen en te certificeren.

Toetsen die tot doel hebben leerlingen te classificeren, zijn bijvoorbeeld de verschillende eindtoetsen in het basisonderwijs die bedoeld zijn leerlingen een onafhankelijk advies te geven over het best passende type voortgezet onderwijs. Soms worden toetsen ook ingezet voor selectiedoeleinden, bijvoorbeeld in het kader van de toelating van een leerling tot één van de vormen van (voortgezet) speciaal onderwijs (op grond van landelijk bepaalde indicatiecriteria). Daarnaast worden toetsresultaten soms gebruikt om leerlingen te plaatsen. Het gaat dan niet zozeer om het volgen van verschillende programma's die tot hetzelfde diploma of certificaat leiden. Immers het PO wordt niet afgesloten met een diploma of certificaat. Maar we zouden wel over plaatsen kunnen spreken als we binnen een groep daadwerkelijk differentiëren in leerstofaanbod (bijvoorbeeld verschillende niveaugroepen) op basis van toetsresultaten. Van certificeren is in het PO geen sprake: alle leerlingen stromen na het primair onderwijs door naar een vorm van voortgezet onderwijs ongeacht de door hen behaalde toetsresultaten.

In paragraaf 1.2 bespreken we toetsen die tot doel hebben het verloop van het onderwijs-leerproces te beoordelen en op basis van de verzamelde informatie het onderwijs eventueel aan te passen. In het PO worden hier bijvoorbeeld de LVS-toetsen van het Cito Volgsysteem primair en speciaal onderwijs voor ingezet. De resultaten van individuele leerlingen op deze toetsen kunnen vervolgens op groeps- en schoolniveau worden samengenomen om de onderwijsprestaties van groepen leerlingen en scholen in kaart te brengen. Dit wordt in paragraaf 1.3 besproken aan de hand van de verschillende rapportages die gekoppeld zijn aan de toetsen van het Cito Volgsysteem primair en speciaal onderwijs. Ten slotte bespreken we in paragraaf 1.4 op welke wijze toetsen in het PO ingezet worden om de kwaliteit van het onderwijs te beoordelen, onder andere op basis van het peilingsonderzoek in het (speciaal) basisonderwijs.

1.1 Toetsen op leerlingniveau

In het PO is, zoals hierboven aangegeven, vooral sprake van toetsen die tot doel hebben leerlingen te classificeren. Maar ook het gebruik van toetsen voor selectiedoeleinden en ten behoeve van plaatsing komt aan de orde. Omdat er in het PO geen toetsen zijn op basis waarvan beslist wordt of leerlingen een onderwijsprogramma al dan niet met succes afgesloten hebben, komt het certificeren van leerlingen in dit hoofdstuk niet aan de orde.

1.1.1 Het selecteren van leerlingen

Zoals in hoofdstuk 1 van het algemene deel van Toetsen op School aangegeven is, spreken we van selectie als niet alle leerlingen die zich voor een opleiding of onderwijsprogramma aanmelden, toegelaten worden.

In strikte zin is er van selectie geen sprake in het PO: immers alle leerlingen volgen (een vorm van het) primair onderwijs. Alleen bij het gebruik van toetsresultaten om te bepalen welke leerlingen toelaatbaar zijn tot het (voortgezet) speciaal onderwijs is sprake van selectie.

Tot 1 augustus 2014 bepaalde een speciaal ingestelde commissie, de Commissie voor de Indicatiestelling (Cvl), of een kind in aanmerking kwam voor een indicatie (voortgezet) speciaal onderwijs. Deze Cvl's maakten onderdeel uit van een regionaal expertisecentrum (REC) dat regionaal zorg droeg voor speciale onderwijszorg voor de leerlingen die daarop zijn aangewezen. De Cvl's waren onafhankelijk van de verschillende scholen voor speciaal onderwijs en overwogen hun beslissing over de indicatie aan de hand van landelijk geldende criteria die verschilden voor de diverse schoolsoorten binnen het speciaal onderwijs. Voor de meeste schoolsoorten was één van die criteria het aantonen van een leerachterstand met behulp van toetsresultaten voor bijvoorbeeld spellen, begrijpend lezen en rekenen. Een ander veelgebruikt criterium was de score op een intelligentietest.

Met de invoering van passend onderwijs per 1 augustus 2014 is de Cvl-indicatiestelling verdwenen. Alle scholen van de voormalige clusters 3 (kinderen met verstandelijke en/of lichamelijke beperkingen en langdurig zieke kinderen) en 4 (kinderen met ernstige gedragsstoornissen, ontwikkelingsstoornissen of een psychiatrisch probleem) van het speciaal onderwijs zijn samen met reguliere basisscholen onderdeel van een regionaal samenwerkingsverband¹. Binnen dit samenwerkingsverband spreken de scholen zelf af welke leerlingen ze doorverwijzen naar het speciaal onderwijs. Er is dus niet langer sprake van landelijke criteria voor toelating tot deze schoolsoorten; elk samenwerkingsverband bepaalt de eigen systematiek. Voor leerlingen die volgens deze systematiek toelaatbaar zijn tot het speciaal onderwijs, geeft het samenwerkingsverband een zogenaamde 'toelaatbaarheidsverklaring' (tlv) af. Zo'n verklaring geeft recht op een plek in het speciaal onderwijs. Er zijn samenwerkingsverbanden die de criteria, zoals voorheen door de Cvl gehanteerd, als leidraad gebruiken bij het afgeven van een toelaatbaarheidsverklaring.

Toelating tot een instelling in cluster 1 (blinde, slechtziende kinderen) en cluster 2 (dove, slechthorende kinderen en kinderen met ernstige spraakmoeilijkheden) gebeurt wel landelijk door een zogenaamde Commissie van Onderzoek (voor elk van deze clusters één). Deze commissie beoordeelt aan de hand van criteria of een leerling is aangewezen op onderwijs op een speciale instelling of op ondersteuning vanuit een dergelijke instelling in het regulier onderwijs. De commissies bepalen zelf welke criteria gehanteerd worden. Bijvoorbeeld bij het vaststellen of er sprake is van een meervoudig gehandicapte dove leerling vormt het gebruik van resultaten op een intelligentietest onderdeel van de indicatiecriteria van de Commissie van Onderzoek (in dit geval van cluster 2²). Naast een criterium voor de gehoorstoornis, dient op basis van psychodiagnostisch onderzoek (dat individueel is afgenomen en rekening houdt met de kenmerken van de leerling) bij deze leerlingen een intelligentiequotiënt lager dan 70 te zijn vastgesteld. Ook wordt bij bijvoorbeeld de indicatiecriteria voor leerlingen met een taalontwikkelingsstoornis (TOS) mede gebruik gemaakt van leerachterstanden als indicatie voor een ernstige structurele beperking in de onderwijsparticipatie. Het gaat dan om de prestaties van de leerling op het gebied van lezen, spellen en rekenen. Deze prestaties dienen op tenminste twee van deze drie gebieden te behoren tot de 10% zwakste in vergelijking met andere leerlingen van dezelfde didactische leeftijd. Zowel bij het bepalen van de leerachterstand als de intelligentie is hier dus sprake van het gebruik van toetsresultaten voor selectie-

1 Zie swv.passendonderwijs.nl voor een overzicht van alle samenwerkingsverbanden

2 Zie de brochure Indicatiecriteria auditief en/of communicatief beperkte leerlingen van Simea op www.simea.nl > dossiers > brochure indicatiecriteria

doeleinden. Voor het bepalen van leervorderingen en intelligentie kan gebruik gemaakt worden van diverse instrumenten, zie voor een overzicht Toetsgids (onderdeel van www.toetswijzer.nl).

1.1.2 Het classificeren van leerlingen

In hoofdstuk 1 van het algemene deel van Toetsen op School wordt van classificeren van leerlingen gesproken als leerlingen afhankelijk van de resultaten op een toets verschillende programma's volgen en die onderwijsprogramma's tot verschillende diploma's of certificaten leiden.

In het PO is sprake van één classificatiemoment, namelijk in het laatste leerjaar. Naast het schooladvies van de basisschool is het sinds schooljaar 2014/2015 voor alle schoolverlaters verplicht een eindtoets te maken die hun taal- en rekenvaardigheid meet. Scholen bepalen zelf welke eindtoets zij gebruiken. Voor de toelating tot het voortgezet onderwijs is het schooladvies van de basisschool leidend. De score op de eindtoets is een 'second opinion', een onafhankelijk tweede gegeven. Indien een leerling op de eindtoets hoger scoort dan verwacht, dan moet de basisschool het schooladvies heroverwegen. De basisschool kan dan het schooladvies naar boven toe bijstellen (maar dit hoeft niet). Het nieuwe advies geldt dan als leidend voor de toelating. Voor meer informatie over de verplichte eindtoets en de consequenties voor de overgang van primair naar voortgezet onderwijs met ingang van schooljaar 2014/2015, zie www.vanponaarvo.nl

Vanaf schooljaar 2015/2016 mogen scholen voor de afname van de eindtoets alleen nog gebruik maken van een door de minister van Onderwijs, Cultuur en Wetenschap (OCW) toegelaten eindtoets. Op dit moment betreft dit de volgende toetsen:

- Centrale Eindtoets PO: in opdracht van het ministerie van OCW brengt het College voor Toetsen en Examens (CvTE) de Centrale Eindtoets PO uit. De toets wordt in samenwerking met Stichting Cito gemaakt.
- ICE Eindevaluatie Primair onderwijs (IEP Eindtoets): deze eindtoets wordt uitgebracht door bureau ICE.
- ROUTE 8: deze eindtoets wordt uitgebracht door A-VISION B.V.
- Dia-Eindtoets: deze eindtoets wordt uitgebracht door Diataal.
- CESAN: deze eindtoets wordt uitgebracht door SM&C Internet Services.
- AMN Eindtoets: deze eindtoets wordt uitgebracht door AMN.

Naast deze eindtoetsen laat de minister mogelijk nog andere eindtoetsen toe. Het ministerie maakt eind november van elk jaar bekend welke toetsen zijn toegelaten. Alle eindtoetsen worden in de periode tussen 15 april en 15 mei afgenomen.

Hieronder worden de drie eindtoetsen met de meeste afnamen nader toegelicht.

Centrale Eindtoets PO

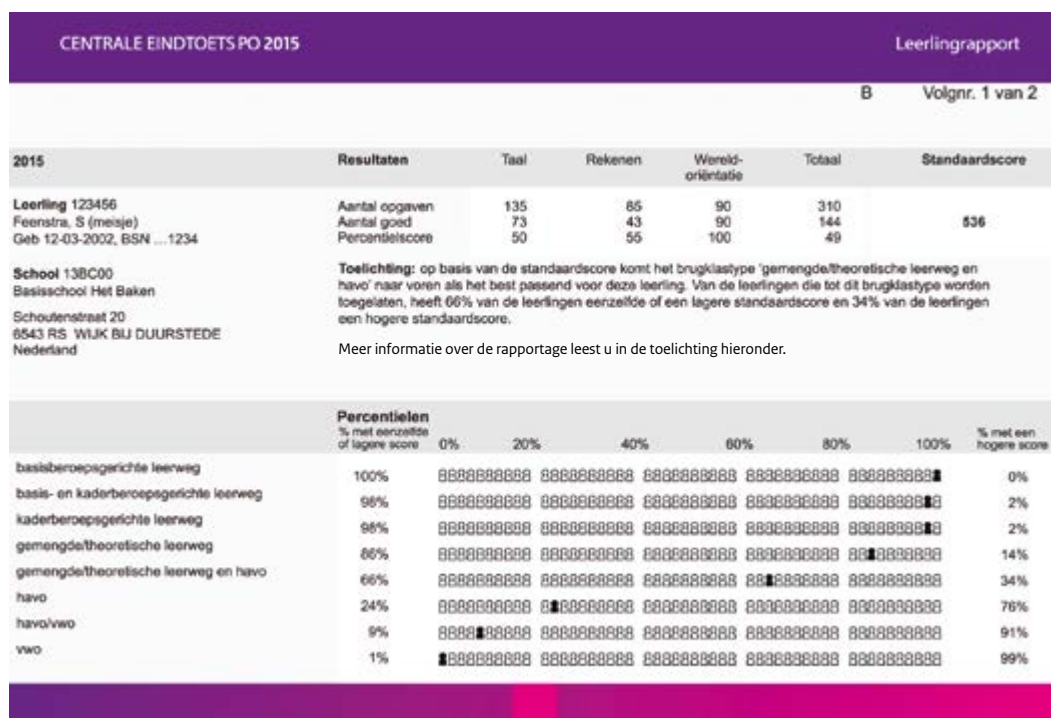
De Centrale Eindtoets PO bouwt voort op de Eindtoets Basisonderwijs die Cito tot schooljaar 2014/2015 uitbracht. De Centrale Eindtoets meet wat een leerling in vergelijking met andere leerlingen in acht jaar basisonderwijs geleerd heeft op het gebied van Taal (135 opgaven), Rekenen-Wiskunde (85 opgaven) en -facultatief- Wereldoriëntatie (90 opgaven). Als onderdeel van taal worden de domeinen lezen, woordenschat, schrijven en taalverzorging getoetst. Voor rekenen betreft dit getallen, verhoudingen, meten en meetkunde en verbanden. Voor wereldoriëntatie gaat het om ruimte (aardrijkskunde), tijd (geschiedenis) en natuur en techniek. Alle vragen zijn meerkeuzevragen; de antwoorden worden op een antwoordblad genoteerd. De toets wordt op drie ochtenden afgenomen. Een school kan ervoor kiezen de Centrale Eindtoets PO digitaal af te nemen.

Van de Centrale Eindtoets PO wordt elk jaar een nieuwe versie gemaakt. Op basis van het totaal aantal goed op de onderdelen Taal en Rekenen-Wiskunde wordt de zogenaamde standaard-score berekend. Dit is een getal tussen de 501 en 550. Omdat de opgaven van de Centrale Eindtoets elk jaar weer helemaal nieuw zijn, is de toets niet ieder jaar precies even moeilijk. Daardoor kunnen de prestaties in de verschillende jaren via de 'aantallen goed' niet direct vergeleken worden. Via een omzetting naar de standaardscores kan dat wel. De standaardscores liggen op een schaal waarnaar de 'aantallen goed' op de verschillende jaargangen van de toets te vertalen zijn. Ongeveer net zoals bedragen uitgedrukt in de vroegere Europese munteenheden (bijvoorbeeld guldens, Duitse marken en Franse francs) via de euroschaal gemakkelijk te vergelijken zijn. Door de standaardscoreschaal zijn we in staat de scores van jaar tot jaar te vergelijken ondanks verschillen in moeilijkheid tussen de diverse jaargangen van de toets.

De standaardscore vormt de basis voor de advisering van het best passende brugklatype voortgezet onderwijs. De interpretatie van de standaardscore is mede gebaseerd op het zogenaamde Toelatings- en doorstroomonderzoek Eindtoets Basisonderwijs dat periodiek wordt uitgevoerd. In dit onderzoek worden de gegevens verzameld van leerlingen die in een bepaald jaar aan de eindtoets hebben deelgenomen en wordt nagegaan tot welk brugklatype deze leerlingen toegelaten zijn en naar welke type tweede en derde leerjaar leerlingen doorstromen. Deze gegevens vormen de basis voor de interpretatie van de standaardscore. Voor de Centrale Eindtoets PO 2015 vormde de volgende tabel het uitgangspunt voor de interpretatie van de standaardscore:

Brugklatype	Interval
basisberoepsgerichte leerweg	501-520
basis- en kaderberoepsgerichte leerweg	519-525
kaderberoepsgerichte leerweg	523-528
gemengde / theoretische leerweg	529-533
gemengde / theoretische leerweg en havo	533-536
havo	537-540
havo / vwo	540-544
vwo	545-550

De betekenis van de standaardscore wordt grafisch weergegeven in de zogenoemde poppetjes-grafiek op het Leerlingrapport van de Centrale Eindtoets PO (zie figuur 1.1). In het bovenste gedeelte van het rapport is per onderdeel en voor het totaal het aantal opgaven, het aantal goed en de percentielscore weergegeven. De percentielscore geeft aan welk percentage leerlingen landelijk gezien een even hoog of lager aantal goed heeft behaald. De percentielscore van de betreffende leerling, Sanne Feenstra, voor het totaal is 49. Dat betekent dat 49% van de leerlingen eenzelfde of lagere score heeft gehaald en 51% van de leerlingen een hogere score.



Figuur 1.1: Voorbeeld leerlingrapport centrale eindtoets

Helemaal rechts is de standandscore opgenomen, in dit geval 536. In de poppetjesgrafiek onderaan wordt op basis van de behaalde standandscore weergegeven welke positie een leerling zou innemen tussen zijn medeleerlingen in de verschillende brugklatypes.

In de poppetjesgrafiek op het leerlingrapport van Sanne zien we dat zij met haar score behoort tot de hoogst scorende leerlingen in de basisberoepsgerichte leerweg. Van de leerlingen die zijn toegelaten tot een brugklas van dit type heeft niemand een hogere score dan 536. Als Sanne in dit brugklatype zou worden geplaatst zou er van onderschatting sprake kunnen zijn. Aan de andere kant lijkt de kans op overschatting van de mogelijkheden van Sanne bij plaatsing in het brugklatype 'havo' tamelijk groot: slechts 24% van de leerlingen heeft een score van 536 of lager; 76% scoort hoger dan Sanne. Het brugklatype 'gemengde/theoretische leerweg en havo' lijkt voor Sanne op basis van haar score op de centrale eindtoets het meest geschikt: hier behoort zij tot de middenmoot.

Naast het in figuur 1.1 getoonde leerlingrapport ontvangen scholen bij de Centrale Eindtoets PO ook een rapportage waarin voor rekenen, lezen en taalverzorging wordt aangegeven op welk referentieniveau de leerling deze onderdelen beheerst³. Voor meer informatie over de Centrale Eindtoets PO zie www.centraleeindtoetspo.nl.

IEP Eindtoets

De IEP Eindtoets meet de vaardigheden taal (95 opgaven) en rekenen (65 opgaven). Bij taal betreft het de onderdelen lezen en taalverzorging. Bij rekenen gaat het om de onderdelen getallen, verhoudingen, meten en meetkunde en verbanden. De toets wordt afgenomen op

³ De referentieniveaus beschrijven wat kinderen op verschillende momenten in hun schoolloopbaan op het gebied van taal en rekenen moeten kennen en kunnen. Voor het einde van de basisschool zijn twee referentieniveaus van belang, te weten het fundamentele niveau 1F en het streefniveau 2F (voor taal) of 1S (voor rekenen). Zie voor meer informatie over de referentieniveaus paragraaf 2.2. van deze uitgave.

twee ochtenden van maximaal twee uur. De opgaven hebben de vorm van een meerkeuzevraag of een korte-open-antwoord-vraag. Omdat de toets ook open vragen bevat, noteren de leerlingen de antwoorden in het toetsboekje. Vooralsnog kan de toets alleen op papier worden afgenomen.

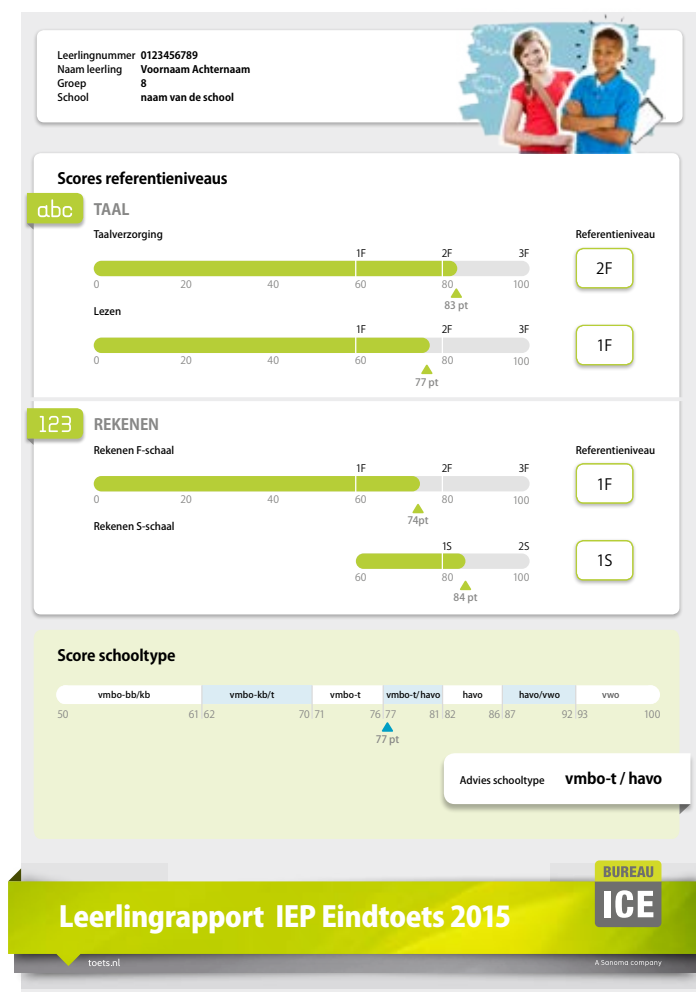
Op basis van de ruwe score per vaardigheid wordt vastgesteld wat het vaardigheidsniveau van een leerling is voor de onderdelen lezen, taalverzorging en rekenen. Dit niveau wordt gerapporteerd door (met behulp van een genormeerde score) aan te geven op welk referentieniveau de leerling in groep 8 deze onderdelen beheerst. Voor het uitbrengen van een schooladvies worden de (gewogen) vaardigheidsscores van de leerlingen op de onderdelen (lineair) omgezet naar een standaardscore waarbij rekening wordt gehouden met de moeilijkheidsgraad van de toets. De standaardscore op de IEP Eindtoets is een score op een schaal van 50 tot en met 100 en wordt schooladviescore genoemd. Het schooladvies op de IEP Eindtoets zal in de toekomst worden gebaseerd op longitudinaal doorstroomonderzoek. Omdat de IEP Eindtoets een nieuw ontwikkelde eindtoets is, is de dataverzameling daarvoor pas in schooljaar 2014/2015 gestart. Voor het bepalen van de schooladviezen op de IEP Eindtoets die in schooljaar 2014/2015 is afgenomen, is gebruik gemaakt van een normgerichte interpretatie waarbij de score is vergeleken met de score van meerdere referentiegroepen aan het begin van het eerste leerjaar in het voortgezet onderwijs. Dit heeft geresulteerd in de volgende indeling:

Leerweg	Standaardscorebereik
vmbo-b en vmbo k/b	50 – 61
vmbo-k en vmbo-k/t	62 – 70
vmbo-t	71 – 76
vmbo-t/havo	77 – 81
havo	82 – 86
havo-vwo	87 – 92
vwo	93 – 100

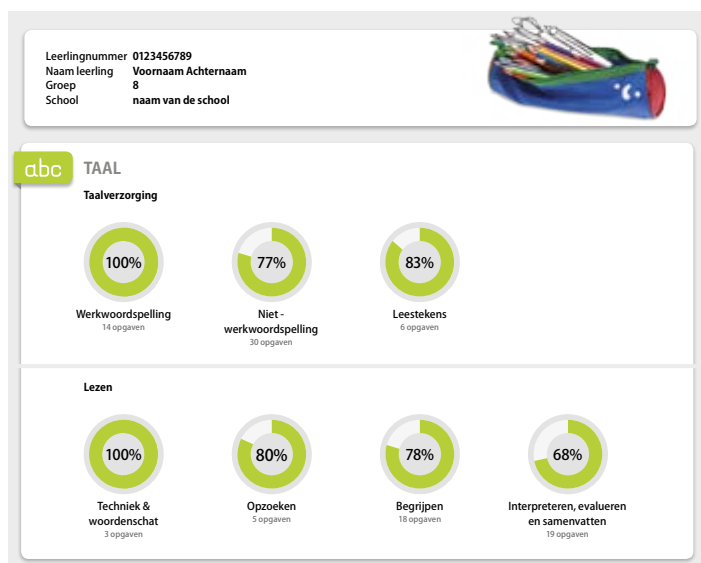
In figuur 1.2 is een voorbeeld opgenomen van een leerlingrapport bij de IEP Eindtoets. Na het blokje met basisgegevens over de leerling, wordt in het bovenste gedeelte van het leerlingrapport aangegeven op welk referentieniveau de leerling de onderdelen taalverzorging, lezen en rekenen beheerst. In de voorbeeldrapportage is te zien dat deze leerling voor taalverzorging niveau 2F beheerst en voor lezen al wel op weg is naar niveau 2F, maar dit nog niet beheerst. In de praktijk betekent dit dat de leerling niveau 1F beheerst en daarnaast een aantal (maar niet alle) deelvaardigheden van niveau 2F. Voor rekenen wordt naast het fundamentele niveau (F-niveau) ook het streefniveau (S-niveau) gemeten. De fundamentele niveaus van rekenen richten zich op basale kennis en inzichten en zijn gericht op een meer toepassinggerichte benadering van rekenen. De streefniveaus bereiden al voor op de meer abstracte wiskunde. In het voorbeeld is zichtbaar dat deze leerling voor rekenen de niveaus 1F en 1S beheerst. In het onderste gedeelte van het leerlingrapport is de schooladviescore opgenomen. Voor deze leerling is dat 77 met bijbehorend advies vmbo-t/havo.

Op de achterzijde van het leerlingrapport is opgenomen uit hoeveel opgaven de verschillende onderdelen bestaan. Daarnaast is zichtbaar hoeveel van deze opgaven de betreffende leerling goed beantwoord heeft. In figuur 1.3 is een gedeelte van de achterzijde van het leerlingrapport van de IEP Eindtoets opgenomen. In dit voorbeeld is onder meer zichtbaar dat het onderdeel niet-werkwoordspelling uit 30 opgaven bestaat en dat de leerling 77% daarvan (23 opgaven) goed heeft beantwoord.

Voor meer informatie over de IEP Eindtoets zie www.toets.nl > ipeindtoets



Figuur 1.2: Voorbeeld voorzijde leerlingrapport IEP Eindtoets



Figuur 1.3: Gedeelte achterzijde leerlingrapport IEP Eindtoets

ROUTE 8

ROUTE 8 is een digitale, adaptieve eindtoets die op de computer, laptop of tablet via internet wordt afgenomen. Aangezien de toets adaptief is en zich aanpast aan het niveau van de leerling is de afnameduur en het aantal opgaven dat elke leerling maakt verschillend. De afnameduur van de toets ligt tussen de 2 à 3 uur. ROUTE 8 meet de vaardigheden taal, rekenen en optioneel ook functioneren. Deze onderdelen bestaan uit de volgende deelttoetsen:

- Taal: luistervaardigheid, leesvaardigheid, taalverzorging, woordenschat en begrippenlijst
- Rekenen: getallen, verhoudingen, meten & meetkunde en verbanden
- Functioneren (optioneel): zelfconcept en werkhouding

In figuur 1.4 is een voorbeeld opgenomen van het scorerapport van ROUTE 8 voor de onderdelen taal en rekenen. Over het optionele deel functioneren wordt apart gerapporteerd (hier niet opgenomen).



Figuur 1.4: Voorbeeld scorerapport ROUTE 8

In het scorerapport staat de score voor alle deoltoetsen, net als de totaalscore voor de domeinen taal en rekenen (de oranje lijnen in het scorerapport). Alle scores worden weergegeven op een schaal van 100-300 en gemarkeerd met een rode cirkel. Voor de onderdelen taalverzorging, leesvaardigheid en rekenen wordt bovendien aangegeven welk referentieniveau de leerling beheerst. In figuur 1.4 is bijvoorbeeld zichtbaar dat de score van Cato op taalverzorging 262 is en dat dit betekent dat zij voor dit onderdeel niveau 2F beheerst.

De scores van de afzonderlijke deoltoetsen worden vervolgens samengenomen tot een adviescore (A-score) die eveneens op een schaal van 100-300 wordt gerapporteerd. Voor de advisering wordt de volgende indeling gehanteerd:

Advies	Scorebereik
Praktijkonderwijs (PrO)	<141
Basisberoepsgerichte leerweg (BBL)	141 –168
Kaderberoepsgerichte leerweg (KBL)	169 –190
Theoretische leerweg (TL)	191 –210
HAVO	211 –234
VWO	>234

In het voorbeeld is zichtbaar dat Cato met haar totaalscore van 212 op de toets het advies havo krijgt. Voor meer informatie over ROUTE 8, zie www.route8.nl.

1.1.3 Het plaatsen van leerlingen

Zoals in hoofdstuk 1 van het algemene deel van Toetsen op School aangegeven is, spreken we van plaatsen van leerlingen als leerlingen op basis van de resultaten op een toets verschillende onderwijsprogramma's volgen, maar die verschillende programma's wel tot hetzelfde diploma of certificaat leiden.

In het PO is geen sprake van het plaatsen van leerlingen zoals hierboven beschreven. Maar we zouden wel over plaatsen kunnen spreken als we binnen een groep daadwerkelijk differentiëren in leerstofaanbod op basis van door leerlingen behaalde toetsresultaten. Dit is het uitgangspunt van bijvoorbeeld het Onderwijscontinuüm van de CED-groep, een werkwijze waarmee schoolteams, directieleden en schoolbesturen datagestueerd en opbrengstgericht kunnen werken⁴. Bij het Onderwijscontinuüm wordt uitgegaan van een beperkt aantal onderwijsarrangementen in één groep. Het passend maken van het onderwijs voor alle individuele leerlingen blijkt qua uitvoering binnen de klas namelijk nauwelijks uitvoerbaar te zijn; het is realistisch om uit te gaan van drie, maximaal vier instructiegroepen in een groep. Daarom formuleert een school die werkt volgens het Onderwijscontinuüm een basisaanbod voor alle leerlingen. Dit basisaanbod moet voor de meeste leerlingen, zo'n 75%, voldoende zijn om de gewenste resultaten te halen. Voor de leerlingen voor wie op basis van de behaalde toetsresultaten (uiteraard in vergelijking met de gestelde doelen) het basisaanbod niet voldoende blijkt, wordt er naast een basisarrangement een van de andere arrangementen aangeboden. Het verdiepte arrangement is bedoeld voor de ca. 25% leerlingen die bovenop het basisarrangement extra uitdaging kunnen gebruiken. Het intensieve arrangement is bedoeld voor de ca. 15% leerlingen die extra oefening en begeleiding nodig hebben om de geformuleerde

4 Het Onderwijscontinuüm is vervolgens weer vertaald in diverse praktische hulpmiddelen, instrumenten en publicaties onder de noemer Opbrengstgericht werken 4D: Data, Duiden, Doelen, Doen (zie <http://www.opbrengstgerichtwerken4d.nl/>).

minimumdoelen te halen. En het zeer intensieve arrangement (plus zorg) is bedoeld voor de allerzwakste ca. 10% van de leerlingen die de minimumdoelen naar verwachting niet zullen behalen.

Op deze manier kunnen resultaten op toetsen, zoals de LVS-toetsen van Cito, gebruikt worden om binnen een groep de leerinhouden indien nodig te differentiëren. Door vooraf de grenzen van de arrangementen te definiëren in termen van toetsscores, kunnen leerlingen op basis van hun toetsscores worden geplaatst in het (zeer) intensieve, basis- of verdiepte arrangement. Er is daarbij geen sprake van verschillende diploma's of certificaten aan het einde van de rit; alle leerlingen stromen na het primair onderwijs door naar een vorm van voortgezet onderwijs.

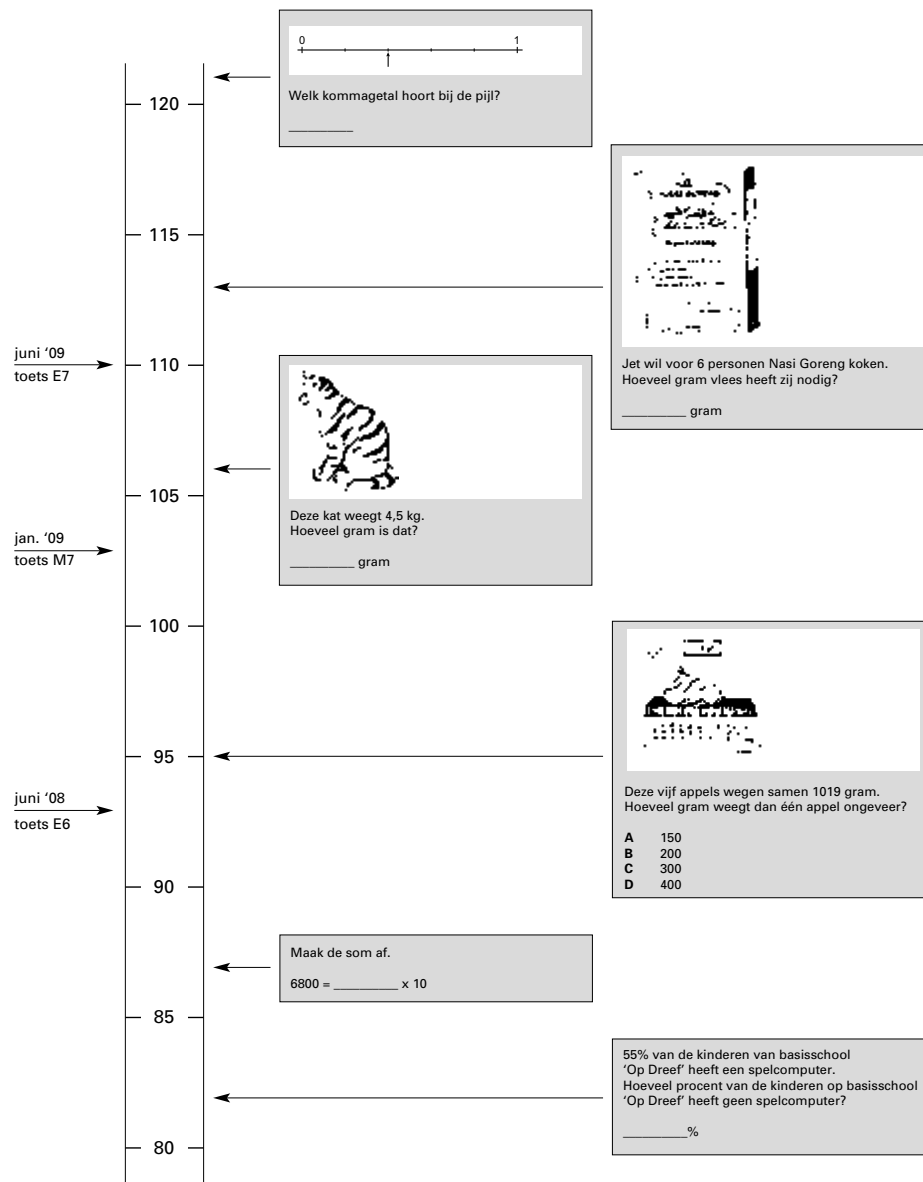
1.2 Toetsen voor het beoordelen van het onderwijsleerproces

Toetsen voor het beoordelen van het onderwijsleerproces hebben tot doel de leerkracht te informeren over het verloop van het leerproces bij de leerlingen. Op basis van die informatie kan het onderwijs aan individuele leerlingen of groepen leerlingen indien nodig worden aangepast.

In het PO wordt voor dit doel veel gebruik gemaakt van methodegebonden toetsen (zie ook de Inleiding). Methodegebonden toetsen hebben over het algemeen betrekking op de leerinhouden die in de afgelopen les of lessen aan de orde zijn geweest. Op basis van de resultaten op deze toetsen kan het onderwijs indien nodig bijgesteld worden zodat het beter aansluit bij de behoeften van de leerling.

Naast methodegebonden toetsen wordt er in het PO veel gebruik gemaakt van methode-onafhankelijke toetsen. Dit zijn toetsen die onafhankelijk zijn van de gebruikte lesmethode. Met deze toetsen kan op vaste momenten in het schooljaar, over het algemeen één of twee keer per schooljaar, op een betrouwbare manier en aan de hand van objectieve standaarden informatie verzameld worden over de vaardigheid van leerlingen op basis van grotere leerstofeenheden dan bij methodegebonden toetsen. In het PO zijn de meest gebruikte methodeonafhankelijke toetsen de LVS-toetsen van Cito. Deze toetsen maken het niet alleen mogelijk de vaardigheid van een leerling vast te stellen, maar ook om de ontwikkeling van deze vaardigheid gedurende de schoolloopbaan te volgen. Dit is mogelijk doordat de scores op de verschillende LVS-toetsen voor eenzelfde vaardigheid 'vertaald' kunnen worden naar dezelfde onderliggende meetschaal, de vaardigheidsschaal.

Op de volgende bladzijde is een voorbeeld opgenomen van een gedeelte van de vaardigheidsschaal Rekenen-Wiskunde.



Figuur 1.5: Vaardigheidsschaal Rekenen-Wiskunde

In de figuur is aan de rechterkant van de schaal een aantal opgaven uit de LVS-toetsen Rekenen-Wiskunde opgenomen. Met behulp van opgaven op deze schaal kan de vaardigheid van leerlingen worden gemeten. De getallen van de schaal geven aan hoe moeilijk de opgaven zijn. Hoe hoger de opgaven op de schaal liggen, en hoe groter de getallen op de schaal zijn, hoe moeilijker de opgaven zijn en hoe meer vaardigheid een leerling nodig heeft om de opgaven succesvol op te kunnen lossen. Aan de linkerkant van de schaal is de score van een leerling op verschillende LVS-toetsen Rekenen-Wiskunde opgenomen. Deze toetsen zijn afgenomen aan het eind van groep 6 (E6), halverwege groep 7 (M7) en aan het eind van groep 7 (E7). Om de scores op deze qua inhoud, omvang en moeilijkheidsgraad van elkaar verschillende toetsen te kunnen vergelijken, wordt het aantal goed (de toetsscore) van een leerling omgezet in een getal op de rekenvaardigheidsschaal: de vaardigheidsscore. Hoe hoger de vaardigheidsscore, hoe groter de vaardigheid op dat onderdeel is. De rekenvaardigheidsschaal heeft dus een dubbele functie: de opgaven kunnen er naar moeilijkheid op worden gerangschikt en de vaardigheid van de leerlingen kan er op worden aangegeven. Daardoor kunnen we op de schaal zien wat

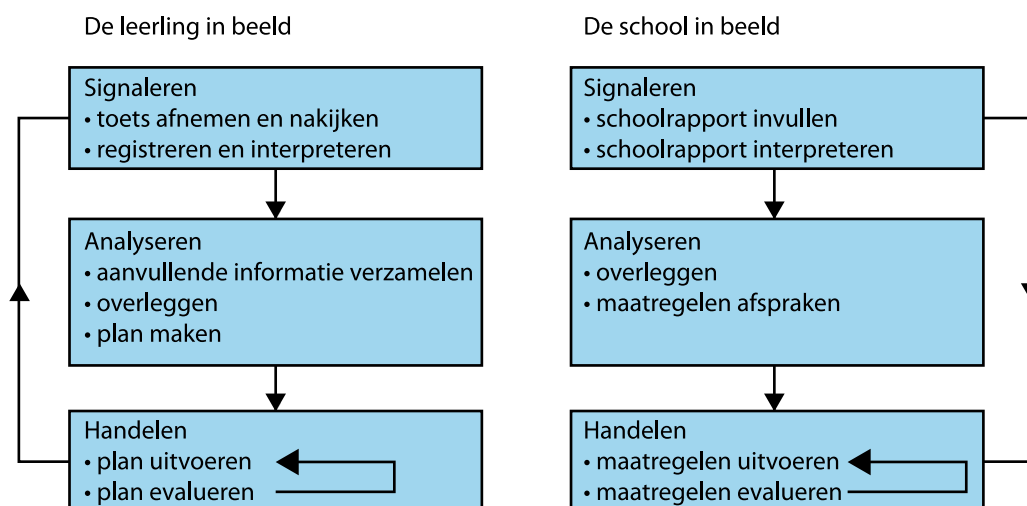
leerlingen met een bepaalde vaardigheid zoal kunnen. Van elke opgave is op de schaal af te lezen hoe vaardig een leerling moet zijn om met een grote kans op succes die opgave goed te maken. Een grote kans op succes definiëren we als gemiddeld minstens acht van de tien opgaven van dit type goed maken, dat wil zeggen een kans van 80%. We spreken dan van goede beheersing.

Eind groep 6 beheerst de leerling in het voorbeeld alleen de onderste twee opgaven in de figuur goed; de vier andere opgaven worden op dat moment matig of onvoldoende beheerst. Een half jaar later, halverwege groep 7, is de vaardigheid van de leerling toegenomen en beheerst hij ook de opgave over het gewicht van de appels. Als voor elk afnamemoment ook de scores van een landelijke vergelijkingsgroep op de schaal zouden worden afgebeeld, kan ook de relatieve positie van de leerling in vergelijking met leeftijdgenoten worden bepaald. Op deze manier wordt het mogelijk de score van een leerling op drie manieren te interpreteren:

- Ten opzichte van zichzelf (self-referenced)
De mate waarin de leerling vooruit is gegaan kan worden vergeleken met de eerdere vorderingen van een leerling. Belangrijke vragen zijn dan: Is de mate van vooruitgang van de afgelopen periode te vergelijken met de mate van vooruitgang in daaraan voorafgaande periodes? Is er nog steeds sprake van groei of stagneren de vorderingen van de leerling?
- In vergelijking met leeftijd- of groepsgenoten (normgericht)
De vorderingen van de leerling kunnen worden vergeleken met de vorderingen van andere vergelijkbare leerlingen. Belangrijke vragen zijn dan: Hoe valt de vergelijking met de andere leerlingen uit het land over een langere periode uit? Wordt een eventuele achterstand steeds groter of zien we dat de leerling zijn achterstand inloopt?
- In relatie tot de te realiseren doelen (domein- of inhoudsgericht)
Op basis van de positie van de leerling op de vaardigheidsschaal kunnen conclusies worden getrokken over de beheersing van de betreffende vaardigheid. Ook kan worden nagegaan in hoeverre een leerling als hij zich in hetzelfde tempo blijft ontwikkelen, de gestelde tussen- en einddoelen bereikt.

Werken met de LVS-toetsen

Werken met de LVS-toetsen gaat in fasen, zie figuur 1.6. De eerste fase, het signaleren, start met het afnemen en nakijken van de toetsen. De resultaten daarvan worden verwerkt tot zogenaamde leerlingrapporten. Met dit rapport kan het vaardigheidsniveau en de ontwikkeling van een leerling worden vastgesteld. Wordt geconstateerd dat een leerling - en dat geldt zowel voor vaardige als minder vaardige leerlingen - onvoldoende vooruitgaat of een lage score haalt, dan kan worden vastgesteld op welk vlak er precies problemen zijn. Dit gebeurt door in de fase van het analyseren (zie ook hoofdstuk 3) met behulp van bijvoorbeeld diagnostische gesprekken, hulpboeken of materialen uit de methode, meer gegevens te verzamelen om het niveau van de leerling en zijn leerbehoeften verder in kaart te brengen. In de fase van het plannen die hierop volgt, bepaalt de leerkracht welke leerdoelen bij een leerling passen en op welke wijze hij met die leerling aan de doelen kan werken. Hierbij wordt de koppeling gemaakt van toetsresultaten naar de leerstof in de methode. Daarna volgt de fase van het handelen. In deze fase voert de leerkracht het handelingsplan uit waarbij gebruik kan worden gemaakt van aanwijzingen en suggesties in de handleiding bij een toets, in een apart hulpboek bij het pakket of in andere publicaties. Bij een volgende toetsafname kan dan worden bepaald welk effect deze actie heeft gehad (evalueren). Dan begint weer een nieuwe cyclus van signaleren, analyseren en handelen.



Figuur 1.6: Fasen van werken met LVS-toetsen

Naast het volgen van de voortgang van individuele leerlingen zijn er nog twee niveaus van gebruik van de LVS-toetsen: het niveau van de groep en het niveau van de school. Ook op deze niveaus wordt gewerkt volgens het principe signaleren, analyseren en handelen. Hieronder bespreken we een aantal rapportages bij de LVS-toetsen die informatie bieden op grond waarvan het onderwijs aan individuele leerlingen of groepen leerlingen indien nodig kan worden aangepast (toetsen voor het beoordelen van het onderwijsleerproces).

Deze rapportages kunnen handmatig worden gemaakt, maar verschillende computerprogramma's zoals het Computerprogramma LOVS⁵ van Cito, ParnasSys, Esis B en SchoolOAS bieden ook mogelijkheden om dergelijke rapportages op te vragen.

Op www.schoolaanzet.nl zijn handreikingen, kwaliteitskaarten en praktijkvoorbeelden te vinden die aanvullende informatie bieden over het werken met toetsresultaten⁶.

Leerlingrapport

Een leerlingrapport is een grafiek waarin het vaardigheidsniveau van een individuele leerling op meer afnamemomenten kan worden aangegeven tegen de achtergrond van de prestaties van een landelijke vergelijkingsgroep.

5 Een demoversie van het Computerprogramma LOVS waarmee zelf allerlei rapportages gemaakt kunnen worden, is te downloaden via www.cito.nl > Primair en speciaal onderwijs > Computerprogramma LOVS. Voorbeeldrapportages van ParnasSys zijn te vinden op www.parnassys.nl > Producten > ParnasSys > Presentaties (knop onderaan de pagina over ParnasSys). Voorbeeldrapportages van Esis zijn te zien op www.rovict.nl > Producten > ESIS. Hier staat bijvoorbeeld een video "Toetsoverzichten maken", met daarin verschillende rapportages. Meer informatie over de applicatie SchoolOAS van Dotcomschool is te vinden op www.dotcomschool.nl.

6 De handreikingen, kwaliteitskaarten en praktijkvoorbeelden zijn op www.schoolaanzet.nl te vinden onder het tabblad 'Thematische inhoud' door te selecteren op onderwijssector 'primair onderwijs', thema 'opbrengstgericht werken' en subthema 'gebruik data'.

In het voorbeeld van een leerlingrapport in figuur 1.7 is de landelijke vergelijkingsgroep in vijf niveaugroepen ingedeeld:

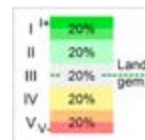
Niveau	%	Interpretatie
I	20%	Ver boven het gemiddelde
II	20%	Boven het gemiddelde
III	20%	De gemiddelde groep leerlingen
IV	20%	Onder het gemiddelde
V	20%	Ver onder het gemiddelde

Naast de niveau-indeling I t/m V, kan er in de rapportages bij de LVS-toetsen ook gebruik gemaakt worden van de niveau-indeling A t/m E (zie tabel hieronder). Deze indeling hoorde bij de eerste LVS-toetsen, maar wordt nog steeds door veel scholen gebruikt. De indeling I t/m V is door Cito geïntroduceerd omdat deze symmetrisch is opgebouwd, een gemiddelde groep bevat (niveau III) en aansluit bij de niveau-indeling van andere toetsinstrumenten van Cito. Voorlopig stelt Cito bij de LVS-toetsen beide niveau-indelingen beschikbaar.

Niveau	%	Interpretatie
A	25%	De 25% hoogst scorende leerlingen
B	25%	De 25% leerlingen die net boven tot ruim boven het landelijk gemiddelde scoren
C	25%	De 25% leerlingen die net onder tot ruim onder het landelijk gemiddelde scoren
D	15%	De 15% leerlingen die ruim onder het landelijk gemiddelde scoren
E	10%	De 10% laagst scorende leerlingen

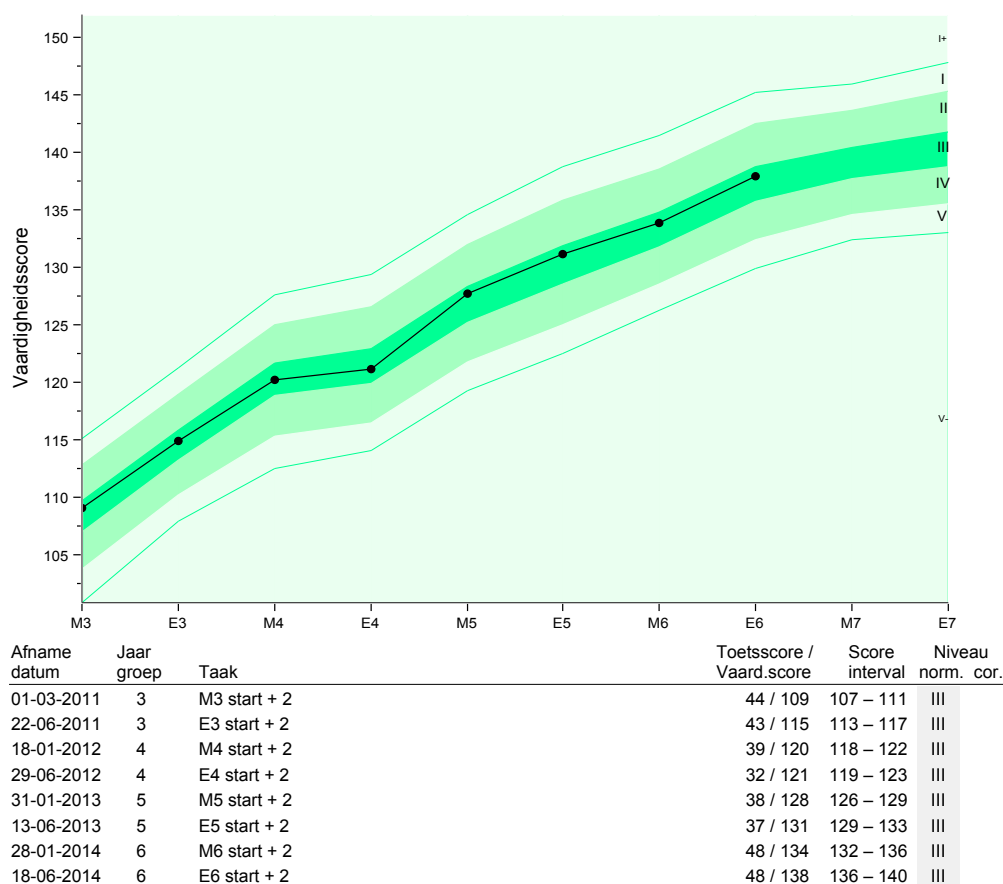
In figuur 1.7 staat een voorbeeld van een leerlingrapport van de LVS-toets Spelling. Dit leerlingrapport laat de ontwikkeling van de spellingvaardigheid van een leerling zien gedurende zijn schoolloopbaan. Die ontwikkeling wordt weergegeven door een lijn tussen de scores die de leerling op de verschillende toetsen Spelling behaald heeft. Deze lijn geeft aan hoe een leerling zich in een bepaalde periode heeft ontwikkeld. Hoe steiler de lijn, hoe groter de vooruitgang. Als de lijn nauwelijks stijgt, is er sprake van een stagnatie. Doordat op de leerlingrapporten de gegevens van de landelijke vergelijkingsgroep zijn vermeld (in dit voorbeeld de niveaus I t/m V), ontstaat een beeld van de vorderingen van de leerling in vergelijking met die van de vergelijkingsgroep. Op basis van de leerlingrapporten is het dus mogelijk de voortgang van individuele leerlingen te beoordelen ten opzichte van zichzelf maar ook in vergelijking met leeftijdgenoten.

In het voorbeeld in figuur 1.7 kunnen we zien dat Roland in groep 3 scores in niveau III behaalde en daarmee behoorde tot de gemiddelde groep leerlingen op het gebied van Spelling. Ook in groep 4, 5 en 6 scoort hij consistent in niveau III en ontwikkelt hij zich binnen dat niveau conform verwachting.



Leerlingrapport

Leerling: **Roland Schäfu (7 - 7B)**
 Toets: **Spelling 2012**



Figuur 1.7: Leerlingrapport uit het Computerprogramma LOVS

Om vast te stellen of de vorderingen van een leerling bevredigend zijn, moeten de vorderingen worden afgezet tegen een bepaald criterium. Dat kan door de vorderingen van de leerling te vergelijken met zichzelf, met de vorderingen van leeftijdgenoten en met de te realiseren doelen (zie bladzijde 18). Op basis van deze informatie kan worden bepaald of de vorderingen van de leerling nader geanalyseerd moeten worden en of voor deze leerling extra actie moet worden ondernomen (stappen analyseren en handelen).

In figuur 1.7 kunnen we zien dat Roland zich conform verwachting ontwikkelt in vergelijking met voorgaande periodes en in vergelijking met leeftijdgenoten. Er is dus geen directe noodzaak tot nadere analyse en aanpassing van het onderwijsaanbod.

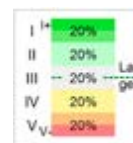
Groepsrapport

Een groepsrapport is een overzicht dat op grafische wijze de resultaten van een groep leerlingen op één afnamemoment weergeeft. We nemen als voorbeeld het groepsrapport E6 Spelling met de niveau-indeling I tot en met V (zie figuur 1.8 op de volgende bladzijde).

Op het groepsrapport is de niveau-indicatie I t/m V aan de bovenkant aangegeven. Met een

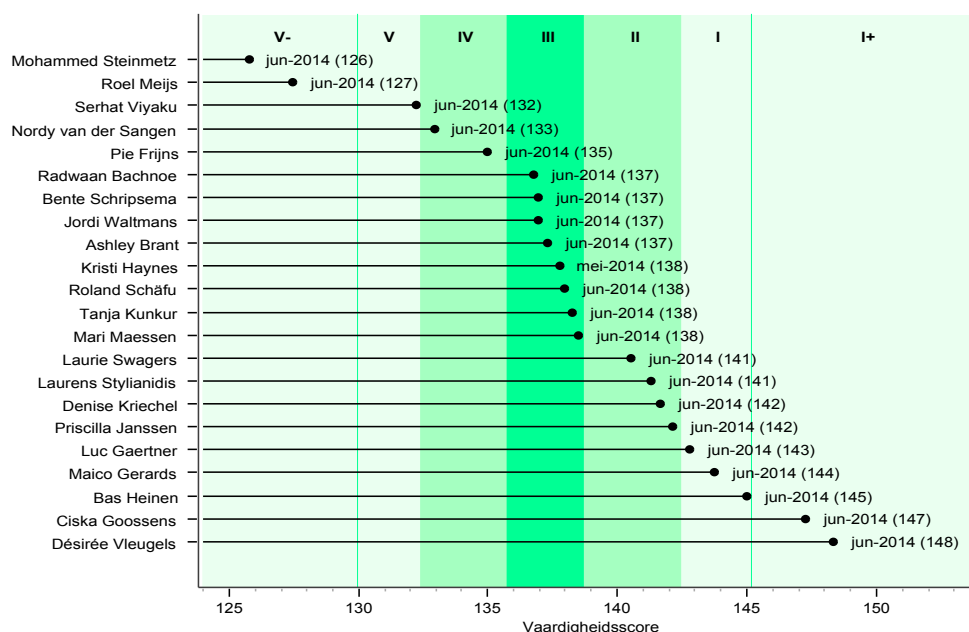
puntje is aangegeven welke vaardigheidsscore elke leerling heeft behaald. Achter het puntje is tussen haakjes de behaalde vaardigheidsscore vermeld met daarbij het moment waarop de toets is afgenomen. In het voorbeeld heeft Mohammed een vaardigheidsscore van 126. Deze vaardigheidsscore behoort bij Spelling E6 tot niveau V en binnen dat niveau tot de 10% laagst scorende leerlingen (vandaar de aanduiding V-).

Cito - DEMO versie Arnhem



Groepsrapport - grafisch

Groep: **6 - 6B (22 lln)**
Toets: **Spelling 2012 - E6**



Afname datum	Leerling	Toetsscore / Vaard.score	Score interval	Niveau norm. cor.
10-06-2014	Mohammed Steinmetz	28 / 126	124 – 127	V-
06-06-2014	Roel Meijs	33 / 127	126 – 129	V-
03-06-2014	Serhat Viyaku	38 / 132	131 – 134	V
17-06-2014	Nordy van der Sangen	36 / 133	131 – 134	IV
06-06-2014	Pie Frijns	43 / 135	133 – 137	IV
06-06-2014	Radwaan Bachnoe	45 / 137	135 – 138	III
11-06-2014	Bente Schripsema	47 / 137	135 – 139	III
06-06-2014	Jordi Waltmans	45 / 137	135 – 139	III
11-06-2014	Ashley Brant	45 / 137	136 – 139	III
27-05-2014	Kristi Haynes	46 / 138	136 – 140	III
18-06-2014	Roland Schäfu	48 / 138	136 – 140	III
14-06-2014	Tanja Kunkur	49 / 138	136 – 140	III
04-06-2014	Mari Maessen	50 / 138	137 – 140	III
03-06-2014	Laurie Swagers	52 / 141	138 – 143	II
06-06-2014	Laurens Stylianidis	52 / 141	139 – 143	II
05-06-2014	Denise Kriechel	53 / 142	139 – 144	II
06-06-2014	Priscilla Janssen	52 / 142	140 – 144	II
11-06-2014	Luc Gaertner	54 / 143	140 – 146	I
10-06-2014	Maico Gerards	54 / 144	141 – 147	I
26-06-2014	Bas Heinen	56 / 145	142 – 148	I
15-06-2014	Ciska Goossens	57 / 147	143 – 151	I+
06-06-2014	Désirée Vleugels	57 / 148	145 – 152	I+

Figuur 1.8: Groepsrapport uit het Computerprogramma LOVS

Op basis van de door alle leerlingen van de groep behaalde vaardigheidsscores, kan de gemiddelde vaardigheidsscore voor de groep worden berekend door het optellen van de vaardigheidsscores van alle leerlingen en de som te delen door het aantal leerlingen. De gemiddelde vaardigheidsscore van de groep krijgt betekenis door deze met de gemiddelde vaardigheidsscore van andere groepen te vergelijken. Daarom zijn bij de LVS-toetsen de gemiddelde vaardigheidsscores van een steekproef van scholen verzameld en als vergelijkingsgegevens in de handleiding opgenomen. Het is belangrijk om te realiseren dat leerlingen met een erg hoge of lage vaardigheidsscore grote invloed hebben op het gemiddelde, zeker bij kleine groepen. Wanneer deze leerlingen niet aan de afname meedoen, kan dit een opmerkelijke stijging of daling van het gemiddelde tot gevolg hebben.

Op basis van de informatie in het groepsrapport kunnen beslissingen worden genomen over de voortgang van het onderwijs en kan de leerstof beter worden afgestemd op het niveau van de groep of van individuele leerlingen. Hoe dat precies in zijn werk gaat, is uitgewerkt in hoofdstuk 3.

1.3 Toetsen voor het beoordelen van groepen leerlingen en scholen

Zoals in hoofdstuk 1 van het algemene deel van Toetsen op School aangegeven is, zijn toetsen die voor het beoordelen van groepen leerlingen en scholen gebruikt worden gewoonlijk dezelfde toetsen als die gebruikt worden voor het beoordelen van individuele leerlingen. De resultaten die individuele leerlingen van een bepaalde groep of school op een toets behaald hebben, worden dan gebruikt om informatie over de groep of de school te geven.

Als leerkrachten nagaan of hun onderwijs beantwoordt aan de doelen die ze nastreven, stellen ze zich vragen als:

- Wat brengt mijn onderwijs op oftewel hebben mijn leerlingen geleerd wat ik ze wilde leren?
- Hoe verloopt de feitelijke ontwikkeling van mijn leerlingen?
- Welk effect hebben bepaalde maatregelen en inspanningen die ik heb genomen?

Wanneer een school de doelen die ze nastreeft niet (geheel) realiseert, is het zoeken naar de oorzaken daarvan een logisch vervolg. Licht dat bijvoorbeeld aan:

- De organisatie van het onderwijs, bijvoorbeeld of het onderwijs meer of juist minder geïndividualiseerd moet worden.
- De wijze van lesgeven en de didactische aanpak.
- De inhoud van het onderwijsprogramma, oftewel wordt er voldoende tijd besteed aan het betreffende leergebied of onderdelen daarvan?
- De gehanteerde leermiddelen.

Met de discussie over deze aspecten neemt een team, in goed onderling overleg, de kwaliteit van het onderwijs onder de loep. Dat kan de aanzet vormen tot kwaliteitsverbeteringen.

Bij schoolzelfevaluatie wordt op elk niveau in het onderwijs gekeken of de beoogde kwaliteit en het gewenste niveau worden bereikt en hoe die kwaliteit zo nodig kan worden verbeterd. Alle toetsen die deel uitmaken van het Cito Volgsysteem primair en speciaal onderwijs voorzien naast rapportages op leerlingniveau ook in rapportages op groeps- en op schoolniveau. Een voorbeeld van een rapportage op groepsniveau bij de LVS-toetsen zagen we in paragraaf 1.2 in de vorm van het groepsrapport. De informatie die dit rapport bevat, is zowel van belang bij het beoordelen van de voortgang van het onderwijsleerproces als bij het beoordelen van groepen leerlingen. Ook bij de Entreetoets en bij sommige van de in paragraaf 1.1.2 genoemde eindtoetsen, zijn dit soort rapportages beschikbaar. Bij de LVS-toetsen bestaan er daarnaast ook rapportages over de verschillende jaargroepen heen, rapportages op schoolniveau dus. Hieronder zal eerst worden ingegaan op de schoolrapportage bij de Entreetoets. Vervolgens

zullen twee rapporten op schoolniveau bij de LVS-toetsen besproken worden, de dwars-doorsnede en de trendanalyse.

Schoolrapportage Entreetoets

Schoolrapporten zijn een belangrijk hulpmiddel bij schoolzelfevaluatie. Scholen kunnen met behulp van de rapporten zien hoe hun school presteert ten opzichte van voorgaande jaren en in vergelijking met andere scholen. In de schoolrapporten worden gemiddelde scores van de school vergeleken met gemiddelde scores van andere scholen.

Bij de Entreetoets kunnen scholen met behulp van het Rapportageprogramma Entreetoets (RET, beschikbaar via Cito Portal op <http://portal.cito.nl>) twee soorten schoolrapporten maken op basis van alle mogelijke selecties van leerlingen:

- Schoolrapport zonder correctie
- Schoolrapport met correctie Leerlinggewicht (LG)

In het Schoolrapport zonder correctie worden de gemiddelde scores van de school vergeleken met de gemiddelde scores van alle deelnemende scholen. Hierbij wordt geen rekening gehouden met kenmerken van de school of de leerlingen.

Het is maar de vraag of zo'n vergelijking zinvolle informatie oplevert. Dat zou zo zijn als we het op een school gegeven onderwijs als enige verklaring voor verschillen tussen schoolgemiddelden zouden kunnen beschouwen. Dat is echter niet zo. Verschillen tussen schoolgemiddelden worden voor een deel bepaald door factoren waarop het onderwijs niet of nauwelijks invloed heeft, maar die wel van grote betekenis zijn voor de onderwijsresultaten.

Factoren die systematische verschillen tussen scholen veroorzaken maar waar het onderwijs niet of nauwelijks invloed op heeft, zijn onder te verdelen in twee categorieën: factoren op schoolniveau en factoren op leerlingniveau. Een factor op schoolniveau is de sociaal-economische achtergrond van de leerlingen. Scholen die veel leerlingen met een zwakke(re) sociaal- economische achtergrond hebben, moeten een grotere inspanning leveren om landelijk gemiddelde prestaties neer te zetten dan scholen die weinig leerlingen met een zwakke(re) sociaal- economische achtergrond hebben. Een factor op leerlingniveau is de intelligentie van leerlingen. Leerlingen die intelligenter zijn, behalen bij dezelfde onderwijsinspanning gewoonlijk hogere scores dan leerlingen die meer moeite hebben met leren.

In het Schoolrapport correctie LG wordt daarom een schatting gegeven van de gemiddelde score op de onderdelen wanneer alle scholen die aan de Entreetoets meedoen gelijk zouden zijn geweest wat betreft het percentage gewichtenleerlingen. Het percentage gewichtenleerlingen op een school wordt daarbij beschouwd als een indicator voor de sociaal-economische achtergrond van de leerlingpopulatie.

Een andere factor waar het onderwijs geen invloed op heeft, is toeval. Een enkele leerling met een extreem hoge of lage score kan grote invloed hebben op de gemiddelde prestatie van een school, zeker bij een kleine groep leerlingen. Bij grotere groepen leerlingen vallen extreem hoge en lage scores makkelijker tegen elkaar weg. Het gemiddelde van grotere groepen leerlingen geeft daarom beter de prestaties van een school weer dan het gemiddelde van een kleine groep leerlingen. Bij het schoolrapport met instroomcorrectie voor leerlinggewicht wordt daarom rekening gehouden met de omvang van de groep.

In figuur 1.9 op de volgende bladzijde is een voorbeeld opgenomen van het Schoolrapport met instroomcorrectie voor leerlinggewicht (LG).

De schoolrapporten zijn op de volgende manier opgebouwd: In de linkerkolom staan van boven naar beneden:

- de schoolgegevens;
- de totaalscore op de Basismodule van de Entreetoets groep 7. De Basismodule bestaat uit 180 opgaven op het gebied van lezen, taalverzorging en rekenen-wiskunde;
- de totaalscores voor de onderdelen rekenen, lezen, taalverzorging, luisteren, schrijven, woordenschat en wereldoriëntatie;
- de scores op de verschillende subonderdelen waaruit elk onderdeel is opgebouwd, bijvoorbeeld grammatica.

Van links naar rechts zijn weergegeven:

- de kolom met de onderdelen en subonderdelen waaruit de Entreetoets groep 7 bestaat;
- de kolom met per onderdeel en subonderdeel het aantal opgaven ('aantal opgaven');
- de kolom met het 'gemiddeld aantal goed' of 'gecorrigeerd aantal goed';
- een kolom met een grafische weergave van de scores: de (gecorrigeerde) gemiddelde score (sterretjes *) en eventueel het betrouwbaarheidsinterval rondom die gemiddelde score (het kader).

Jaar:	2016	Schoolscores			Landelijk gemiddelde		
School:	99957	Aantal leerlingen 14 van 14	Gem. aantal goed zonder correctie 85,8	Gem. aantal goed met correctie LG 110,8	Zonder correctie	Correctie LG	
	De Regenboog					111,2	
	Hoofdweg 9999XX						
	Demostad	2016			110,9		
Printdatum: 17-06-2016							
Instroomcorrectie voor leerlinggewicht (LG)							
	Aantal opgaven	Gecorrigeerd aantal goed	V	IV	III	II	I
Totaal basis	180	110,8			*		
Rekenen	140	82,1			*		
• Getallen	50	33,8			*		
• Meten en Meetkunde	35	26,9			*		
• Verbanden	25	17,6			*		
• Verhoudingen	30	16,5			*		
Lezen	80	51,4			*		
• Begrijpend lezen	40	24,1			*		
• Opzoeken	20	13,6			*		
• Samenvatten	20	12,3			*		
Taalverzorging	90	57,5			*		
• Grammatica	20	13,4			*		
• Interpunctie	20	11,6			*		
• Spelling niet-werkwoorden	25	17,1			*		
• Spelling werkwoorden	25	16,9			*		
Luisteren	30	18,6			*		
Schrijven	30	22,1			*		
Woordenschat	40	25,7			*		
Wereldoriëntatie	60	39,8			*		
• Aardrijkskunde	20	12,2			*		
• Geschiedenis	20	12,3			*		
• Natuur en techniek	20	11,9			*		

Figuur 1.9: Schoolrapport correctie LG

Gecorrigeerd aantal goed en gecorrigeerd gemiddelde

Bij het schoolrapport met instroomcorrectie voor leerlinggewicht wordt een 'gecorrigeerd aantal goed' gerapporteerd. Dit is het verwachte aantal goed gemaakte opgaven na correctie voor leerlinggewicht. Een gecorrigeerd gemiddelde is het verwachte gemiddelde dat een school behaald zou hebben als het een school met doorsnee-instroom zou zijn.

Het verschil tussen het gecorrigeerde gemiddelde van een school en het ongecorrigeerde gemiddelde is groter naarmate de school qua instroom meer afwijkt van de aan de Entreetoets deelnemende scholen. Indien de school een relatief hoog percentage gewichtenleerlingen heeft, zullen de gemiddelden in de Schoolrapporten met correctie naar boven zijn bijgesteld.

Omgekeerd, indien een school een relatief laag percentage gewichtenleerlingen heeft, zullen de gemiddelden in de Schoolrapporten met correctie naar beneden zijn bijgesteld.

In het schoolrapport met correctie voor leerlinggewicht worden dus schattingen gerapporteerd. Daarom wordt ook wel van een verwacht gemiddelde gesproken als het over het gecorrigeerd gemiddelde gaat.

Voorbeeld

In Schoolrapport met correctie voor leerlinggewicht in het voorbeeld is het gecorrigeerd aantal goed na correctie voor LG 110,8. Het ongecorrigeerde aantal goed is 85,8.

Na correctie voor leerlinggewicht is er dus sprake van een hogere gemiddelde score.

Dat komt doordat deze school in vergelijking met andere aan de Entreetoets deelnemende scholen een hoger percentage gewichtenleerlingen heeft. Daarom is het gecorrigeerd aantal goed naar boven bijgesteld.

Onder- en bovengrens

Elke schatting heeft een bepaalde mate van onzekerheid. De mate van onzekerheid hangt af van allerlei factoren, maar een van de belangrijkste factoren is het aantal leerlingen waarop een gemiddelde berekend is. Hoe groter het aantal leerlingen, hoe nauwkeuriger het gecorrigeerde gemiddelde. Om de mate van onzekerheid aan te geven, rapporteren we de ondergrens en de bovengrens van het 90% betrouwbaarheidsinterval. Het voor de school gecorrigeerd gemiddelde ligt met een kans van 90% tussen deze onder- en bovengrens (zie ook hoofdstuk 3 van het algemene deel van Toetsen op School). Het kader in de grafische weergave geeft het 90%-betrouwbaarheidsinterval voor de score op de onderdelen grafisch weer. Als de waarde III in het kader valt, valt het landelijk gecorrigeerde gemiddelde tussen de onder- en bovengrens van het betrouwbaarheidsinterval. Als dit het geval is, heeft de school niet uitzonderlijk gepresteerd op het betreffende onderdeel. Als het kader geheel onder de waarde III ligt, heeft de school na correctie onder het landelijk gemiddelde gepresteerd. Als de donkere balk geheel boven de waarde III ligt, heeft de school na correctie op dit onderdeel boven het landelijk gemiddelde gepresteerd.

Voorbeeld

De school uit ons voorbeeld heeft niet extreem goed of slecht gepresteerd op de Basismodule in vergelijking met het landelijk gemiddelde. Het landelijk gemiddelde voor de Basismodule valt binnen het kader van het 90%-betrouwbaarheidsinterval.

Dwarsdoorsnede en trendanalyse bij de LVS-toetsen

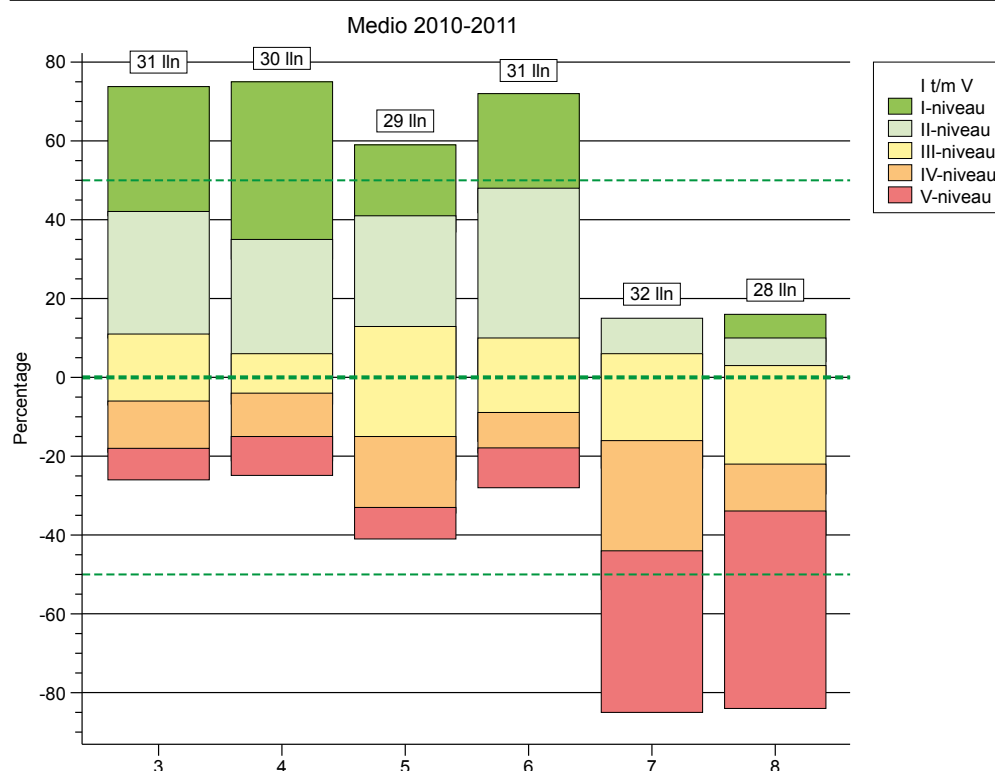
Bij de LVS-toetsen zijn verschillende rapportages op groeps- en schoolniveau mogelijk. Deze rapportages kunnen net als de rapportages op leerlingniveau handmatig worden samengesteld, maar ook hier bieden verschillende computerprogramma's mogelijkheden om deze rapportages te genereren. Hieronder bespreken we de interpretatiemogelijkheden van de dwarsdoorsnede en trendanalyse die gemaakt kunnen worden bij de LVS-toetsen van Cito met het Computerprogramma LOVS. Beide overzichten zijn bedoeld voor een teamdiscussie over kwaliteitszorg op school en over maatregelen om de kwaliteit van het onderwijs te verbeteren. De gegevens mogen echter niet te absoluut worden geïnterpreteerd. Het is op de eerste plaats namelijk erg lastig om 'de onderwijskwaliteit' in cijfers te vangen. Het is verder bekend dat de sociaal-economische achtergrond van de schoolbevolking voor een belangrijk deel het niveau van de leerresultaten bepaalt. Voor de ene school zal het behalen van een bepaald trendniveau een onmogelijke opgave betekenen, terwijl een andere school zich daar nauwelijks voor hoeft in te spannen.

Daarom moet bij de interpretatie van de dwarsdoorsnede en de trendanalyse rekening worden gehouden met de kenmerken van de leerlingen.

Dwarsdoorsnede

De dwarsdoorsnede toont in één overzicht voor alle verschillende jaargroepen van een school de verdeling van de leerlingen over de vijf verschillende niveaus op één of meerdere leerstofgebieden op een bepaald moment in de tijd (halverwege of aan het einde van het jaar). Uit de dwarsdoorsnede blijkt hoe de jaargroepen presteren in vergelijking met andere groepen in het land, of er in de school groepen zijn die op een laag niveau presteren maar ook of er groepen zijn met een prestatiepatroon dat afwijkt van wat op de school gebruikelijk is. Een dwarsdoorsnede maakt in feite een foto van een school op een bepaald moment in de tijd.

Dwarsdoorsnede - Rekenen-Wiskunde



Figuur 1.10: Dwarsdoorsnede uit het Computerprogramma LOVS

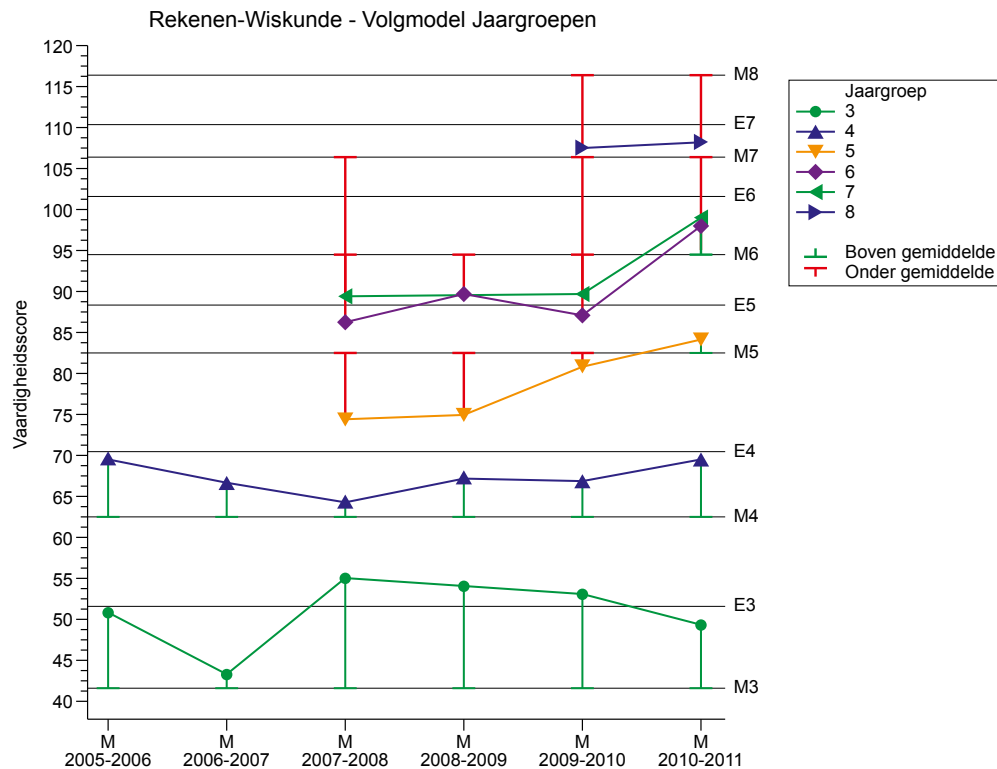
Figuur 1.10 is een dwarsdoorsnede voor Rekenen-Wiskunde medio schooljaar 2010/2011. De 0%-lijn stelt het landelijk gemiddelde voor. Boven deze lijn zijn de niveaus I en II afgebeeld, onder deze lijn de niveaus IV en V. Niveau III ligt om de 0%-lijn heen. Op basis van de landelijke verdeling zou 50% van de leerlingen een score behalen boven het gemiddelde (niveau I en II en een deel van niveau III) en 50% eronder (een deel van niveau III en niveau IV en V). In bovenstaand voorbeeld is te zien dat de groepen 3 t/m 6 in vergelijking met de landelijke verdeling hoog scoren: het gedeelte van de staaf boven de 0%-lijn is aanzienlijk groter dan onder de 0%-lijn. In groep 7 en 8 is het gedeelte van de staaf onder de 0%-lijn echter aanzienlijk groter dan boven de 0%-lijn: het percentage leerlingen dat een IV of een V-score haalt, is hoger dan we landelijk gezien zouden verwachten. Dat is een signaal dat om een nadere analyse vraagt. Wat is de reden dat de leerlingen van groep 7 en 8 lager scoren? Is dat een incidenteel verschijnsel omdat

bijvoorbeeld het onderwijs aan groep 7 en 8 door een aantal min of meer toevallige oorzaken in schooljaar 2010/2011 anders is verlopen dan de bedoeling was? Zijn er wijzigingen in de groepssamenstelling opgetreden? Op scholen waar gedurende het schooljaar veel nieuwe leerlingen bijkomen of vertrekken kunnen opeenvolgende dwarsdoorsneden een minder betrouwbaar beeld van de ontwikkeling van de leerprestaties van een jaargroep geven. Dat geldt ook voor een dwarsdoorsnede van een kleine groep leerlingen (< 15 leerlingen).

Als er veel wijzigingen in de groepssamenstelling zijn opgetreden of als het een (zeer) kleine groep leerlingen betreft, moeten maatregelen die betrekking hebben op het onderwijsprogramma met grote terughoudendheid worden genomen. Uitval veroorzaakt door incidentele oorzaken hoeft voor het onderwijs in dat leerjaar niet direct consequenties te hebben. Wel zal het zo zijn dat de volgende leerkracht met deze beginsituatie rekening zal moeten houden. Bestudering van de dwarsdoorsneden van voorgaande jaren kan ook aan het licht brengen dat bijvoorbeeld de rekenprestaties in groep 7 en 8 stelselmatig tegenvallen. Dat kan wijzen op een mogelijke tekortkoming in het programma in leerjaar 7 en 8. Het kan ook een aanwijzing zijn dat er te weinig tijd besteed wordt aan rekenen. Door de dwarsdoorsneden van voorgaande jaren te bekijken, is het mogelijk om vast te stellen dat een bepaalde groep stelselmatig lager presteert dan de andere. Zo'n 'zwakke groep' schuift als het ware op elke dwarsdoorsnede een jaar op. Wanneer dit wordt geconstateerd, hoeft dit niet direct te leiden tot structurele aanpassing van het onderwijsprogramma. Wel zal het duidelijk zijn dat het programma voor deze jaargroep moet worden aangepast. Immers de resultaten van het ene jaar vormen tevens het beginniveau voor het daaropvolgende jaar.

Trendanalyse

Dwarsdoorsneden zullen vaak illustreren dat de jaargroepen op een school wisselende resultaten halen. De ene keer wat beter dan de andere keer. Op zichzelf is dat niet vreemd. De ene jaargroep leerlingen is nu eenmaal beter dan de andere. Dat maakt het lastig om te beoordelen of een gevonden resultaat terug te voeren is tot 'de kwaliteit van het gegeven onderwijs' of tot de 'toevallige' samenstelling van de jaargroep. Om daar enig zicht op te krijgen, moeten gegevens van meer groepen en van achtereenvolgende jaren met elkaar in verband worden gebracht. Iedere dwarsdoorsnede is als het ware te beschouwen als een momentopname. Telkens komen er echter nieuwe resultaten bij die, gevoegd bij de resultaten van eerdere jaren, eventuele ontwikkelingen in kaart brengen en een beeld geven van de effectiviteit van het onderwijsprogramma. Dat is de kern van een trendanalyse. De trendanalyse laat de resultaten zien die de leerlingen van een school door de jaren heen behaalden op een bepaalde toets in een bepaalde jaargroep of in bepaalde jaargroepen. Een trendanalyse geeft informatie over de gemiddelde opbrengst van het onderwijs op basis waarvan beslissingen over het onderwijs kunnen worden genomen en gericht gewerkt kan worden aan verbeteringen.



Figuur 1.11: Trendanalyse volgmodel jaargroepen Computerprogramma LOVS

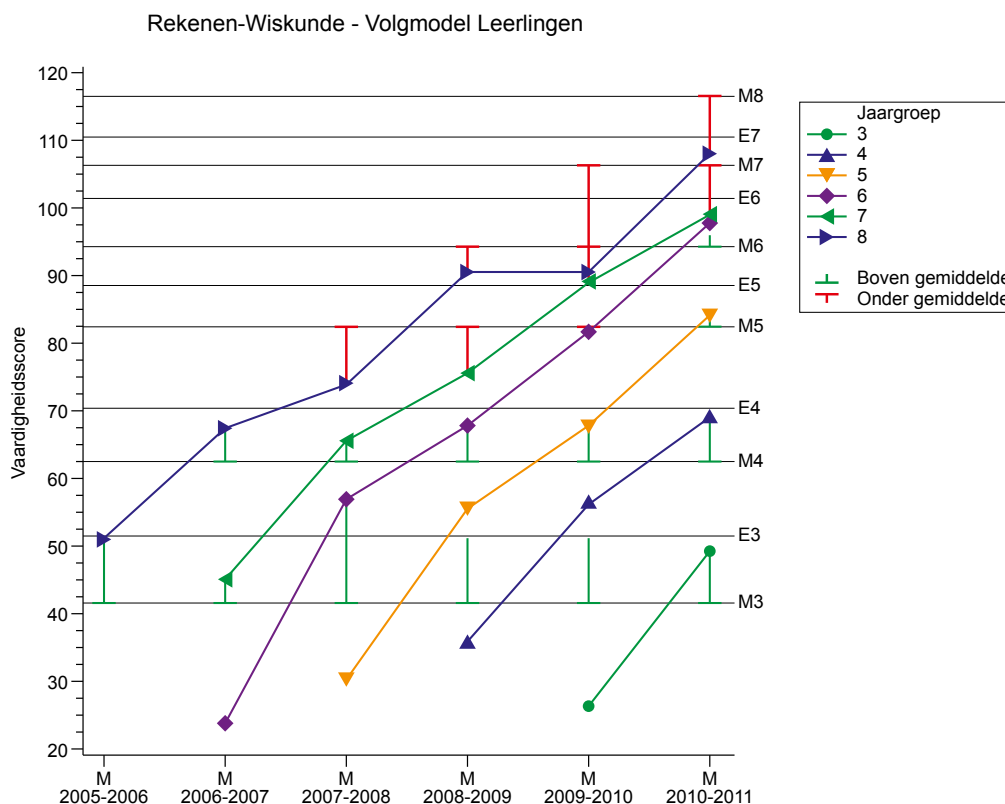
Figuur 1.11 is een voorbeeld van een trendanalyse volgmodel jaargroepen. Dit rapport geeft een beeld van het niveau van bepaalde jaargroepen door de jaren heen op één bepaald leergebied.

In het voorbeeld gaat het om een trendanalyse Rekenen-Wiskunde voor de jaargroepen 3 t/m 8. Als we kijken naar de trend van groep 3 dan wordt de gemiddelde score van groep 3 in schooljaar 2010/2011 vergeleken met de gemiddelde score van groep 3 van schooljaar 2009/2010 en alle voorgaande jaren (t/m schooljaar 2005/2006). Het gaat dus om de gemiddelde scores van verschillende groepen leerlingen. In dit voorbeeld is in groep 3 in zes opeenvolgende jaren halverwege het leerjaar een toets Rekenen-Wiskunde afgenomen. Op de linkeras van de figuur staan de vaardigheidsscores (ook wel schaalscores genoemd) afgebeeld. De horizontale lijnen met daarnaast M3, E3 en zo verder geven de gemiddelde vaardigheidsscores van de landelijke referentiegroep weer.

In het voorbeeld kunnen we zien dat, hoewel de resultaten over de jaren heen variëren, de gemiddelde vaardigheidsscores in de groepen 3 en 4 steeds boven het landelijke gemiddelde liggen (respectievelijk boven de M3-lijn en boven de M4-lijn). In groep 5, 6, 7 en 8 liggen de resultaten echter vrijwel alle jaren (ver) beneden het landelijk gemiddelde. Voor al deze groepen geldt dat er in schooljaar 2010/2011 sprake is van een (lichte) stijging in de scores; in groep 5 en 6 liggen de resultaten in dat schooljaar boven het landelijk gemiddelde. In groep 7 en 8 liggen de resultaten nog steeds beneden het landelijk gemiddelde maar er is wel sprake van een lichte stijging. De vraag is dus wat er veranderd is in het rekenonderwijs in dit schooljaar en nog belangrijker, hoe de school deze veranderingen op een dusdanige wijze kan borgen dat de resultaten blijven(d) verbeteren.

Naast een trendanalyse van jaargroepen is het ook interessant om te zien hoe eenzelfde groep leerlingen (een cohort) zich door de jaren heen ontwikkeld heeft. De leerlingen in de huidige groep 8 worden dan in beeld gebracht met hun prestaties in de voorgaande jaren, oftewel toen zij in groep 7 en 6 en eerder zaten. Dit is zichtbaar in het rapport trendanalyse volgmodel leerlingen waarvan een voorbeeld is opgenomen in figuur 1.12.

Trendanalyse



Figuur 1.12: Trendanalyse volgmodel leerlingen Computerprogramma LOVS

Net als in de trendanalyse jaargroepen zijn in deze figuur op de horizontale as de schooljaren opgenomen terwijl op de verticale as de vaardigheidsscores afgebeeld zijn. En ook hier geven de horizontale lijnen met daarnaast M3, E3 en zo verder de gemiddelde vaardigheidsscores van de landelijke referentiegroep weer.

Als we in het voorbeeld de resultaten van groep 8 in schooljaar 2010/2011 bekijken, zien we dat groep 8 vrijwel in alle schooljaren (ver) beneden het landelijk gemiddelde scoort. De gemiddelde score van de leerlingen van deze groep lag boven het landelijk gemiddelde toen zij in groep 3 (schooljaar 2005/2006) startten. Dat geldt ook voor de gemiddelde score van dit cohort in groep 4 (schooljaar 2006/2007). Ditzelfde geldt voor de leerlingen in groep 7 in schooljaar 2010/2011. Zij startten in 'hun' groep 3 (schooljaar 2006/2007) en groep 4 (schooljaar 2007/2008) met een bovengemiddelde score, maar scores sinds schooljaar 2008/2009 beneden het gemiddelde, respectievelijk beneden de M5-lijn, beneden de M6-lijn en beneden de M7-lijn.

Als we de gegevens uit de trendanalyse volgmodel leerlingen naast de dwarsdoorsnede Rekenen-Wiskunde leggen, kunnen we met zekerheid zeggen dat de veronderstelling dat de groepen 7 en 8 'zwakke' groepen zouden zijn, verworpen kan worden. Immers de leerlingen in de huidige groepen 7 en 8 stroomden in met een bovengemiddelde score en zakten vanaf

groep 5 beneden het gemiddelde. Kijken we naar de resultaten van de leerlingen in groep 5 en 6 in schooljaar 2010/2011, dan zien we dat zij in vergelijking met het landelijk gemiddelde alle jaren boven of op het landelijk gemiddelde scoren. Maar we zien ook dat zij hoger scoorden in 'hun' groep 3 en 4 dan ze in schooljaar 2010/2011 scoorden.

Het lijkt er dus op dat de resultaten dalen als de leerlingen doorstromen naar de hogere jaargroepen (vanaf groep 5). Om te zien of dat waar is, kunnen we de trendanalyse volgmodel jaargroepen bekijken. Deze bevestigt het beeld dat verschillende cohorten in dezelfde school vanaf groep 5 beneden het gemiddelde scoren (met de eerder genoemde uitzonderingen in groep 5 en 6 in schooljaar 2010/2011). Zoals ook bij de trendanalyse volgmodel jaargroepen is opgemerkt, is er in schooljaar 2010/2011 in groep 5 t/m 8 wel een (lichte) stijging zichtbaar. Dat bevestigt de eerder gemaakte opmerking dat goed zou moeten worden nagegaan wat er in schooljaar 2010/2011 veranderd is in het rekenonderwijs en hoe de school deze veranderingen verder kan doorzetten en borgen om het onderwijs te verbeteren.

Het rapport trendanalyse volgmodel jaargroepen leent zich er overigens ook uitstekend voor om een zogenaamde schooleigen standaard aan te ontleen op grond waarvan een meer evenwichtige beoordeling van de kwaliteit van het onderwijs tot stand kan komen.

Als bijvoorbeeld jaar na jaar en over een langere periode (minimaal 3 of 4 jaar) blijkt dat de jaargroepen 3 ruim boven het landelijk gemiddelde scoren, kan dat als minimale doelstelling genomen worden voor het onderwijs in deze jaargroep. Ook kunnen eventueel acties worden ondernomen om de onderwijskwaliteit verder te verbeteren. Zo verleent de trendanalyse elke school een reële basis voor het werken aan kwaliteitsverbetering. Het gaat daarbij niet zozeer om de vraag of er een, al dan niet van buitenaf opgelegd, criterium wordt gehaald maar om het nemen van maatregelen die er toe leiden dat een merkbare verhoging van de trendscore optreedt.

1.4 Toetsen voor het beoordelen van de kwaliteit van het onderwijs

Zoals ook is aangegeven in hoofdstuk 1 van het algemene deel van Toetsen op School vindt er zowel nationaal als internationaal onderzoek plaats naar de kwaliteit van het primair onderwijs in Nederland. De resultaten van deze onderzoeken zijn niet bedoeld voor het beoordelen van leerlingen en scholen maar om betrokkenen bij het onderwijs te informeren over de stand van zaken van het onderwijs. De resultaten van deze onderzoeken zijn van belang voor iedereen die betrokken is bij de discussie over en de vormgeving van het primair onderwijs. Allereerst natuurlijk de overheid als opdrachtgever van het onderzoek en daarnaast onderwijsorganisaties, onderzoekers en ontwikkelaars van methoden, onderwijsbegeleiders en lerarenopleiders, leraren basisonderwijs en ouders.

Tot eind 2014 werd in opdracht van het ministerie van OCW door Cito voor de meeste inhoudsgebieden in het primair onderwijs zogenaamd periodiek peilingonderzoek (Periodieke Peiling van het OnderwijsNiveau, PPON) verricht. Het belangrijkste doel van peilingonderzoek is om met een zekere regelmaat (voor de meeste inhoudsgebieden elke 5 jaar) gegevens te verzamelen over het onderwijsaanbod en de onderwijsresultaten in het basisonderwijs en het speciaal (basis)onderwijs. Deze onderzoeksresultaten bieden een basis voor een algemeen maatschappelijk debat over de inhoud en het niveau van het onderwijs.

Het onderzoek richt zich op drie vragen:

- Waaruit bestaat het onderwijsaanbod in een bepaald leer- en vormingsgebied?
- Welke resultaten in termen van kennis, inzicht en vaardigheden zijn er gerealiseerd?
- Welke veranderingen of ontwikkelingen in aanbod en opbrengst zijn er in de loop van de tijd te traceren?

Bij het uitvoeren van peilingonderzoek wordt een reeks toetsen voor verschillende onderwerpen van een inhoudsgebied, bijvoorbeeld Rekenen-Wiskunde, ontwikkeld en afgenomen bij een groot aantal leerlingen van basisscholen. Elke leerling maakt maar een klein deel van alle opgaven die voor een inhoudsgebied worden getoetst omdat peilingsonderzoek niet tot doel heeft inzicht te geven in de resultaten van individuele leerlingen.

De resultaten van elk peilingsonderzoek zijn door Cito vastgelegd in een onderzoeksrapport, ook wel 'Balans' genoemd. Daarnaast zijn de resultaten van ieder onderzoek samengevat in een folder die naar alle scholen wordt gestuurd: PPON informeert. Alle verschenen balansen en de schoolfolders zijn te raadplegen op www.cito.nl > Onderzoek en wetenschap > Participatie nationaal onderzoek > Periodieke Peiling van het Onderwijsniveau.

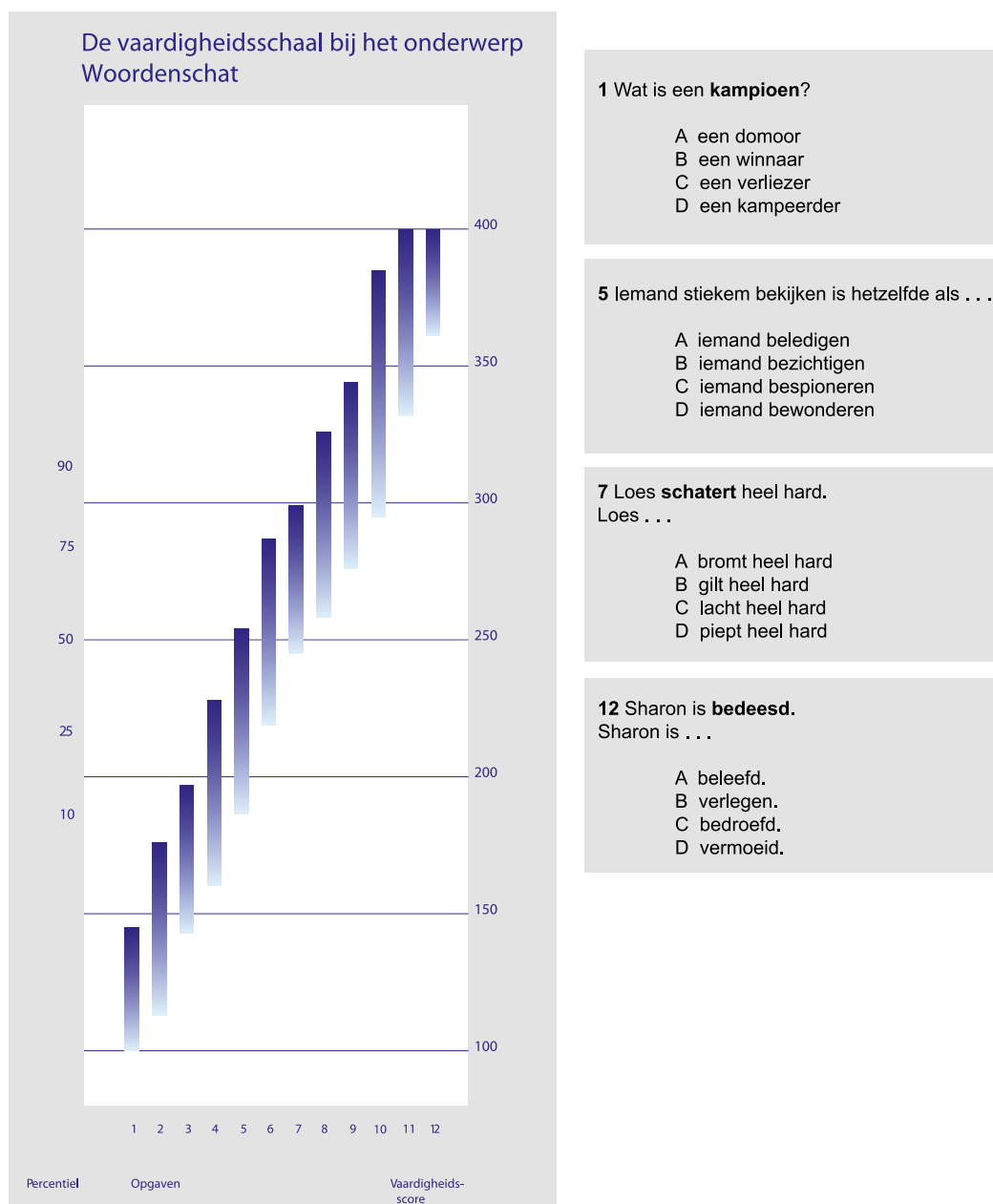
Om antwoord te geven op de vraag welke resultaten er in termen van kennis, inzicht en vaardigheden zijn behaald, zijn in de balansen de vaardigheden van de leerlingen aan de hand van voorbeeldopgaven beschreven. De voorbeeldopgaven illustreren wat leerlingen op verschillende niveaus van vaardigheid goed, matig of in onvoldoende mate weten of kunnen. Op de volgende pagina is in figuur 1.13 een voorbeeld opgenomen uit de Balans van het leesonderwijs halverwege de basisschool (2008). Halverwege de basisschool wil in dit geval zeggen eind groep 5.

In figuur 1.13 zien we aan de rechterkant de vaardigheidsschaal (zie paragraaf 1.2) staan met vaardigheidsscores van (beneden) 100 tot (boven) 400. Aan de linkerkant staan percentielen. Percentielen geven weer hoeveel procent van de leerlingen eenzelfde of lagere score heeft gehaald. Percentiel 25 betekent bijvoorbeeld dat 25% van de leerlingen een lagere score heeft en 75% van de leerlingen dus hoger scoort.

Op de horizontale as staan nummers vermeld. Deze nummers horen bij de balkjes in de grafiek. Deze balkjes representeren opgaven (een aantal opgaven is naast de figuur afgebeeld) en geven aan hoe moeilijk een opgave is. De bovenkant van het balkje komt overeen met de vaardigheidsscore waarbij een leerling een kans van 80% heeft om de opgave goed te beantwoorden. De onderkant van het balkje geeft de vaardigheidsscore aan waarbij een leerling een kans van 50% heeft om de opgave goed te maken. Als een opgave moeilijker is, ligt het balkje hoger op de vaardigheidsschaal. De opgaven zijn dus geordend naar moeilijkheidsgraad met links de gemakkelijkste opgave en rechts de moeilijkste.

Aan de hand van de balkjes onderscheiden we drie niveaus in beheersing van een opgave:

- We spreken van een goede beheersing wanneer de kans op een goed antwoord groter is dan 80%. De leerling heeft dan een vaardigheidsscore die hoger ligt dan het balkje. Een leerling met een vaardigheidsscore van 250 beheerst opgave 1 dus goed.
- Wanneer de kans op een goed antwoord tussen de 50% en 80% ligt (dat is dus het gebied van het balkje), spreken we van een (zeer) matige tot bijna goede beheersing. Een leerling met een vaardigheidsscore van 250 beheerst opgave 7 (zeer) matig (de vaardigheidsscore ligt aan de onderkant van het balkje) en opgave 5 bijna goed (de vaardigheidsscore ligt aan de bovenkant van het balkje).
- We spreken van onvoldoende beheersing van een opgave wanneer de kans op een goed antwoord kleiner is dan 50%. De leerling heeft dan een vaardigheidsscore die lager ligt dan het afgebeelde balkje. Een leerling met een vaardigheidsscore van 250 beheerst opgave 12 dus onvoldoende.



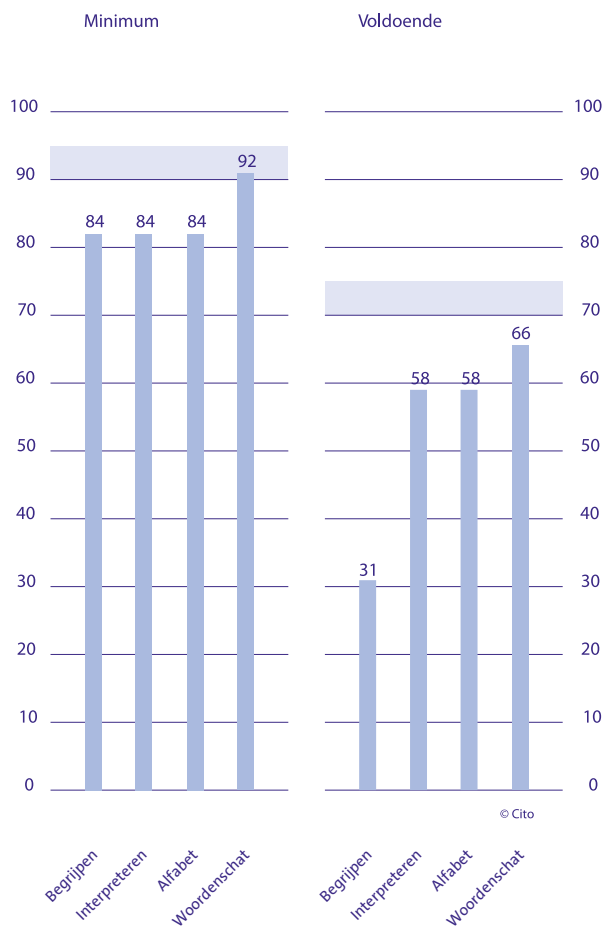
Figuur 1.13: Vaardigheidsschaal Woordenschat

Om te bepalen wat leerlingen halverwege de basisschool zouden moeten weten of kunnen en in hoeverre leerlingen dit resultaat ook bereiken, geven deskundige beoordelaars (zoals leerkrachten, ontwikkelaars en opleiders van pabo's) aan wat naar hun oordeel het gewenste niveau van vaardigheid is. Ook hierover wordt in de balansen gerapporteerd.

De deskundige beoordelaars bepalen niveaus voor minimale en voldoende beheersing, de zogenoemde standaarden. De standaard Voldoende geeft het niveau aan waarop – volgens de beoordelaars – de doelen van het basisonderwijs in voldoende mate beheerst worden. In het geval van de Balans van het leesonderwijs halverwege de basisschool gaat het hierbij om de tussendoelen aangezien het gaat om een peiling halverwege de basisschool. De beoordelaars gaan er bij het bepalen van de standaard Voldoende van uit dat 70 tot 75% van de leerlingen dit niveau zou moeten kunnen bereiken. Voor zover leerlingen de standaard Voldoende niet bereiken, dient het onderwijs te streven naar een minimaal vaardigheidsniveau. Dit niveau

wordt geformuleerd met de standaard Minimum. De beoordelaars gaan er bij het bepalen van de standaard Minimum vanuit dat 90 tot 95% van de leerlingen dit niveau zou moeten kunnen bereiken.

In figuur 1.14, afkomstig uit de publicatie 'Balans van het leesonderwijs halverwege de basisschool' (2008), zien we dat voor Woordenschat 92% van de leerlingen de standaard Minimum bereikt en 66% van de leerlingen de standaard Voldoende. Aangezien de beoordelaars er bij het vaststellen van de standaard Voldoende van uitgaan dat 70 tot 75% van de leerlingen dit niveau zou moeten kunnen bereiken (het grijze balkje in de rechter grafiek), kunnen we op basis van de gegevens in figuur 1.14 concluderen dat voor Woordenschat de standaard Voldoende door minder leerlingen wordt gehaald dan waar de deskundigen naar streven. Het aantal leerlingen dat de standaard Minimum behaalt (92%), ligt ongeveer op het niveau dat de deskundigen beogen (90 tot 95%, het grijze balkje in de linker grafiek). Voor de standaard Minimum kan dus gezegd worden dat de leerlingen presteren conform verwachting terwijl de leerlingen enigszins onder verwachting presteren bij de standaard Voldoende.

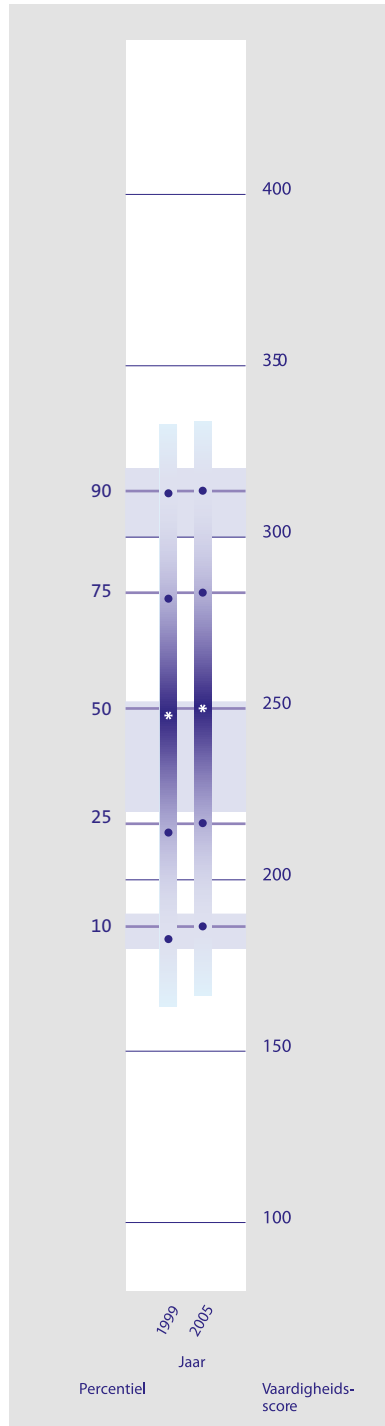


Figuur 1.14: Percentage leerlingen dat halverwege de basisschool de standaarden bereikt

Tot slot wordt in de balansen gerapporteerd hoe de resultaten van leerlingen zich in de loop van de tijd ontwikkelen. Dat is zichtbaar in figuur 1.15, afkomstig uit de publicatie 'Balans van het leesonderwijs halverwege de basisschool' (2008).

In figuur 1.15 is net als in figuur 1.13 aan de rechterkant wederom de vaardigheidsschaal afgebeeld en aan de linkerkant wederom de percentielen. Verder zien we twee balkjes met

daaronder de jaartallen 1999 en 2005. Deze balkjes laten de vaardigheidsverdelingen van de leerlingen in respectievelijk 1999 en 2005 zien. Met de sterretjes wordt de gemiddelde vaardigheidsscore van de leerlingen weergegeven, de punten verbeelden van onder naar boven de percentielscores 10, 25, 75 en 90. In de figuur zien we dat in 2005 alle scores iets hoger liggen dan in 1999. De leerlingen lijken in 2005 dus iets te zijn vooruit gegaan ten opzichte van 1999.



Figuur 1.15: Vaardigheidsverdelingen woordenschat in 1999 en in 2005

Vanaf 2014 is de regie over het peilingsonderzoek in handen gekomen van de Inspectie van het Onderwijs. Onder de naam Peil.onderwijs laat de Inspectie ieder schooljaar twee peilingen uitvoeren. Doel van de peilingen is (net als bij PPON) om in kaart te brengen wat kinderen aan het einde van het primair onderwijs kennen en kunnen op verschillende inhoudsgebieden. Voor taal en rekenen vormen de referentieniveaus taal en rekenen (zie www.taalenrekenen.nl) de basis voor de peilingen; voor de overige inhoudsgebieden zijn dat de kerndoelen (zie www.slo.nl/primair/kerndoelen). Omdat de kerndoelen vrij globaal zijn, worden ze door het Nationaal Expertisecentrum Leerplanontwikkeling (SLO) nader uitgewerkt in domeinbeschrijvingen. Die vormen het uitgangspunt voor de peilingen. De peilingen geven vervolgens zicht op de kennis en vaardigheden van leerlingen op een bepaald inhoudsgebied, mede in relatie tot het onderwijsaanbod en over de tijd heen.

Hoe de rapportage over de peilingsonderzoeken onder regie van de Inspectie er exact uit komt te zien (en of bovenstaande rapportageoverzichten worden gehandhaafd), is op dit moment nog niet bekend. Over de eerste peilingsonderzoeken moet nog worden geapporteerd. Naast deze peilingen rapporteert de Inspectie ook jaarlijks over de taal- en rekenvaardigheid van leerlingen einde basisonderwijs op basis van de gegevens van de eindtoets. De eerste rapportage van deze zogenaamde jaarlijkse meting taal en rekenen 2015 is te vinden op www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs. Naast het nationaal peilingsonderzoek PPON vindt er in het PO ook internationaal onderzoek naar de kwaliteit plaats: PIRLS en TIMSS. PIRLS staat voor Progress in International Reading Literacy Study en is een onderzoek naar de leesvaardigheid bij leerlingen na vier jaar leesonderwijs, leerlingen van 9 en 10 jaar (zie <http://www.expertisecentrumnederlands.nl/projecten/pirls/> en www.pirls.org). Naast het feit dat PIRLS uitspraken doet over de bereikte resultaten in termen van kennis, inzicht en vaardigheden en over hoe deze resultaten zich in de loop van de tijd ontwikkelen, laat PIRLS ook zien hoe Nederlandse leerlingen lezen in vergelijking met andere aan het onderzoek deelnemende landen. Dat is een belangrijk verschil met het nationaal peilingsonderzoek waar alleen Nederlandse leerlingen aan deelnemen. Die vergelijking tussen Nederlandse leerlingen en leerlingen uit andere aan het onderzoek deelnemende landen vindt ook plaats bij TIMSS. TIMSS staat voor Trends in International Mathematics and Science Study en is een onderzoek naar de leerlingprestaties voor rekenen en natuuronderwijs bij leerlingen in groep 6 (zie voor de resultaten van het meest recente TIMSS-onderzoek www.utwente.nl/ibr/timssenpirls of www.pirls.org > TIMSS).

Vaak wordt in internationaal onderzoek aangegeven op welke plek Nederland staat in de ranglijst van deelnemende landen. Die plek is meestal over de jaren heen niet precies hetzelfde. Als Nederland in de ranglijst gedaald is in vergelijking met een eerdere afname betekent dat nog niet per definitie dat bijvoorbeeld de leesvaardigheid van de leerlingen in Nederland ook gedaald is. Het kan bijvoorbeeld ook zijn dat in andere landen de leerlingen beter zijn gaan lezen, terwijl de leesvaardigheid in Nederland gelijk is gebleven. Een vergelijkbare redenering geldt uiteraard ook als Nederland stijgt in de ranglijst. Ook kan het zijn dat er nieuwe landen hebben deelgenomen die hoger scoren dan Nederland op, in dit geval, de leesvaardigheid. Dit is iets om rekening mee te houden als er over internationaal onderzoek wordt gerapporteerd.

2 De inhoud en constructie van toetsen

2 De inhoud en constructie van toetsen

Judith Hollenberg en Marleen van der Lubbe

Een leerkracht toetst niet om te toetsen, maar wil specifieke informatie verzamelen. Bij elke toets is het daarom belangrijk om in beeld te hebben wat de toets beoogt te meten. In paragraaf 2.1 staan enkele aandachtspunten bij de vraag 'Wat meet de toets?' beschreven. Daarbij worden enkele algemene punten genoemd om na te kunnen gaan of een toets ook bij specifieke groepen leerlingen meet wat de toets beoogt te meten. In het primair onderwijs sluiten de meeste toetsen aan bij de kerndoelen en de referentieverhalen naar onderwijsaanbod voor elk leerjaar, worden deze uitgewerkt in leerlijnen en tussendoelen. Leerlijnen en tussendoelen vormen niet alleen de basis voor het onderwijsaanbod, maar ook voor een toets. Een voorbeeld van een leerlijn staat beschreven in paragraaf 2.3. Elke opgave in een toets is zorgvuldig gekozen en geconstrueerd. We willen immers met zo min mogelijk toetsvragen zoveel mogelijk informatie over de vaardigheid van een leerling verzamelen. Hoe een evenwichtige toets geconstrueerd wordt die lang genoeg is om een betrouwbare uitspraak te doen over de vaardigheid van een leerling, maar zo beknopt mogelijk is om de toetslast te beperken, wordt beschreven in paragraaf 2.4.

2.1 Visie op leren en toetsen

De meeste toetsen in het PO meten de vaardigheid op één leergebied. Zo meet de (methode-onafhankelijke) volgtoets 'Begrijpend lezen van Kapinga', de vaardigheid begrijpend lezen. Een methodegebonden toets uit 'Veilig leren lezen' meet de vaardigheid technisch lezen. Methodegebonden toetsen hebben als doel om na te gaan of een leerling de zojuist aangeboden lesstof beheerst. Hiervoor volstaan enkele opgaven die aansluiten bij die lesstof. Om een uitspraak te doen over een vaardigheid als geheel zijn meerdere opgaven nodig die representatief zijn voor een grotere leerstofeenheid. Een methodeonafhankelijke toets wordt volgens een uitgebreide, zorgvuldige procedure geconstrueerd (zie paragraaf 2.4). Na constructie wordt de toets eerst bij een grote groep leerlingen uitgeprobeerd, ook wel pretesten of proeftoetsen genoemd, voordat de definitieve, genormeerde versie wordt samengesteld. Een methodegebonden toets wordt doorgaans niet uitgeprobeerd en genormeerd. Dit is gezien het primaire doel van een methodegebonden toets, namelijk directe bijsturing van het onderwijsleerproces, ook minder noodzakelijk. Ook methodeonafhankelijke toetsen hebben dit doel, maar ze worden daarnaast ook gebruikt voor belangrijke beslissingen zoals de indicatiestelling voor speciaal onderwijs of het bepalen van een ontwikkelingsperspectief met uitstroombestemming (zie pagina 75). Methodeonafhankelijke toetsen dienen betrouwbaar en valide te zijn (zie ook de hoofdstukken 3 en 4 over betrouwbaarheid en validiteit in het algemene deel van ToS) zodat de leerkracht op basis van de resultaten verstrekken beslissingen op verantwoorde wijze kan nemen. In welke niveaugroep past de leerling bijvoorbeeld? En maakt de leerling voldoende groei door of is voor deze leerling extra aandacht voor het leergebied als geheel nodig? De verschillende wijzen waarop de methodegebonden en methodeonafhankelijke toetsen zijn samengesteld en ontwikkeld, sluit dus aan bij de verschillende doelen waarvoor de toetsen worden gebruikt. Naast methodegebonden en methodeonafhankelijke toetsen die een specifieke vaardigheid

meten, zijn er ook methodeonafhankelijke toetsen die zich richten op een combinatie van vaardigheden. In de Eindtoets Basisonderwijs (zie hoofdstuk 1) zijn bijvoorbeeld opgaven over meerdere vakgebieden opgenomen, omdat de leervorderingen op verschillende inhoudsgebieden een belangrijke voorspeller zijn voor toekomstig schoolsucces. Ook intelligentietests richten zich niet op een vakgebied, maar bekijken een breder geheel. Zo komen in intelligentietests zowel opgaven rondom ruimtelijk inzicht als meer talige opgaven aan bod.

Behalve toetsen die zich richten op een specifiek vakgebied en toetsen die meerdere inhoudsgebieden meten, worden in het PO ook instrumenten gebruikt om de sociaal-emotionele ontwikkeling in kaart te brengen. Hier betreft het vaak een combinatie van vaardigheden, zoals sociale vaardigheden, en gedragskenmerken, zoals teruggetrokken gedrag.

Rekening houden met speciale behoeften

Elke toets meet een eigen vaardigheid. Belangrijk is dat de toets daadwerkelijk meet wat we willen meten. Met een rekentoets willen we bijvoorbeeld rekenvaardigheid meten, niet de aandachtsspanne. Wanneer de toets voor een leerling te lang duurt, is concentratie een beperkende factor en brengen we dus niet meer alleen de rekenvaardigheid in kaart, maar ook hoe lang een leerling geconcentreerd kan werken. De toets meet dan dus niet meer wat we beogen te meten.

Met een begrijpend leestoets willen we de vaardigheid op het gebied van begrijpend lezen meten. Wanneer de leerkracht de toets voorleest, meet de toets geen leesvaardigheid meer, maar luistervaardigheid.

Bij elke toets is het dus van belang om vooraf goed te realiseren wat er met de toets gemeten moet worden. Heeft de leerling specifieke beperkingen, zoals een beperktere aandachtsspanne of visusproblemen, dan is het extra belangrijk om na te gaan of de toets ook bij deze leerling meet wat we willen meten.

Zie voor het toetsen van leerlingen met extra onderwijsbehoeften:

www.cito.nl > Primair en speciaal onderwijs > Toetsen speciale leerlingen. Hier is een pdf 'Wegwijzer toetsgebruik voor leerlingen met specifieke onderwijsbehoeften' te downloaden. Hierin staat meer informatie over welke aanpassingen mogelijk zijn en welke aangepaste versies van toetsen beschikbaar zijn.

Bij de ontwikkeling van nieuwe toetsen houdt Cito zoveel mogelijk rekening met leerlingen met extra onderwijsbehoeften. Informatie hierover is te vinden op www.cito.nl > Primair en speciaal onderwijs > LVS-toetsen > derde generatie.

Veelgebruikte toetsen

In de notitie 'Analyse en waardering van opbrengsten primair onderwijs' van de Inspectie van het Onderwijs worden verschillende toetsen genoemd. Zie www.onderwijsinspectie.nl > Publicaties. Deze zijn ingedeeld in:

- Toetsen die eindresultaten in beeld brengen, oftewel de eindtoetsen voor groep 8 (zie hoofdstuk 1).
- Toetsen die tussenresultaten in beeld brengen, oftewel volgtoetsen zoals de DLE-toetsen en LVS-toetsen van Cito.
- Meetinstrumenten voor sociale competentie.

Bij toetsen kunnen verschillende visies/invullingen betreffende het leergebied van toepassing zijn. Zo leggen sommige aanbieders van toetsen rekenen-wiskunde een groter accent op kale opgaven, terwijl anderen een groter accent op rekenen in contexten leggen.

Om een toets te kiezen die zo goed mogelijk aansluit bij het onderwijsaanbod, is het van belang een idee te hebben welke visie op het betreffende inhoudsgebied aan de toets ten grondslag ligt. In figuur 2.1 staat naast een beschrijving van de inhoud van de LVS-toets Rekenen-Wiskunde, ook een beknopte beschrijving van de toets Taal voor kleuters. In de toets Taal voor

kleuters voor groep 1 zijn alleen de onderdelen passieve woordenschat en kritisch luisteren opgenomen. Vanaf groep 2 worden ook andere taalaspecten getoetst zoals klank en rijm. Deze opbouw van de toetsen is conform de leerlijn Taal voor kleuters van de SLO (zie www.slo.nl > PO > thema's > Het jonge Kind > Leerplankader Jonge Kind) waarbij in groep 1 het accent ligt op woordenschat en begrijpend luisteren. Vanaf groep 2 komen er meer vaardigheden op het gebied van voorbereidend lezen bij, zoals schriftoriëntatie en auditieve synthese.

Toets	Wat meet de toets?	Hoe wordt getoetst?
Taal voor kleuters	<i>Vanaf groep 1 bevat de toets:</i> • Passieve woordenschat • Kritisch luisteren <i>Vanaf groep 2 bevat de toets ook:</i> • Klank en rijm • Eerste en laatste woord horen • Schriftoriëntatie • Auditieve synthese	De leerkracht leest de vragen voor. De leerlingen hebben bij elke vraag de keuze uit enkele plaatjes. De leerling geeft het antwoord door een streep onder het juiste plaatje te zetten. De specifieke onderdelen worden getoetst. De leerling kiest het plaatje: • waarop het genoemde woord staat afgebeeld • dat past bij de voorgelezen zin • dat rijmt op een genoemd woord of dat een genoemde klank bevat (woord dat begint met B) • waarop het laatste/eerstgenoemde woord staat afgebeeld • waarop de genoemde letter staat • waarop het genoemde woord staat afgebeeld, de leerkracht spreekt niet het gehele woord maar de losse klanken uit (b-oe-k)
Rekenen-Wiskunde	De toetsen bevatten opgaven uit de verschillende rekendomeinen, aansluitend bij de kerndoelen en referentieniveaus. Op basis van uitgewerkte leerlijnen en een analyse van het onderwijsaanbod zijn per toets opgaven gemaakt en geselecteerd bij de stof die de leerlingen tot dan toe gehad (zouden moeten) hebben. De toetsen meten de vaardigheid Rekenen-Wiskunde als totaal, dus zowel rekenen in contexten als het uitrekenen van kale opgaven.	De toets bestaat grotendeels uit open (kort-antwoord) opgaven. De toets bevat enkele meerkeuze opgaven. In de lagere niveaus leest de leerkracht of de computer de opgaven voor. In de hogere niveaus leest de leerling zelf de opgaven. Bij digitale afname heeft de leerling hierbij de mogelijkheid de opgave te laten voorlezen. De leerling schrijft bij papieren afname de antwoorden in een antwoordboekje of op een antwoordblad. Bij de digitale versie van de toets, voert de leerling zijn antwoord in met het bedieningspaneel dat naast de opgave op het beeldscherm staat afgebeeld.

Figuur 2.1: Wat en hoe meten LVS-toetsen?

2.2 Aansluiting toetsen bij kerndoelen en referentieniveaus

De inhoud van het PO, en daarmee ook de inhoud van de meeste toetsen, is gebaseerd op de kerndoelen (www.slo.nl/primair/kerndoelen/) en referentieniveaus (www.taalenrekenen.nl).

Kerndoelen

Kerndoelen zijn opgesteld door de SLO en vastgesteld door de overheid. In de kerndoelen is vastgelegd wat in het PO aan bod moet komen.

Er zijn kerndoelen voor de volgende vakgebieden:

- Nederlands;
- Engels;
- Fries (voor scholen in Friesland);
- Rekenen-Wiskunde;
- Oriëntatie op jezelf en de wereld;
- Kunstzinnige oriëntatie;
- Bewegingsonderwijs.

Een voorbeeld van een kerndoel Nederlands:

De leerlingen leren informatie te achterhalen in informatieve en instructieve teksten, waaronder schema's, tabellen en digitale bronnen.

Een voorbeeld van een kerndoel Rekenen-Wiskunde:

De leerlingen leren schattend tellen en rekenen.

Een voorbeeld van een kerndoel bewegingsonderwijs:

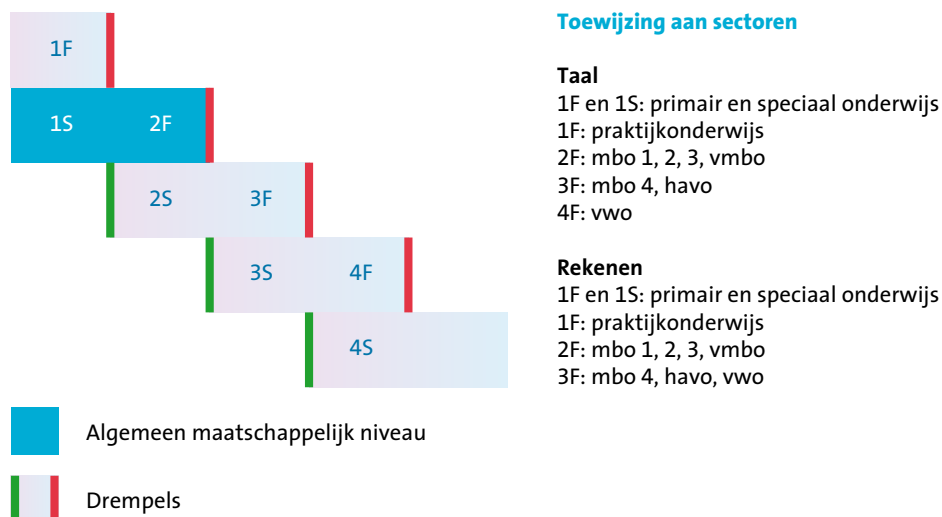
De leerlingen leren op een verantwoorde manier deelnemen aan de omringende bewegingscultuur en leren de hoofdbeginselen van de belangrijkste bewegings- en spelvormen ervaren en uitvoeren.

Deze formuleringen geven richting aan wat er in het onderwijs, en dus ook in toetsen, aan bod kan komen, maar geven ook nog ruimte voor interpretatie. Zo is het duidelijk dat er aandacht moet zijn voor schattend rekenen, maar zijn er geen regels over schattend rekenen in welk getalgebied. En moet het gaan om schattend rekenen met kale opgaven of juist in een context? Wat betreft bewegingsonderwijs is het duidelijk dat leerlingen verschillende sporten moeten beoefenen. Maar of de leerkracht wel of geen aandacht besteed aan trefbal, ligt niet vast. In deze voorbeelden is zichtbaar dat de kerndoelen globale beschrijvingen geven van de onderdelen waar een leerkracht aandacht aan moet besteden, maar dat het niveau en de exacte inhoud niet vastligt.

Referentieniveaus

Naast de kerndoelen zijn er in opdracht van het ministerie van OC&W referentieniveaus opgesteld voor Nederlandse taal en rekenen.

In de referentieniveaus is beschreven wat leerlingen moeten kennen en kunnen op verschillende drempelmomenten (zoals eind PO, eind vmbo, etcetera) in hun schoolloopbaan. Er zijn vier fundamentele niveaus en vier streefniveaus (1F t/m 4F en 1S t/m 4S) over alle onderwijssectoren (PO, SO, VO en MBO) heen. Niveau 1F en 1S zijn van toepassing op het einde van het basis- en speciaal (basis)onderwijs (zie figuur 2.2). Natuurlijk is het de bedoeling dat zoveel mogelijk leerlingen een zo hoog mogelijk niveau halen, maar ten minste 1F. Voor leerlingen die méér kunnen, kan toegewerkt worden naar niveau 1S en hoger. Het verschil tussen referentieniveaus en kerndoelen is dat kerndoelen meer globaal zijn geformuleerd en alleen de eisen voor het aanbod in het onderwijs vastleggen. Kerndoelen zijn dus aanbodsdoelen terwijl de referentieniveaus beheersingsdoelen zijn en specificeren wat leerlingen moeten kennen en kunnen.



Figuur 2.2: Opbouw referentieniveaus

De referentieniveaus vormen een nuttig instrument om opbrengstgericht werken in de praktijk te realiseren en op deze manier de taal- en rekenprestaties te verbeteren. Met de referentieniveaus kunnen doelen worden vastgesteld en leerprestaties van individuele leerlingen worden gemeten, gevolgd en waar nodig bijgesteld. Bovendien bevorderen de referentieniveaus een soepelere aansluiting tussen de verschillende onderwijssectoren doordat in elke fase van de opleiding van een leerling zijn of haar niveau met behulp van de referentieniveaus op eenduidige wijze kan worden vastgelegd en kan worden doorgegeven aan het vervolgonderwijs. Het VO weet zo dus beter welk niveau ze kan verwachten van een leerling die uit het PO komt en kan hier op deze manier beter op aansluiten.

Voor Taal zijn referentieniveaus opgesteld voor de volgende domeinen:

- Mondelinge taalvaardigheid, met de drie subdomeinen gespreksvaardigheid, luistervaardigheid en spreekvaardigheid.
- Leesvaardigheid, met de twee subdomeinen lezen van zakelijke teksten en lezen van fictionele, narratieve en literaire teksten.
- Schrijfvaardigheid. Schrijven wordt ruim opgevat; het gaat om het produceren van creatieve en zakelijke teksten.
- Begrippenlijst en taalverzorging. In de begrippenlijst staan termen en concepten die leerlingen en docenten nodig hebben om over taalvaardigheid van gedachten te wisselen. Bij taalverzorging gaat het om zaken die in dienst staan van een verzorgde, schriftelijke taalproductie waarbij men kan denken aan het leergebied spelling.

De vier domeinen van taal zijn ieder op vier niveaus beschreven. De niveaus geven een opklimmende moeilijkheidsgraad in basiskennis en vaardigheden aan. Elk fundamenteel niveau omvat het voorgaande niveau: leerlingen op niveau 2F beheersen dus ook niveau 1F. Aangezien voor taal niveau 1S en 2F gelijk zijn, kan bij het behalen van een fundamenteel niveau het volgende niveau als streefniveau worden gezien.

Binnen rekenen worden vier domeinen onderscheiden:

- Getallen
- Verhoudingen
- Meten en Meetkunde
- Verbanden

Elk domein is bij rekenen opgebouwd uit de onderdelen:

- notatie, taal en betekenis, waarbij het gaat om de uitspraak, schrijfwijze en betekenis van getallen, symbolen en relaties en om het gebruik van wiskundetaal;
- met elkaar in verband brengen, waarbij het gaat om het verband tussen begrippen, notaties, getallen en dagelijks spraakgebruik;
- gebruiken, waarbij het gaat om rekenvaardigheden in te zetten bij het oplossen van problemen.

Anders dan bij taal gaat het bij rekenen om van elkaar onderscheiden fundamentele niveaus en streefniveaus. Deze zijn beide op drie niveaus beschreven, dus voor het fundamentele niveau 1F, 2F en 3F en voor het streefniveau 1S, 2S en 3S. De fundamentele niveaus richten zich op basale kennis en inzichten en zijn gericht op een meer toepassinggerichte benadering van rekenen. De streefniveaus bereiden al voor op de meer abstracte wiskunde. Bij Rekenen is er geen vierde niveau, omdat dat het domein van de wiskunde is.

Referentieniveaus zijn geen toetsscores, maar kunnen wel vertaald worden naar toetsscores. Dit kan door de in de referentieniveaus beschreven inhoud te koppelen aan vaardigheidsscores voor de verschillende leergebieden rondom Taal en Rekenen. Zo kan met een toets gemeten worden in hoeverre leerlingen de referentieniveaus beheersen. Referentieniveaus zelf zijn dus geen 'getal', maar zijn doelen. Voor de toetsen worden deze vertaald naar een meetbaar getal.

In figuur 2.3 staat een uitwerking van de referentieniveaus voor Rekenen, onderdeel getallen en getalrelaties. Dit betreft een uitwerking van onder meer het volgende kerndoel:

De leerlingen leren structuur en samenhang van aantallen, gehele getallen, kommagetallen, breuken, procenten en verhoudingen op hoofdlijnen te doorzien en er in praktische situaties mee te rekenen. Voor alle leerlingen geldt dat zij inzicht moeten hebben in de getallenlijn, met daarop ook eenvoudige decimale getallen zoals 12,5. Dit is het 1F-niveau. In de tweede kolom is zichtbaar dat het 1S-niveau ook breuken en decimale getallen in het algemeen als doel weergeeft. Leerlingen die aan het einde van het primair onderwijs 1S beheersen, moeten dus ook getallen zoals 16,25 en $\frac{1}{8}$ op een getallenlijn kunnen weergeven.

12 jaar	1 - fundament	1 - streef
B Met elkaar in verband brengen – Getallen en getalrelaties – Structuur en samenhang	Paraat hebben – tienstructuur – getallenrij – getallenlijn met gehele getallen en eenvoudige decimale getallen	Paraat hebben – getallenlijn, ook met decimale getallen en breuken
	Functioneel gebruiken – vertalen van eenvoudige situatie naar berekening – afronden van gehele getallen op ronde getallen – globaal beredeneren van uitkomsten – splitsen en samenstellen van getallen op basis van het tientallig stelsel	1F en ook: – Moeilijkere getallen – Complexere situaties – Abstracter/formeler – vertalen van eenvoudige situatie naar berekening – afronden van gehele getallen op ronde getallen – afronden binnen gegeven situatie: 77,6 dozen berekend dus 78 dozen kopen
	Weten waarom – structuur van het tientallig stelsel	Weten waarom – opbouw decimale positiestelsel – redeneren over breuken, bijvoorbeeld: is er een kleinste breuk?

Figuur 2.3: 1F- versus 1S-niveau - referentieniveaus (bron: www.slo.nl)

Niet alleen de inhoud van de toetsen sluit aan bij de referentieniveaus, bij de volgtoetsen van Cito voor de hogere leerjaren wordt vanaf schooljaar 2014-2015 ook de koppeling met de referentieniveaus gemaakt. Door afname van zogenoemde referentiesets is voor verschillende leergebieden vastgesteld welke vaardigheidsscore hoort bij de verschillende referentieniveaus. In de rapportages voor het PO kan zo aangegeven worden in hoeverre een leerling referentieniveau 1F of 1S heeft behaald. Scoort de leerling boven de vaardigheidsscore die hoort bij referentieniveau 1F? Dan heeft hij dit niveau behaald. Scoort de leerling eronder? Dan is dat een indicatie dat deze leerling de doelen bij 1F nog onvoldoende beheerst.

Kerdoelen en referentieniveaus versus het S(B)O

De kerndoelen en referentieniveaus voor het PO zijn ook van toepassing op leerlingen uit het SBO. Voor het speciaal onderwijs zijn er aparte kerndoelen; de referentieniveaus zijn echter integraal van toepassing op het SO met uitzondering van zeer moeilijk lerende (ZML) en meervoudig gehandicapte (MG) leerlingen. Voor alle clusters van het SO behalve voor ZML en MG-leerlingen komen de kerndoelen grotendeels overeen met de doelen voor het primair onderwijs. De scholen dienen dus hetzelfde onderwijsaanbod te verzorgen als binnen het primair onderwijs. Voor het SO zijn er naast de genoemde vakgebieden nog enkele aanvullingen op de kerndoelen die onder andere gericht zijn op het vergroten van zelfredzaamheid. Voor blinde en slechtziende leerlingen zijn er bijvoorbeeld in de kerndoelen extra doelen opgenomen over verplaatsen in de ruimte.

Omdat de inhoud van het onderwijs in het SO in hoofdlijnen overeenkomt met het reguliere onderwijs, komen ook de toetsinhouden van het SO overeen. Uitzondering zijn (wederom) de ZML en MG-leerlingen. Omdat de kerndoelen voor deze groep leerlingen duidelijk andere doelen aangeven, zijn er speciale toetsen in ontwikkeling die aansluiten bij deze kerndoelen.

Zie voor de kerndoelen speciaal onderwijs: www.slo.nl > (V)SO > (V)SO > Kerndoelen

2.3 Van kerndoelen en referentieniveaus naar leerlijnen

In de kerndoelen ligt wettelijk vast wat gedurende 8 jaar basisonderwijs aan bod moet komen. In de referentieniveaus ligt vast wat een leerling aan het einde van het primair onderwijs moet beheersen.

De uitgewerkte inhoud en wat de leerling in verschillende leerjaren tot en met groep 8 moet beheersen, is niet wettelijk vastgelegd. Scholen, en dus ook ontwerpers van methodes en constructeurs van toetsen, zijn vrij om zelf te bepalen op welke wijze en in welke volgorde zij de leerstof aanbieden en welke leerstof wordt getoetst. Daarbij kan gebruik worden gemaakt van de uitwerking van kerndoelen en referentieniveaus in tussendoelen en leerlijnen. De tussendoelen en leerlijnen verdelen de totale leerstof van een inhoudsgebied over de schooljaren op zo'n manier dat wordt aangesloten bij de verschillende ontwikkelingsstadia van leerlingen en er een logische volgorde in het leerstofaanbod wordt gehanteerd. Er zijn verschillende uitwerkingen van leerlijnen beschikbaar, zie bijvoorbeeld:

www.leerlijntaal.nl voor een uitwerking van de leerlijnen taal in relatie tot de referentieniveaus;

www.rekenlijn.nl voor een uitwerking van de leerlijnen rekenen in relatie tot de referentieniveaus rekenen;

www.cedgroep.nl voor een uitwerking van de kerndoelen voor het speciaal onderwijs;

<http://tule.slo.nl/> voor een uitwerking van kerndoelen in tussendoelen.

Tussendoelen vormen een basis voor leerlijnen. Kerndoelen zijn breed geformuleerd, tussendoelen zijn iets specifiekere. Tussendoelen bestaan uit inhouden en activiteiten die een uitwerking zijn van de kerndoelen. Voor het PO en SO zijn er aparte tussendoelen, uitgewerkt als beschrijvingen van inhouden en activiteiten per leerjaar, om zo handreikingen te geven over

de organisatie van het onderwijsaanbod in een bepaald leerjaar op een zodanige wijze dat het aansluit bij de kerndoelen. Leerlijnen bestaan uit duidelijk gespecificeerde leerdoelen. Leerdoelen beschrijven wat een leerkracht de leerlingen aan kan bieden, zodat de leerkracht vervolgens kan besluiten hoe hij dit het beste kan aanbieden.

Kerndoel 6 (Nederlands, schriftelijke taal) luidt: *De leerlingen leren informatie en meningen te ordenen bij het lezen van school- en studieteksten en andere instructieve teksten, en bij systematisch geordende bronnen, waaronder digitale bronnen.*

Dit kerndoel is op de Tule-site uitgewerkt zoals weergegeven in figuur 2.4.



Figuur 2.4: Uitwerking Kerndoel 6 (Nederlands) op de Tule-site

Per twee leerjaren (voor groep 1/2, groep 3/4, groep 5/6, groep 7/8) staan er aanwijzingen voor het onderwijsaanbod. In groep 1/2 kan aandacht besteed worden aan voorbereidende vaardigheden om te kunnen ordenen en sorteren, zoals het rubriceren van plaatjes. In groep 3 en 4 leren de leerlingen een tekst systematisch te lezen door de tekst nogmaals te herlezen, bijvoorbeeld door het maken van een woordspin of door belangrijke zinnen te onderstrepen. Ook is er aandacht voor signaalwoorden die een tijdsvolgorde aangeven en de leerlingen leren vraagwoorden te herkennen.

In groep 5 en 6 wordt dit uitgebreid met o.a. het leren maken van diverse soorten schema's en door het maken van een uittreksel. Wat betreft signaalwoorden komen daar ook tegenstellingen bij, maar leren de leerlingen ook om teksten te lezen waarin verschillende tekens zoals dubbele punt worden gebruikt.

In groep 7 en 8 leren de leerlingen nog complexere schema's en tabellen te maken aan de hand van een tekst en leren ze een echte samenvatting te maken. Ook komen er extra signaalwoorden bij zoals voorwaardelijke structuren (als...dan).

De uitwerkingen van de tussendoelen leerlijnen (TULE) zijn herkenbaar in zowel toetsen als in lesmateriaal. In figuur 2.5a staat een fragment van Begrijpend lezen uit Nieuwsbegrip voor groep 5 en 6, waarin zichtbaar is dat de leerlingen leren werken met schema's om tegenstellingen aan te geven (zoals genoemd in de eerste rij, derde kolom van de tabel in figuur 2.4). In figuur 2.5b is in het fragment voor groep 7/8 zichtbaar dat de leerlingen zelf een samenvatting moeten maken (zoals genoemd in de eerste rij, vierde kolom van de tabel in figuur 2.4).

Opdracht 3 (samen): Een schema invullen

In de tekst vertelt Bernice veel over haar tocht naar de Zuidpool. Welke dingen vond ze leuk? Welke dingen vond ze moeilijk? Vul het schema in. Er staan al een paar antwoorden.

Leuk aan haar reis vond Bernice:

Dit was best wel zwaar:

1. *Aandacht voor het klimaat vragen.*

1. *Ze kreeg een longinfectie.*

.....

.....

2.....

2.....

.....

.....

3.....

3.....

.....

.....

4.....

4.....

.....

.....



Figuur 2.5a: Fragment uit Nieuwsbegrip niveau A (groep 5/6)

Opdracht 1 (klas): Tekst lezen

1. Lees de tekst met het stappenplan.
2. Schrijf twee moeilijke woorden uit de tekst op. Wat betekenen deze woorden?
Gebruik de woordhulp op het stappenplan.

Opdracht 2 (klas en alleen): Samenvatten

Uitleg

Een tekst bestaat meestal uit een titel, een inleiding en andere stukjes tekst. Boven elk stukje tekst staat een kopje. Het kopje dat boven een stukje tekst staat geeft vaak al aan waar dat stukje over gaat en wat de belangrijkste informatie in dat stuk tekst is.

De tekst 'Eerste Nederlandse vrouw op de Zuidpool' heeft een inleiding en vier stukjes tekst. Boven de stukjes tekst staan kopjes. Maar dit zijn geen goede kopjes. In de kopjes staat niet het belangrijkste van het stukje tekst. In de opdrachten ga je zelf betere kopjes bedenken.

Tipl

Je kunt zelf kopjes bedenken voor stukjes tekst. Lees het stukje tekst dan goed. Zoek de belangrijkste informatie in het stukje. Dat zet je in het kopje.

1. Lees het stukje tekst onder 'Op ski's'. Zoek de belangrijkste informatie. Wat is een beter kopje voor dit stukje? Kruis aan.
 - ☐ Naar de Zuidpool
 - ☐ Op de Noordpool en Siberië
 - ☐ IJskappen

2. Lees het stukje tekst onder 'Mars'. Bedenk zelf een beter kopje voor dit stukje. Schrijf op.

.....

3. Wat is een beter kopje voor het stukje tekst onder 'Longontsteking'? Schrijf op.

.....

4. Wat is een beter kopje voor het stukje tekst onder 'Contact met scholen'? Schrijf op.

.....



Figuur 2.5b: Fragment uit Nieuwsbegrip niveau B (groep 7/8)

In figuur 2.6a en 2.6b is zichtbaar dat de inhoud van TULE ook terugkomt in de toetsen. In de voorbeelden komen de inhoudsdoelen terug zoals die in de tussendoelen bij het betreffende leerjaar staan beschreven. In het fragment uit de toets voor groep 6 is bijvoorbeeld het woord 'doordat' gebruikt. Dit correspondeert met de omschrijving zoals genoemd in rij 2, kolom 3 van figuur 2.4 waarbij het gaat om het begrijpen van een oorzaak-gevolgrelatie. In het fragment uit de toets groep 7 wordt de structuur als...dan gebruikt, zoals weergegeven in rij 2, kolom 4 van figuur 2.4 waarbij het gaat om het kunnen hanteren van voorwaardelijke structuren.

Ochtend

Zodra de lichtcellen alle drie weer voldoende licht meten, gaat er opnieuw een signaal het elektriciteitsnet op: de lichten gaan weer uit! De hele aansturing gebeurt dus 3, afhankelijk van hoe donker of licht het is. Voor een lantaarnpaal maakt het **daardoor** ook niet uit of het zomer of winter is.

Figuur 2.6a: Fragment uit toets Begrijpend lezen M6, gebruik van verwijswoord 'daardoor'.

Sommige moleculen botsen wat minder vaak en die zijn veel eerder bij jou. Maar helaas: onze neus is niet zo goed. Aan een paar 5 heeft hij niets. Het moeten er veel meer zijn, voor je iets ruikt. Een hond heeft een neus die een miljoen keer beter is dan die van jou. Als een hond een paar moleculen in zijn neus krijgt, 6. Dus **als** je hond opeens gestrekt op de grond ligt, met zijn poten in de lucht, **dan** ruik jij waarschijnlijk een minuutje later dat de koekjes ernstig zijn aangebrand ...

Figuur 2.6b: Fragment uit toets Begrijpend lezen M7, gebruik van voorwaardelijke structuren zoals als...dan³

Aan elke methode ligt een eigen leerlijn, gebaseerd op de beschikbare uitwerkingen van kerndoelen en tussendoelen, ten grondslag. Aan elk doel uit de leerlijn koppelt de methode voldoende lesmateriaal om zo alle leerdoelen voldoende aan bod te laten komen en de inhoudsdoelen uit de kerndoelen te dekken.

De leerlijnen van de methodes vormen de basis voor methodegebonden toetsen. Een leerkracht kan na het afnemen van een methodegebonden toets zien welke lesdoelen de leerlingen wel en welke ze (nog) niet beheersen.

Ook methodeonafhankelijke toetsen zijn, zoals zichtbaar in het voorbeeld van figuur 2.6, gebaseerd op de kerndoelen en de uitwerking daarvan in tussendoelen en leerlijnen. Bij het maken van deze toetsen wordt eerst een domeinbeschrijving, een beschrijving van het inhoudsgebied aan de hand van een lijst van leer- en vormingsdoelen, opgesteld. Deze beschrijving is gebaseerd op de kerndoelen basisonderwijs, de uitwerking daarvan in tussendoelen en leerlijnen, analyses van onderwijsmethodes en vakinhoudelijke en vakdidactische publicaties.

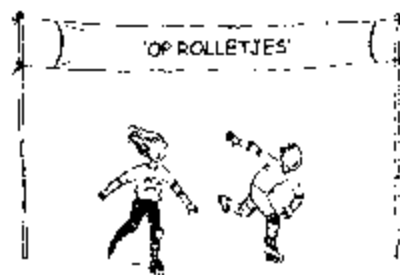
In figuur 2.7 staat een gedeelte van de inhoudsbeschrijving van de toets Rekenen-Wiskunde voor midden groep 5. Hierin staat onder andere beschreven dat de opgaven in de toets t/m 1000 gaan en welke strategieën ingezet kunnen worden om de opgaven op te kunnen lossen. Naast de omschrijving staat een voorbeeldopgave om een indicatie te geven welk type opgave bij de omschrijving past.

³ Figuur 2.6a en 2.6b bevatten open-plaatsvragen waarbij de leerling op de streep met het cijfer een passend woord in moet vullen. De leerling heeft hierbij keuze uit vier alternatieven.

2 Hoofdrekenen

2.1 Optellen

- Optellen in het getallengebied tot 1000, waarbij werkwijzen gebruikt kunnen worden als hergroeperen, splitsen, doortellen met sprongen en één of meer getallen veranderen en daarvoor zonodig een correctie toepassen. De lastigste typen uit het getallengebied tot 100 (zoals $34 + 58$) komen nu ook voor.
- Toepassen van het optellen in diverse contexten: samennemen, toevoegen en vergelijken.



De moeder van Karel is 39 jaar. Zijn oma is 25 jaar ouder.
Hoe oud is de oma van Karel?

_____ jaar

Figuur 2.7: Fragment inhoudsbeschrijving van de toets Rekenen-Wiskunde M5, gedeelte hoofdrekenen

Aan de hand van de leerlijnen kan een leerkracht de resultaten op een methodeonafhankelijke toets eenvoudig vertalen naar onderwijsaanbod. Hierbij is stap 1 om *indien er een gespecificeerde leerlijn aan de toets ten grondslag ligt, deze te vergelijken met de leerlijn uit de eigen methode*. In veel gevallen zal deze qua inhoud vrijwel overeenkomen. Uiteraard zullen er accentverschillen zijn omdat sommige methodes wat uitgebreider stilstaan bij het ene leerstofonderdeel (bijvoorbeeld springen op de getallenlijn), en andere weer bij een andere leerstofeenheid (zoals handig rekenen). De grootste verschillen tussen methodes bestaan bij Spelling. Doordat de methodemakers de verschillende spellingscategorieën in een andere volgorde aanbieden, is de spellingscategorie in de LVS-toets nog niet altijd aan bod geweest in de methode. Als leerlingen op een specifiek onderdeel van de toets uitvallen en dit nog in de methode aan bod moet komen, is het niet meteen noodzakelijk om extra actie te ondernemen, maar kan de reguliere methode gevolgd blijven worden zodat de categorie alsnog aan bod komt. Alleen indien een categorie wel in het onderwijsaanbod is behandeld maar nog onvoldoende wordt beheerst, is extra actie noodzakelijk.

Stap 2 is om met hulp van het toetsresultaat het niveau van de leerling te bepalen. Door te kijken naar het functioneringsniveau (zie hoofdstuk 3) krijgt de leerkracht een indicatie van welk niveau op de leerlijn ongeveer aansluit bij het functioneren van de leerling. Het functioneringsniveau is echter niet hetzelfde als het beheersingsniveau. Het is daarom belangrijk om het functioneringsniveau dat uit de toets komt altijd in relatie te bekijken tot eigen observaties en resultaten op methodegebondentoetsen. Functioneringsniveau 'M5' geeft een indicatie dat een leerling lesstof tot en met de eerste helft van groep 5 vlot zal oppikken. Het is raadzaam om met methodegebondentoetsen te checken of de leerling deze stof daadwerkelijk beheerst.

Wanneer een leerling uit groep 3 een functioneringsniveau M5 haalt, betekent dit niet dat de lesstof van groep 4 overgeslagen kan worden. Een functioneringsniveau M5 geeft in dit geval aan dat de leerling de lesstof van groep 4 waarschijnlijk vlot kan doorlopen, maar het volgen van de leerlijn, en dus instructie en oefening (al dan niet versneld) van deze lesstof, is voor een goede ontwikkeling noodzakelijk.

Zie voor meer informatie over het gebruik van functioneringsniveau bij het vaststellen van het aanbod de flyer 'Functioneringsniveaus, zo gebruikt u ze' op www.cito.nl

Stap 3 is om in de leerlijn te *bekijken welke specifieke doelen de leerling wel en niet beheerst*. Hiervoor kunnen afhankelijk van de toets analyses van de LVS-toetsen als aanknopingspunt gebruikt worden. Wanneer bijvoorbeeld uit een categorieënanalyse Rekenen-Wiskunde (zie pagina 70) blijkt dat een leerling uitvalt bij optellen en aftrekken, kan de leerkracht in de leerlijn bij optellen en aftrekken ook bij doelen uit lagere niveaus kritisch kijken welke de leerling wel en niet beheerst. Om op doelniveau vast te stellen wat een leerling wel en niet kan, zijn methodegebonden toetsen een goed hulpmiddel. Hiervoor is het waardevol om vast te leggen bij welke doelen uit de methode de verschillende opgaven horen.

2.4 De constructie van LVS-toetsen

Zoals genoemd in paragraaf 2.1 komt een methodeonafhankelijke toets volgens een zorgvuldige procedure tot stand. Om te illustreren hoe zorgvuldig de toetsen worden samengesteld beschrijft onderstaand stappenplan globaal de onderdelen van het constructieproces van de meest gebruikte methodeonafhankelijke toetsen, namelijk de LVS-toetsen van Cito.

Stap 1: Toetsdoel bepalen

Een school 'toetst niet om te toetsen'. Toetsen hebben altijd een doel. Bij het bepalen van het toetsdoel kijken we naar twee vragen:

- **Waarom willen we meten?**

In hoofdstuk 1 staan verschillende doelen om een toets af te nemen, denk aan het classificeren van leerlingen met een eindtoets of het selecteren van leerlingen wanneer toetsresultaten worden ingezet voor de indicatiestelling. Sommige toetsen, zoals de LVS-toetsen AVI/DMT of Spelling, geven behalve een indicatie van het niveau van een leerling ook meer diagnostische informatie zoals welk onderdeel van het vakgebied een leerling beter of juist minder goed beheerst. Ook het bieden van dergelijke handvatten voor het eventueel bijstellen van het onderwijsleerproces kan een doel van toetsen zijn.

- **Wat willen we meten?**

Wanneer duidelijk is waarom gemeten wordt, is het belangrijk om goed te definiëren wat precies gemeten moet worden. Spelling als geheel, of alleen actieve spelling? Als we Begrijpend lezen willen meten, wat verstaan we hier dan onder? En definiëren we Rekenen als rekenen in een context of bijvoorbeeld als kunnen automatiseren? Op basis van wat gemeten dient te worden, kiezen we een toetsvorm die daar het beste bij aansluit. Voor de makers van toetsen is het belangrijk om tijdens het maken van een toets steeds te realiseren wat deze beoogt te meten. Zo willen we met de toets LVS-Rekenen-Wiskunde van Cito de rekenvaardigheid in alledaagse situaties vaststellen, niet het maken van kale sommen. De toets moet dus opgaven bevatten in relevante contexten. Voor gebruikers is het ook goed voor ogen te houden waarom en wat ze willen meten. Met de toets kan een leerkracht dus uitspraken doen over de rekenvaardigheid in contexten, niet over het automatiseren van kale opgaven.

Het doel van de toets bepaalt onder andere het type opgaven, de moeilijkheidsgraad van de opgaven en de variatie in moeilijkheidsgraad tussen de opgaven. In een groep zitten bijvoorbeeld leerlingen met uiteenlopende niveaus, een volgtoets bevat daarom ook moeilijkere en makkelijkere opgaven. Een toets om bij een specifieke groep leerlingen, bijvoorbeeld hoogbegaafde leerlingen, vast te stellen welke vaardigheid zij beheersen, kan bestaan uit opgaven van ongeveer gelijk niveau.

Stap 2: Toetsspecificaties

Een toets moet voldoen aan verschillende eisen (zie ook paragraaf 2.1). Een groot deel van de eisen komt voort uit wensen van scholen, kenmerken van de leerlingen en het doel van de toets, een ander deel bestaat uit eisen met betrekking tot bijvoorbeeld betrouwbaarheid (COTAN en Expertgroep PO, zie ook hoofdstuk 4). De volgende aandachtspunten zijn onder andere van belang:

- **Omvang:** Hoe lang mag een toets duren en uit hoeveel opgaven moet de toets bestaan? Een taak uit de Rekentoets voor groep 3 is bijvoorbeeld korter dan een taak voor groep 8. Deels omdat de leerlingen in groep 3 minder lang achter elkaar kunnen werken, deels omdat de stof in groep 8 meer omvat dan in groep 3. Om betrouwbaar te meten is een minimaal aantal opgaven nodig. Om rekening te houden met de concentratie van de leerling en tegelijkertijd toch voldoende informatie te kunnen verzamelen, kan de toets in meerdere delen ('taken') opgeknipt worden.
- **Inhoud:** Welke (sub)onderdelen komen in de toets aan de orde? In een schema leggen we vast hoeveel opgaven van elke soort in de toets komen. Bijvoorbeeld: Welke spellings-categorieën komen in de toets? En hoeveel opgaven van elke categorie? Een dergelijk schema noemen we een toetsmatrijs. Een voorbeeld van een toetsmatrijs van de toets Taal voor Kleuters staat in figuur 2.8. Hierin is zichtbaar dat in de toets van groep 1 bijvoorbeeld 32 opgaven over passieve woordenschat zitten, in groep 2 zijn dit 15 opgaven.

	Aantal opgaven Groep 1	Aantal opgaven Groep 2
Passieve woordenschat	32	15
Kritisch luisteren	16	15
Klank en rijm		8
Eerste en laatste woord horen		6
Auditieve synthese		8
Schriftoriëntatie		8
Totaal	48	60

Figuur 2.8: Toetsmatrijs van de toets Taal voor Kleuters

- **Niveau:** Bij welke leerlingen wordt de toets afgenomen? Hierbij wordt niet alleen gekeken naar het gemiddelde niveau, maar wordt ook bekeken of deze leerlingen uiteenlopende niveaus hebben óf dat de niveaus juist dicht bij elkaar liggen. De standaard LVS-toetsen zoals Rekenen-Wiskunde bevatten opgaven van uiteenlopende niveaus. Zo kunnen minder vaardige leerlingen een aantal opgaven goed maken, maar hebben ook leerlingen die vaardiger zijn dan gemiddeld de kans om te laten zien wat ze kunnen. Andere toetsen, zoals bijvoorbeeld het Screeningsinstrument Dyslexie, zijn juist bedoeld voor minder vaardige leerlingen. Het doel van de toets bepaalt dus dat deze toets voornamelijk opgaven van een lager niveau bevat en nauwelijks opgaven die bedoeld zijn voor leerlingen met een meer dan gemiddelde vaardigheid.
- **Opgavenvorm:** Hoe zien de opgaven eruit? Hierbij worden bijvoorbeeld de volgende keuzes gemaakt: open vragen of meerkeuzevragen, papier of digitaal, met of zonder antwoordblad, met of zonder plaatjes, individuele of groepsgewijze afname, soort vraagstelling etcetera. Meerkeuzeopgaven zijn efficiënt om te scoren en er bestaat de minste verwarring over of

een antwoord als goed of fout beoordeeld moet worden. Indien gekozen wordt voor meerkeuzeopgaven, wordt vervolgens bepaald hoeveel antwoordalternatieven er bij elke vraag horen. In de meeste gevallen bestaat de vraag uit 3 of 4 antwoordalternatieven, naast het juiste antwoord zijn er dus 2 of 3 zogenaamde afleiders. Bij te veel afleiders komen er onwaarschijnlijke antwoorden bij te staan die leerlingen niet zullen kiezen. Bij te weinig afleiders wordt de gokkans te groot.

Bij zowel open als gesloten vragen zijn er verschillende vraagsoorten mogelijk. Denk bij begrijpend lezen bijvoorbeeld aan de 'open-plaatsvragen' (zie figuur 2.9), of de hussel-verhaaltjes (zie figuur 2.10) waarbij de leerling de zinnen in de goede volgorde moet plaatsen.

Een bijzonder knaagdier

Net als konijnen kun je ratten in een hok houden.

Je kunt deze ratten, net als een tam konijn, gerust oppakken.

De meeste mensen 1 .

3 Wat past het **best** op plaats 1?

- A** hebben geen ruimte voor ratten
- B** hebben verkeerde ideeën over ratten
- C** jagen niet op konijnen
- D** lezen nooit een boek over ratten
- E** vinden konijn lekker

Figuur 2.9: Openplaatsvraag uit Begrijpend lezen M5

2 Zoek de eerste zin.

- A** Al mijn vrienden en vriendinnen vinden dat leuk.
- B** Misschien gaan we naar het zwembad.
- C** Voor mijn verjaardag geef ik een feestje.
- D** Wat zullen we dan gaan doen?
- E** Ze kunnen allemaal goed zwemmen.

Figuur 2.10: Husselverhaaltje uit Begrijpend lezen M5

Er wordt steeds gezocht naar een efficiënte toetsvorm die voor de leerlingen aantrekkelijk en begrijpelijk is.

Stap 3: Opgaven maken

Leerkrachten, Pabodocenten en andere ervaringsdeskundigen maken samen met toetsdeskundigen opgaven:

- Degenen die de opgaven maken, worden constructiegroepleden genoemd. Zij krijgen een constructieopdracht. Zo weten zij voor welke inhoud en leeftijdsgroep ze opgaven moeten maken en van welk vraagtype (met/zonder plaatjes, open vraag/meerkeuze).
- De constructiegroepleden maken individueel opgaven en leveren deze in.
- Tijdens een constructiegroepvergadering bespreken de constructiegroepleden elkaars opgaven. Hierbij wordt onder andere kritisch gekeken naar inhoud, taalgebruik, niveau en context van de opgaven.
- Een toetsdeskundige werkt de opgaven verder uit en zet indien nodig tekenopdrachten uit bij tekenaars om de benodigde tekeningen te maken.

Stap 4: Proefafname

Nadat alle opgaven meerdere malen zijn bekeken en verbeterd volgt de proeftoetsing. Van elke opgave willen we weten hoe moeilijk deze opgave is, hoe goed de opgave onderscheid maakt tussen vaardige en minder vaardige leerlingen en of de opgave meet wat we willen meten. Een grote, representatieve groep leerlingen maakt daarom de opgaven.

Stap 5: Opgaven selecteren

De gegevens uit de proefafname vormen de basis voor de selectie van opgaven.

Van elke opgave bepalen we de moeilijkheidsgraad door te kijken naar het percentage leerlingen dat de opgaven goed maakt. Te makkelijke of te moeilijke opgaven worden verwijderd of eventueel bewaard voor een andere toets.

Bij meerkeuze-opgaven bekijken we of het alternatief dat correct is duidelijk door meer leerlingen gekozen wordt dan de overige alternatieven. Springt het 'correcte' alternatief (het juiste of beste alternatief) er niet duidelijk uit, dan is de opgave 'ambigu' en komt deze niet in aanmerking voor opname in de uiteindelijke toets. Daarnaast moeten alle alternatieven waarschijnlijk zijn. Wanneer één van de alternatieven in de proefafname zelden of nooit is gekozen, komt deze opgave ook niet in de toets.

Ook kijken we welke leerlingen een opgave goed of juist fout maken. Maken vaardige leerlingen relatief eenvoudige opgaven bijvoorbeeld inderdaad goed? Hoe vaardiger de leerling, hoe groter de kans moet zijn om de opgave goed te kunnen maken.

Uiteraard nemen we ook reacties uit het veld mee. Welke opgaven zijn volgens leerkrachten minder goed of juist wel goed? Soms verwijderen we opgaven die statistisch heel goed functioneren, omdat leerkrachten ze als 'slechte opgaven' ervaren.

Al deze informatie gebruiken we om te komen tot een verzameling opgaven waaruit een toets kan worden samengesteld.

Het is belangrijk om te realiseren dat het antwoord van een leerling op een individuele opgave weinig zegt over de werkelijke vaardigheid van een kind. Want daarbij kunnen allerlei toevalligheden een grote rol spelen. Het gaat juist om de antwoorden op een grote reeks opgaven. Daardoor worden toevalligheden uitgesloten en ontstaat informatie waar leerkrachten staat op kunnen maken.

Stap 6: Toets samenstellen

Uit de geselecteerde opgaven wordt een toets samengesteld. Hierbij wordt gekeken naar:

- **Gevarieerde inhoud:** De verschillende onderdelen van een vakgebied moeten volgens afgesproken verhoudingen, al of niet vastgelegd met een toetsmatrijs, in de toets naar voren komen.
- **Gevarieerde moeilijkheidsgraad:** De toets bevat gemakkelijke, moeilijkere en opgaven van gemiddelde moeilijkheidsgraad. In de toets worden deze meestal afwisselend aangeboden om te voorkomen dat een leerling verveeld of gefrustreerd raakt.
- **Gevarieerde opgavenvormen:** Door afwisseling in opgavevorm, bijvoorbeeld lange vragen versus korte vragen en/of opgaven met plaatje versus opgaven zonder plaatje, ziet de toets er aantrekkelijk uit en kan de leerling de aandacht beter vasthouden gedurende de hele toets.

Stap 7: Normeringsonderzoek

- Een grote, representatieve groep leerlingen maakt de opgaven uit de definitieve toetsboekjes. Hierbij worden alle aanwijzingen voor de afname uit de handleiding, die speciaal voor het normeringsonderzoek is opgesteld, opgevolgd. Alleen door standaardisatie is het mogelijk de toetsprestaties van leerlingen met elkaar te vergelijken. Dat betekent dat ook bij de uiteindelijke afname van de toets de normeringstabellen alleen gebruikt kunnen worden als de toets op dezelfde wijze en volgens de aanwijzingen in de toetshandleiding wordt afgenomen.

- De leerlinggegevens worden teruggestuurd naar de toetsconstructeurs. Hier berekenen we de vaardigheidsniveaus en stellen we de normgroepen (I t/m V en A t/m E) vast. De normgegevens komen dus tot stand op basis van de gegevens uit het normeringsonderzoek.

Stap 8: Uitgave

De uitgave bestaat uit:

- **Het toetsmateriaal:** Dit zijn de toetsboekjes en eventuele antwoordbladen of voorleesbladen.
- **De registratieformulieren:** Hierop kan een leerkracht bijhouden welke vorderingen een leerling maakt. Dit kan uiteraard ook met behulp van een computerprogramma, zoals het Computerprogramma LOVS.
- **Achtergronddocumenten bij de toets** (al dan niet gecombineerd in één document):
 - *Een handleiding:* Deze bevat een duidelijke omschrijving van hoe de toets moet worden afgenomen. Daardoor is de manier van afname op verschillende scholen en op verschillende toetsmomenten zo gelijk mogelijk: een gestandaardiseerde afname. Dit maakt een eerlijke vergelijking van resultaten mogelijk. Door aanwijzingen over de scoring kunnen scholen de juiste score van de leerling bepalen en kunnen zij de resultaten gebruiken om de leerling te volgen.
 - *Een inhoudsverantwoording:* Hierin staat beschreven hoe de toets tot stand is gekomen. Bij sommige toetsen staat hier ook informatie waarmee scholen betekenis kunnen geven aan een toetsscore. Wat kan een leerling bij een bepaalde vaardigheidsscore? En welke moeilijkheidsgraad hebben de opgaven in de toets?
 - *Een wetenschappelijke verantwoording:* Hierin staan onder andere de normeringsonderzoeken beschreven. De wetenschappelijke verantwoording wordt samen met het overige toetsmateriaal ter goedkeuring voorgelegd aan de Expertgroep Toetsen PO (zie hoofdstuk 4).

Stap 9: Resultaten monitoren, indien nodig update van normering

Voorafgaand aan de uitgave van een toets heeft een grootschalig normeringsonderzoek plaatsgevonden. Op basis van dit onderzoek is de normering, zoals de indeling in vaardigheidsniveaus I t/m V, vastgesteld. Het is een bekend en onvermijdelijk punt dat de normering 'slijt' met de tijd. Dit slijten van de normering is één van de redenen waarom toetsen vernieuwd worden. Wanneer de normen echter te snel slijten zou een toets zo frequent vernieuwd moeten worden dat het voor scholen te kostbaar wordt. Om deze reden kiezen sommige toetsaanbieders ervoor de behaalde resultaten te monitoren. Deze gegevens zijn bruikbaar om vast te stellen wanneer een update van de normering noodzakelijk is. Een update van de normering is noodzakelijk wanneer een leerkracht op basis van toetsresultaten een te rooskleurig of juist te pessimistisch beeld zou krijgen. Wanneer een leerling bijvoorbeeld een vaardigheidsniveau III haalt, trekt de leerkracht de conclusie dat de leerling hoort bij de gemiddelde leerlingen. Wanneer door het slijten van de normering een III zou betekenen dat een leerling behoort bij de leerlingen met een benedengemiddelde score, oftewel passend bij vaardigheidsniveau IV, zou de leerkracht een verkeerde conclusie trekken. Om ervoor te zorgen dat vernieuwing van toetsen niet al te vaak nodig is én dat de leerkracht de juiste conclusies kan trekken, worden normen indien nodig bijgesteld.

Het lijkt bij bijstellen vaak alsof normen strenger worden. Leerlingen moeten immers een hogere vaardigheidsscore behalen om hetzelfde vaardigheidsniveau te krijgen. Hij bijstellen van de normen betekent echter niet strenger worden, maar betekent weer een juist beeld schetsen.

3 Interpretatie en gebruik van toetsresultaten

3 Interpretatie en gebruik van toetsresultaten

Judith Hollenberg en Marleen van der Lubbe

In de handleidingen van toetsen staat beschreven op welke wijze toetsen afgenomen moeten worden. Hoe meer gestandaardiseerd de toetsafname verloopt, hoe betrouwbaarder de toetsscore is. Het opvolgen van de aanwijzingen uit een handleiding is dus van groot belang. In hoofdstuk 2 kwam aan de orde dat in sommige gevallen voor leerlingen met specifieke beperkingen aanpassingen kunnen worden gedaan om zo op deze wijze ook de vaardigheden van deze leerlingen 'eerlijker' te kunnen meten. Omdat het van belang blijft dat de toets meet wat deze beoogt te meten, is het belangrijk alleen aanpassingen te doen als in de handleiding beschreven staat dat, en vooral bij welke leerlingen, deze zijn toegestaan.

Voor een goed gebruik van toetsen gaat het echter niet alleen om de vraag hoe de toetsen afgenomen moeten worden, maar vooral om hoe de resultaten ingezet kunnen worden als handvat voor het bijstellen van het onderwijsaanbod.

Er zijn verschillende typen toetsscores, waarvan de meest gebruikte in paragraaf 3.1 aan de orde komen. Toetsscores vormen de basis voor verschillende rapportages. Door middel van toetsresultaten kunnen leerkrachten kijken naar het niveau, naar de ontwikkeling en naar specifieke aandachtsgebieden van (een groep) leerlingen. In hoofdstuk 1 zijn al verschillende rapportages uit het Computerprogramma LOVS besproken. In paragraaf 3.2 komen in aanvulling hierop nog enkele voorbeelden van rapportages aan de orde. Waar in hoofdstuk 1 het accent lag op het doel van de verschillende soorten rapportages, ligt in deze paragraaf het accent op het interpreteren en analyseren van de resultaten. In paragraaf 3.3 is er speciale aandacht voor het gebruik van toetsresultaten voor het vaststellen van een geplande vaardigheidsgroei als onderdeel van het ontwikkelingsperspectief en voor het volgen van een leerling op basis van het ontwikkelingsperspectief.

3.1 Soorten toetsscores en normeringen

Als de leerkracht de toets heeft nagekeken, kan hij het aantal goed op de toets bepalen.

Het aantal goede antwoorden op een toets wordt ook wel de ruwe score genoemd. Kan een leerkracht bij 21 goede antwoorden bij de rekentoets tevreden zijn? En welke lesstof sluit aan bij een leerling met 26 goede antwoorden op de toets Spelling E5? Is een leerling vooruit gegaan als hij bij de E3-toets 24 goed had en bij de M4-toets 28 goed?

Het aantal goed bij de ene toets is niet zomaar vergelijkbaar met het aantal goed op een andere toets omdat de toetsen verschillen in inhoud, lengte en moeilijkheidsgraad. Om het aantal goed op verschillende toetsen te kunnen vergelijken, moet het aantal goed omgezet worden naar een score op een onderliggende vaardigheidsschaal. Dit gebeurt bijvoorbeeld bij de LVS-toetsen van Cito, bij intelligentietests en bij de centrale eindtoets.

Om betekenis te geven aan toetsresultaten, worden toetsen veelal genormeerd. Hierbij zijn twee soorten normeringen te onderscheiden:

Bij een *absolute normering* is er een duidelijke grens tussen voldoende versus onvoldoende.

Als bij een tafeldiploma is vastgesteld dat de leerling maximaal 5 fouten mag maken, dan wordt er geen onderscheid gemaakt tussen leerlingen die de toets foutloos hebben gemaakt en

leerlingen die 4 fouten in de toets hebben gemaakt. In beide gevallen krijgen ze hun tafeldiploma.

Bij een *relatieve normering* is er een schaal die aangeeft of de leerling sterk of juist zwak scoort ten opzichte van een normgroep. Bij bijvoorbeeld de LVS-toetsen van Cito wordt onder meer een indeling in niveau I t/m V of A t/m E gebruikt (zie hoofdstuk 1). Een leerling met een A-score is vaardiger dan een leerling met een C-score, maar er is geen sprake van het wel of niet halen van een bepaalde norm (of van 'zakken' of 'slagen').

De meeste volgtoetsen en methodegebonden toetsen kennen een relatieve normering. Bij methodegebonden toetsen geldt dat een leerling met 8 goed het beter heeft gedaan dan een leerling met 5 goede antwoorden. Er zijn bij methodegebonden toetsen echter meestal geen gegevens verzameld die aangeven hoe de score van een leerling zich verhoudt tot de score van vergelijkbare leerlingen, met andere woorden hoe goed de leerling scoort in vergelijking met leeftijdsgenoten. Bij bijvoorbeeld de LVS-toetsen van Cito is dit wel het geval. De normering A t/m E of I t/m V komt tot stand door de score van een grote, landelijke representatieve, groep leerlingen, ook wel vergelijkingsgroep of normgroep genoemd, op de toets te verzamelen (zie hoofdstuk 2).

Een leerkracht neemt toetsen veelal af met als doel te bepalen of een leerling op de goede weg is, of de leerling vooruitgaat en/of om te bepalen hoe een leerling scoort ten opzichte van zijn leeftijdsgenoten. Een leerkracht kan toetsen echter ook gebruiken als houvast voor het onderwijsaanbod, namelijk door de toetsscores inhoudelijk te interpreteren. Wat kan een leerling bij een bepaald aantal goed? En welk onderwijsaanbod past bij de leerling? Naast een *normgerichte interpretatie* is in veel gevallen ook een *inhoudelijke interpretatie* van toetsscores mogelijk. Een toets kan veelal op beide manieren worden geïnterpreteerd. Zo wordt bij een LVS-toets middels een normgerichte interpretatie gekeken hoe een leerling scoort ten opzichte van de totale populatie voor wie de toets bedoeld is, zoals met de scores I t/m V. Daarnaast is ook een inhoudelijke interpretatie mogelijk die betekenis geeft aan een score. Wat kan een leerling bij een bepaalde vaardigheidsscore? Bij de bespreking van de verschillende soorten toetsscores komt waar van toepassing naast de normgerichte interpretatie ook de inhoudelijke interpretatie aan de orde.

3.1.1 Aantal goed en percentage goed

Rapportage van het aantal goed of het percentage goede antwoorden wordt veel gebruikt bij methodegebonden toetsen. Methodegebonden toetsen richten zich op de lesstof die zojuist is aangeboden. Doordat methodes vaak richtlijnen geven over bij welk aantal of percentage goede antwoorden een leerkracht de stof als beheerst kan beschouwen, is er sprake van een soort absolute normering. Hierbij geven de methodes ook veelal adviezen over het onderwijsaanbod, direct gekoppeld aan de toetsresultaten.

Normgerichte en inhoudelijke interpretatie op basis van aantal goed

In figuur 3.1 staat een voorbeeld van de voortgangsregistratie op basis van de methodegebonden toetsen van Wizwijs. In week 3 van elk blok maken de leerlingen in deze rekenmethode een voortgangstoets. Overeenkomstig met het aanbod in de methode krijgt de leerling in deze voortgangstoets meer opgaven rondom Getallen en Bewerkingen en wat minder rondom Meten en Meetkunde. In het figuur staat elk vierkantje voor een vraag. Bij een groene plus is de vraag goed beantwoord, bij een rode min is de vraag fout beantwoord. Per onderdeel staat een beoordeling weergegeven. Ook staat in de tweede kolom van rechts een eindniveau waarin alle twaalf vragen samen zijn meegenomen.

Groepsoverzicht: resultaten en adviezen

Laatste wijziging: 23-07-2009 14:49:24

[Wijzigen](#) [Printen en/of opslaan als pdf](#)

Naam leerling	Toetsdatum	Getallen en Bewerkingen								Niveau	Meten en Meetkunde				Niveau	Eindniveau	Opmerkingen
		1	2	3	4	5	6	7	8		9	10	11	12			
Abdullah Ammi	23-07-2009	+	+	+	+	+	-	+	+	B	+	+	+	+	B	B	
Bertine Boers	23-07-2009	+	-	-	+	+	+	+	+	Z	+	+	+	+	P	Z	
Claudia Cardinale	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	
Dirk van Deurne	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	
Elias van het Einde	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	
Fatima Foudazi	23-07-2009	+	+	+	+	+	+	+	+	P	+	-	+	+	Z	Z	
Geert Groen	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	
Lisa la Lopez	23-07-2009	+	+	+	-	+	+	+	+	B	+	+	+	+	P	B	
Mark Moerevelt	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	
Ronja Reijnders	23-07-2009	+	+	+	+	+	+	+	+	P	+	+	+	+	P	P	



Figuur 3.1: Voorbeeld toetsregistratie Wizwijs
Zie voor een werkwijze van registratie van methodegebonden toetsen op:
www.wizwijs.nl

Maakt een leerling alle opgaven goed, dan staat in het voorbeeld van Wizwijs als eindniveau een 'P' (Plus), oftewel de leerling beheerst de lesstof prima en kan de 'plusopgaven' maken. Maakt een leerling één opgave bij getallen en bewerkingen en/of bij meten en meetkunde fout, dan staat er als eindniveau een 'B' (Basis), oftewel de basisstof op dit gebied sluit prima aan bij de leerling.

Maakt een leerling bij een onderdeel twee of meer fouten, dan verschijnt er een rode stip en de aanduiding 'Z' (zorg), oftewel de leerling beheerst het betreffende onderdeel niet en heeft remediëring nodig alvorens verder te kunnen gaan met de lesstof.

Door de aanduiding van de letters weet de leerkracht welke materialen uit de methode hij de leerling kan aanbieden. Door de weergave van de fouten per opgave, weet de leerkracht bij welk type opgaven extra aandacht wenselijk is.

Bij methodegebonden toetsen is het vaak gebruikelijk om niet alleen het aantal goed te bepalen, maar ook vast te stellen welke opgaven goed en fout worden gemaakt. De toetsen zijn immers beheersingstoetsen en gaan er dus vanuit dat de leerling alle opgaven goed zou moeten kunnen maken. Wanneer de leerling dus uitvalt bij een specifieke opgave, is de kans groot dat extra aandacht voor dit type opgaven nodig is ook al kan de fout natuurlijk altijd een toevalstreffer zijn.

Methodeoverstijgende toetsen zijn geen beheersingstoetsen. Deze toetsen bestrijken een bredere inhoud en bevatten opgaven van uiteenlopende niveaus. Sommige opgaven zijn meer geschikt voor de vaardige leerlingen, van andere opgaven verwachten we dat ook minder vaardige leerlingen deze opgaven goed moeten kunnen maken. Hierdoor is het minder gebruikelijk om per opgave na te gaan of deze goed of fout is gemaakt. Een minder vaardige leerling zal immers naar verwachting de moeilijkere opgaven in de toets fout beantwoorden. Dit betekent echter niet dat aan deze inhoud direct aandacht moet worden besteed. Als een leerling heel laag scoort op een toets, is dit eerder een indicatie dat naar de vaardigheid als geheel aandacht moet uitgaan. Per opgave bekijken welke wel en niet beheerst worden, is in dit geval alleen zinvol als een leerling zowel de makkelijke als de moeilijke opgaven van een bepaalde categorie foutief beantwoordt of indien naar patronen gezocht wordt. Maakt de leerling bijvoorbeeld bij een specifieke categorie opgaven meer fouten dan in een andere categorie?

3.1.2 Werken met vaardigheidsscores

Waar verschillende methodegebonden toetsen los van elkaar staan omdat ze elke keer een kleine leerstofeenheid meten, en dus vaak een gedeelte van een bepaalde vaardigheid toetsen zoals bijvoorbeeld het onderdeel breuken wanneer dit zojuist aan bod is geweest, zijn volgt-toetsen juist aan elkaar gekoppeld. Deze toetsen meten grotere leerstofeenheden en brengen een vaardigheid als geheel in beeld. Om leerlingen te kunnen volgen, moeten resultaten van verschillende toetsen met elkaar vergeleken kunnen worden. Dit is mogelijk door het aantal goede antwoorden op een toets om te zetten naar een vaardigheidsscore op een vaardigheids-schaal.

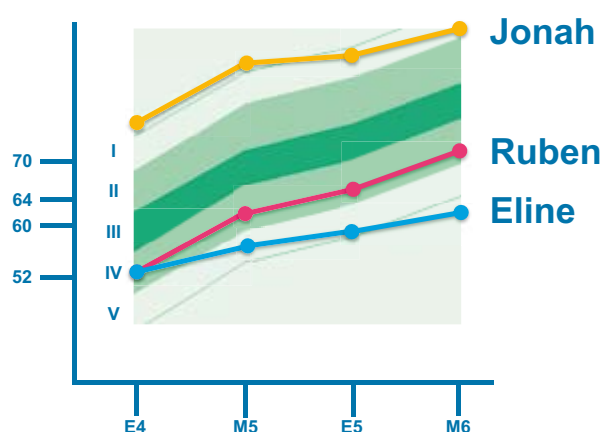
De vaardigheidsschaal maakt het mogelijk om '28 goed' op de ene toets te kunnen vergelijken met '31 goed' op een andere toets. Toetsscores van een leerling op verschillende toets-momenten kunnen dus met elkaar vergeleken worden, net zoals toetsscores van leerlingen die op hetzelfde toetsmoment verschillende toetsen hebben gemaakt, met elkaar vergeleken kunnen worden.

Naast deze mogelijkheden om het niveau en de groei van een leerling vast te stellen, zoals of de leerling vooruit is gegaan en hoe de leerling scoort ten opzichte van andere leerlingen, biedt de vaardigheidsschaal ook mogelijkheden om toetsscores inhoudelijk te interpreteren. Zie hiervoor de uitleg in hoofdstuk 1.

Normgerichte interpretatie op basis van vaardigheidsscores

Op basis van vaardigheidsscores stelt de leerkracht vast of een leerling is vooruitgegaan, maar is deze vooruitgang bevredigend? En hoe functioneert de leerling ten opzichte van zijn leeftijdsgenoten?

Om de eerste vraag of de vooruitgang bevredigend is te beantwoorden, kan gekeken worden naar de gemiddelde groei tussen twee toetsmomenten. In figuur 3.2 is zichtbaar dat Ruben weliswaar beneden het gemiddelde scoort, maar dat zijn groei parallel loopt aan de landelijk gemiddelde groei. De scores van Ruben vallen alle vier de toetsmomenten in vaardigheidsniveau IV, oftewel Ruben hoort bij de zwakste 40% leerlingen. De vaardigheidsscore neemt echter wel toe. Deze gaat immers van een vaardigheidsscore 52 bij E4 naar een score 70 bij M6. Alle punten van Ruben samen vormen een groeicurve. De balk behorend bij vaardigheidsniveau III laat de landelijk gemiddelde groei zien. De groeicurve van Ruben loopt evenwijdig aan de balk die hoort bij niveau III, dit wil zeggen dat zijn groei overeenkomt met de landelijk gemiddelde groei. Hij heeft een achterstand ten opzichte van zijn leeftijdsgenoten, maar deze neemt niet verder toe.



Figuur 3.2: Vaardigheidsgroei van een leerling in vergelijking tot landelijk gemiddelde groei

Om de tweede vraag over hoe de leerling functioneert ten opzichte van leeftijdsgenoten te beantwoorden, kan de vaardigheidsscore omgezet worden naar een normscore I t/m V of A t/m E (zie paragraaf 1.2). Een andere mogelijkheid die, zeker voor leerlingen die afwijken van de reguliere ontwikkeling, aan te bevelen is, is het uitzetten van de vaardigheidsscores in een alternatief leerlingrapport (zie paragraaf 3.2). In het alternatief leerlingrapport staan functioneringsniveaus centraal.

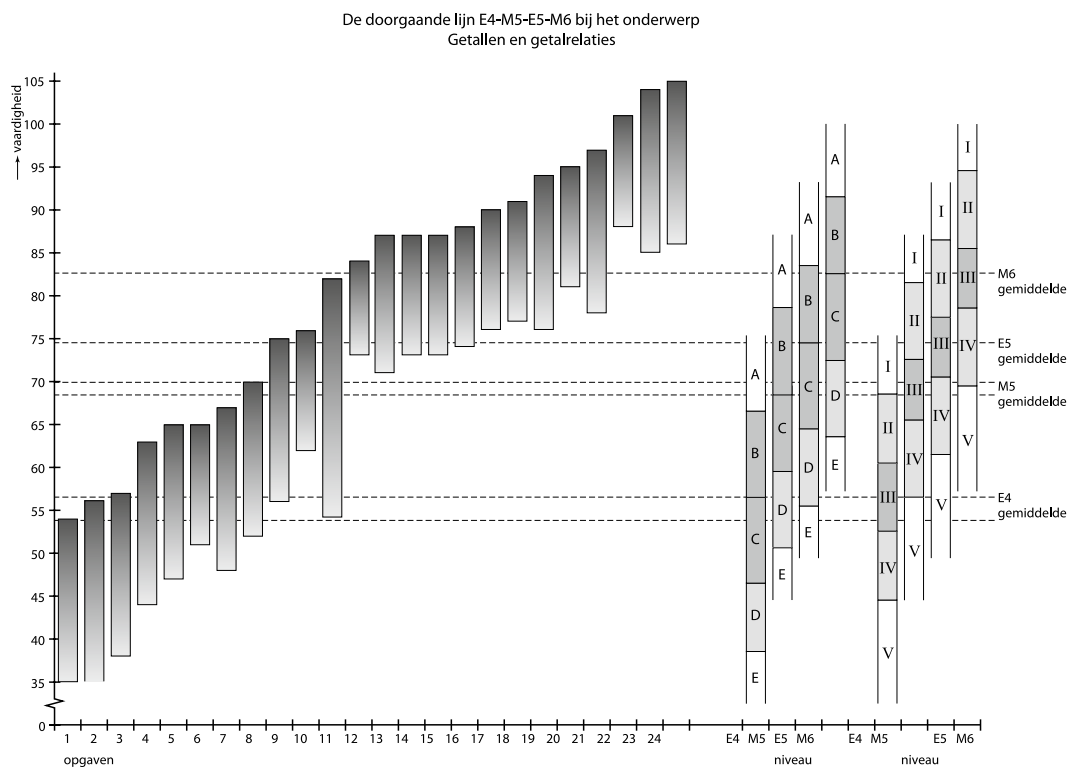
Functioneringsniveaus

Het functioneringsniveau is speciaal ontwikkeld voor leerlingen die werken op hun eigen niveau. Het functioneringsniveau geeft aan op welk niveau een leerling functioneert in vergelijking met leerlingen die het reguliere programma doorlopen.

Een leerling heeft op de LVS-toets Begrijpend lezen M5 een vaardigheidsscore behaald van 8. Deze vaardigheidsscore komt overeen met de vaardigheidsscore van een gemiddelde leerling op het toetsmoment M4. Het functioneringsniveau van de leerling is dus M4. De betekenis hiervan is dat de vaardigheid van de leerling overeenkomt met de vaardigheid van de gemiddelde leerling halverwege groep 4.

Inhoudelijke interpretatie op basis van vaardigheidsscores

Niet alleen leerlingen kunnen geordend worden op de vaardigheidsschaal: hoe hoger de vaardigheidsscore, hoe vaardiger de leerling. Ook bij elke opgave hoort een vaardigheidsscore: hoe hoger de vaardigheidsscore, hoe moeilijker de opgave. Dit maakt het mogelijk om aan de hand van een vaardigheidsscore van een leerling vast te stellen welke opgaven de leerling ongeveer wel en niet beheerst. Hiervoor kan de doorgaande lijn worden gebruikt die in de inhoudsverantwoordingen van de toetsmappen van Spelling en Rekenen-Wiskunde van de LVS-toetsen opgenomen is. Zie voor een voorbeeld figuur 3.3.



Figuur 3.3: Voorbeeld doorgaande lijn Rekenen-Wiskunde uit toetsmap groep 5

Elk balkje representeert een opgave. In figuur 3.3 is zichtbaar dat de bovenkant van het eerste balkje overeenkomt met een vaardigheidsscore 54. Dat wil zeggen dat een leerling met een vaardigheidsscore van 54 ongeveer 80% kans heeft om deze opgave goed te maken. De bovenkant van het achtste balkje zit ongeveer bij vaardigheidsscore 70. Dat betekent dat om opgave 8 goed te maken, dat wil zeggen minimaal een kans van 80% hebben om de opgave goed te beantwoorden, een leerling een vaardigheidsscore van minimaal 70 moet hebben. Om opgave 8 goed te kunnen maken moet een leerling dus een hogere vaardigheidsscore hebben dan om opgave 1 goed te kunnen maken. Dat wil zeggen dat opgave 8 moeilijker is dan opgave 1.

Wanneer een leerling meer dan 80% kans heeft een opgave goed te maken, oftewel het balkje ligt onder de vaardigheidslijn van de leerling, spreken we over beheersing van de opgave. Bij tussen de 50% en 80% kans op het goed beantwoorden van een opgave, is er in termen van Struiksmā/AVI sprake van het instructieniveau. De vaardigheidslijn doorkruist dan het balkje. Bij minder dan 50% kans is er in termen van Struiksmā/AVI sprake van het frustratieniveau. Het balkje ligt dan boven de vaardigheidslijn.

Hoe kan deze inhoudelijke interpretatie van de toetsen Rekenen-Wiskunde worden gebruikt? Het is niet haalbaar en ook niet nodig om voor elke leerling in de klas vast te stellen welke opgaven op beheersings-, instructie- en frustratieniveau liggen.

Het is wel waardevol om voor enkele leerlingen uit de klas de doorgaande lijn te bestuderen.

- Welke opgaven beheerst de leerling met de laagste vaardigheidsscore?
- Welke opgaven beheerst de leerling met een gemiddelde vaardigheidsscore?
- Welke opgaven beheerst de leerling met de hoogste vaardigheidsscore?

Aan de hand van deze informatie kan bijvoorbeeld worden vastgesteld of:

- De niveaus van de verschillende leerlingen voldoende bij elkaar liggen om ze samen instructie te kunnen geven.
- Het niveau van de opgaven op instructieniveau ongeveer overeenkomt met het niveau van de opgaven in de methode.
- Er bepaalde opgaven in de methode zijn waar accenten op gelegd kunnen worden bij zwakkere, gemiddelde of juist sterkere leerlingen.

Let op dat de opgaven behorende bij de doorgaande lijn zijn bedoeld om een indicatie te geven van het niveau van leerlingen met een bepaalde vaardigheidsscore. Het is niet de bedoeling om deze specifieke toetsopgaven te oefenen, maar om de vertaalslag te maken naar opgaven van een vergelijkbaar niveau in de methode.

3.1.3 Andere soorten toetsscores

Naast 'aantal goed' en vaardigheidsscores zijn er vele andere soorten toetsscores. De meest-gebruikte scores worden hier beproven.

Didactisch Leeftijds Equivalent (DLE)

Voor indicatieaanvragen voor bijvoorbeeld LWOO (Leerwegondersteunend onderwijs) en Pro (Praktijkonderwijs) moeten scholen dikwijls DLE's aanleveren. Ook rapporteren veel volg-systemen in DLE's. Zie bijvoorbeeld figuur 3.4 van Dotcomschool.



Figuur 3.4: Rapportage uit Dotcomschool, met als basis DLE's

De afkorting DL staat voor didactische leeftijd. De didactische leeftijd van een leerling komt overeen met het aantal maanden onderwijs dat een leerling vanaf begin groep 3 heeft gevolgd. Daarbij wordt uitgegaan van 10 maanden onderwijs per jaar. Een DL van 20 komt dus overeen met eind groep 4.

DLE staat voor Didactisch Leeftijd Equivalent. DLE zegt hoeveel maanden onderwijs een leerling gemiddeld nodig heeft om tot een bepaald niveau te komen. Ook hierbij tellen de maanden gevolgd onderwijs vanaf groep 3. Een DLE van 20 betekent dus dat de leerling scoort zoals een gemiddelde leerling in juni van groep 4 (10 maanden onderwijs per jaar, startend in september). DLE's worden vaak omgerekend in leerrendement (LR). Wanneer het niveau van een leerling precies past bij de didactische leeftijd van een leerling, is het leerrendement 100%. Met andere woorden: een leerling heeft in januari groep 4 15 maanden onderwijs gehad. Wanneer het niveau van de leerling exact overeenkomt met 15 maanden onderwijs, oftewel DLE = 15, dan is het leerrendement 100%. Ligt de DLE lager dan de didactische leeftijd dan is het leerrendement kleiner dan 100%. Bij een hogere DLE dan DL, is het leerrendement hoger dan 100%.

In figuur 3.4 staat per groep per leerjaar de DLE aangegeven. In groep 3A uit schooljaar 2005/2006 is de didactische leeftijd op het toetsmoment 7. Oftewel de groep heeft 7 maanden onderwijs gehad. Rechts van de balkjes is te zien dat de DLE van deze groep gemiddeld 8 is. Het niveau van deze groep leerlingen komt dus overeen met leerlingen die 8 maanden onderwijs hebben gehad. Oftewel: deze groep leerlingen scoort iets hoger dan verwacht. In de voorlaatste kolom staat de vaardigheidsscore waarop deze berekening gebaseerd is. In de laatste kolom staat het leerrendement (LR). In dit geval staat bij LR 121 wat wil zeggen dat de leerlingen een hoger niveau hebben dan verwacht ($LR = DLE/DL \cdot 100\%$).

Een groot nadeel van het werken met DLE's, is dat deze grotendeels zijn gebaseerd op aannames, zoals dat de ontwikkeling tussen twee toetsmomenten lineair verloopt. Zie ook het artikel Het Drijfzand van DLE's en de flyer DLE hoe zit het daarmee: www.cito.nl > primair onderwijs > Cito volgsysteem primair onderwijs > LVS-toetsen > Toetsen en scores. Het probleem met DLE's is dat toetsen genormeerd worden op vaste momenten, meestal in januari en in juni. Dit betekent dat alleen de DLE's 5, 10, 15, 20, etcetera goed vast te stellen zijn. Deze komen immers overeen met de normeringsmomenten van respectievelijk M3, E3, M4 en E4. Om te bepalen welke vaardigheidsscore bij een DLE van 17 hoort, moet de toets eigenlijk ook in maart aan een grote groep representatieve leerlingen uit groep 4 worden voorgelegd. Bij het ontbreken van deze informatie wordt bij het berekenen van de DLE uitgegaan van een lineaire groei tussen twee normeringsmomenten. Over de groei tussen deze twee momenten is echter weinig bekend. Wel staat vast dat deze allesbehalve lineair verloopt. Dit betekent dus dat de leerling over- of juist onderschat wordt bij het vaststellen van DLE's.

Percentielen

In verschillende rapportages wordt gewerkt met percentielen. Een percentiel van 44 betekent dat de leerling tot de 44% zwakst scorende leerlingen behoort. Oftewel: 44% van de leerlingen heeft dezelfde of een lagere score dan de leerling en 56% van de leerlingen heeft een hogere score dan de leerling.

Percentielen komen onder andere voor in de rapportages van de Centrale Eindtoets en van de Entreetoets. Onderstaande rapportage, figuur 3.5, laat zien dat deze leerling wordt vergeleken met andere leerlingen in groep 7, over het algemeen 30% van de leerlingen hoger scoort en 70% een gelijke of lage score behaalt. Op het gebied van Taal scoort 45% van de leerlingen hoger dan deze leerlingen en 55% gelijk of lager. Bij Studievoordigheden scoort maar 13% van de leerlingen hoger en behaalt 87% een lagere of gelijke score.

	aantal opgaven	aantal goed	percen- tiel	V	IV	III	II	I
Totaal Entreetoets	420	331	70				*	
• Taal	220	163	55			*		
• Rekenen-Wiskunde	120	97	81					*
• Studievoordigheden	80	71	87					*
Taal	220	163	55			*		
• Schrijven	50	37	47			*		
• Grammatica	30	21	48			*		
• Spelling	60	43	57			*		
- Niet-werkwoorden	30	19	32		*			
- Werkwoorden	30	24	75				*	
• Leestekens	25	17	30		*			
• Begrijpend luisteren	50	37	69				*	
• Begrijpend lezen	50	41	69				*	
• Leestempo								
• Woordenschat	60	42	57			*		

Figuur 3.5: Fragment van leerlingprofiel van de Entreetoets

Niveauwaarden

Het leerlingvolgsysteem ParnasSys zet de scores A t/m E om naar niveauwaarden, volgens de procedure zoals weergegeven in figuur 3.6.

Uitleg niveauwaarde:

Niveau omgezet naar waarde tussen 0 en 5

A = 4-5

B = 3-4

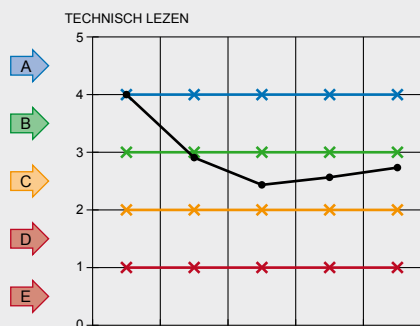
C = 2-3

D = 1-2

E = 0-1

- ✕ Niveau A: niveauwaarde 4 en hoger
- ✕ Niveau B: niveauwaarde 3 en hoger
- ✕ Niveau C: niveauwaarde 2 en hoger
- ✕ Niveau D: niveauwaarde 1 en hoger
- Niveauwaardelij leerling

Gekleurde lijnen zijn de grenslijnen tussen de Cito niveaus



Figuur 3.6: Omzetting A t/m E naar niveauwaarden door ParnasSys

Door deze omzetting zijn scores van leerlingen op verschillende toetsmomenten en verschillende toetsen, dus ook toetsen die rapporteren in termen van DLE en toetsen die rapporteren in vaardigheidsscores, met elkaar te vergelijken. Een leerling met een 'lage B' heeft een niveauwaarde van ongeveer 3. Een leerling met een 'hoge B' zal een niveauwaarde hebben van bijvoorbeeld 3,9 of bijna 4. Heeft de leerling bij Spelling een 'hoge B' en bij Rekenen eveneens een 'hoge B', dan zijn beide scores bijna 4. Heeft een andere leerling bij spelling een 'lage B' en bij rekenen een 'hoge B', dan is er bij enkel bestuderen van de niveauaanduidingen A t/m E geen verschil zichtbaar, terwijl uit de niveauwaarden blijkt dat de leerling bij Rekenen, (niveauwaarde van bijvoorbeeld 3,8), sterker scoort dan bij Spelling, (niveauwaarde is bijvoorbeeld 3,2), ten opzichte van leeftijdsgenoten.

Doordat de vaardigheidsscore is omgezet naar een letter (A t/m E) en vervolgens naar een cijfer (0 t/m 5) is het wel van belang om niet met de cijfers te gaan rekenen. Bij elke omzettingstap gaat immers informatie verloren. Om 'zuivere' berekeningen uit te voeren, moet dus gerekend worden met de vaardigheidsscores.

Gecorrigeerde scores

Bij sommige toetsen, zoals de Centrale Eindtoets, is het mogelijk om gecorrigeerde scores op te vragen. Hierbij wordt de gemiddelde score van de school niet vergeleken met alle scholen, maar met scholen die qua specifieke kenmerken overeenkomen met de leerlingen. Zo is er een correctie voor leerlinggewicht. Hierbij wordt de score van de school vergeleken met de scores van scholen met een vergelijkbaar leerlinggewicht.

Ook bestaan er scores gecorrigeerd voor Begrijpend lezen. De reden hiervoor is dat van alle vaardigheidsscores de score op Begrijpend lezen het sterkste samenhangt met intelligentie. Vandaar dat de score op Begrijpend lezen hier dus gebruikt wordt als maat voor intelligentie.

Door te corrigeren voor de score op Begrijpend lezen, wordt de score van de school dus vergeleken met scholen die leerlingen hebben met een vergelijkbare intelligentie.

Onderwijsresultaten zijn niet alleen afhankelijk van onderwijsaanbod, maar ook van leerlingkenmerken. Door te kijken naar gecorrigeerde scores krijgt men een beeld of de school 'alles uit de leerling haalt wat erin zit'. Van een school met veel leerlingen met een hoge intelligentie mogen dus hogere scores verwacht worden dan wanneer de gemiddelde intelligentie van de leerlingen lager ligt. Overigens geldt dat er ook na de correcties voor leerlinggewicht en Begrijpend lezen natuurlijk nog verschillen tussen scholen bestaan. De correcties zijn weliswaar relevant, maar er zijn meer factoren die invloed hebben op de onderwijsbaarheid van kinderen en de factoren sociaal-economische achtergrond en intelligentie kunnen nog beter gemeten worden dan op basis van leerlinggewicht en scores Begrijpend lezen. De op deze manier voor instroom gecorrigeerde gemiddelden kunnen echter wel aanleiding geven om nog eens goed naar bijvoorbeeld de verdeling van onderwijstijd over de verschillende inhoudsgebieden te kijken.

3.2 Van toetsscore naar handelen in de groep

Een leerkracht toetst niet om te toetsen, maar om een passend onderwijsaanbod voor een leerling te kunnen vaststellen. Hoe komt de leerkracht van een toetsscore naar een handelingsplan? Hiervoor zijn zowel het niveau, de groei als specifieke aandachtsgebieden van belang. In onderstaande paragrafen komen deze drie onderdelen aan de orde.

3.2.1 Kijken naar niveau

Stap 1 bij het kijken naar toetsresultaten, is het bestuderen van het niveau. Hiervoor kan een leerkracht gebruik maken van leerlingrapportages en groepsrapporten (zie voor voorbeelden hoofdstuk 1).

De gegevens op een ingevuld groepsrapport zijn op verschillende manieren te bekijken.

- Allereerst kan een leerkracht de vorderingen van een groep vergelijken met de landelijke groep. In één oogopslag ziet hij hoe de leerlingen uit zijn groep er voor staan in vergelijking met de landelijke vergelijkingsgroep. Belangrijke vragen zijn dan:
 - Hoe verhoudt zich de gemiddelde score van een groep tot die van de landelijke vergelijkingsgroep?
 - Hoe verhoudt de spreiding van de scores van een groep zich tot de spreiding landelijk gezien?
 - Zijn er veel hoog of laag scorende leerlingen en is dat ook overeenkomstig de verwachtingen?
- De resultaten van de leerlingen van een groep zijn ook onderling vergelijkbaar. Belangrijke vragen zijn dan:
 - Van welke leerlingen vallen de resultaten tegen en van welke leerlingen vallen ze mee?
 - Welke leerlingen hebben ongeveer hetzelfde niveau?
 - Hoe groot zijn de verschillen in vaardigheid tussen de leerlingen van de groep?
 - Welke leerlingen hebben onvoldoende van het gegeven onderwijs geprofiteerd?
- Ten slotte kan een leerkracht de vorderingen van een groep ook vergelijken met de onderwijsdoelen die de school wil realiseren. Een belangrijke vraag in dit verband is: Bereiken de leerlingen de gestelde doelen? Om deze vraag te beantwoorden, is het raadzaam het groepsrapport in combinatie met de overzichten in de inhoudsverantwoording bij de toetsen te bekijken. In de inhoudsverantwoording is aangegeven welke doelen getoetst worden en welke opgaven leerlingen van een bepaald vaardigheidsniveau kunnen oplossen.

Al deze informatie is van belang bij het nemen van beslissingen over de voortgang van het onderwijs en kunnen helpen om de leerstof beter af te stemmen op het niveau van de groep of van individuele leerlingen.

Indien een leerkracht toetsresultaten wil interpreteren, moet naast de resultaten van de vergelijkingsgroep ook de context van de eigen situatie in ogenschouw genomen worden. Vragen die dan opkomen zijn bijvoorbeeld: Hoe is de groep samengesteld? Zijn er veel allochtone leerlingen? Zitten er veel zorgleerlingen in de groep? Wanneer uit het groepsrapport blijkt dat het merendeel van een groep beneden het gemiddelde scoort, betekent dat niet automatisch dat aan het gegeven onderwijs de kwalificatie matig of onvoldoende moet worden toegekend. Omgekeerd kan het ook voorkomen dat een school niet tevreden is over de resultaten van een groep die in overeenstemming met de landelijke trend scoort, omdat de groep als geheel meer in haar mars heeft.

De resultaten van een school moeten ook beoordeeld worden tegen de achtergrond van het aanbod. Zo verschillen methoden enigszins van elkaar wat betreft de leerstof en de wijze waarop die wordt aangeboden en de accenten die in een bepaalde periode op onderwerpen worden gelegd. Sommige onderwerpen zijn in de ene methode al diepgaand aan de orde geweest en in andere methoden bijvoorbeeld alleen nog maar geïntroduceerd. Als dat het geval is, moet daar bij de interpretatie rekening mee worden gehouden. De leerkracht kan per leerstofonderdeel globaal nagaan of de betreffende leerstof aan de orde is geweest. Is in de groep geen of weinig aandacht besteed aan bepaalde onderdelen, dan kunnen op dit onderdeel ook minder goede resultaten verwacht worden. Daarmee kan dan rekening worden gehouden bij het interpreteren van de verschillen tussen het niveau van de leerlingen en het niveau van de vergelijkingsgroep.

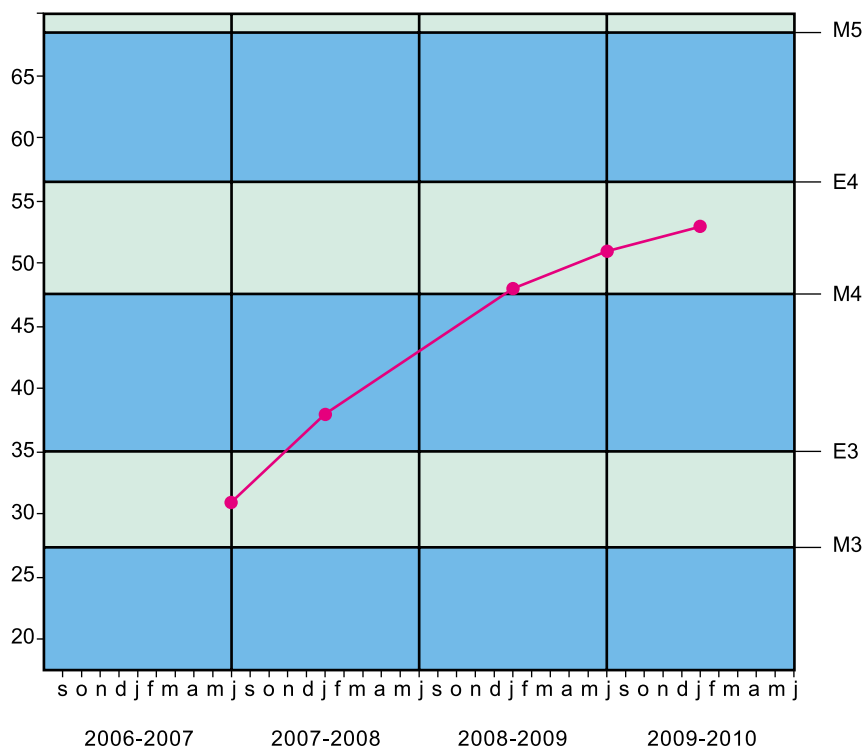
In reguliere rapportages bij de LVS-toetsen van Cito is er in het Computerprogramma LOVS een keuzemogelijkheid voor weergave van de niveaus in A t/m E of I t/m V. In het alternatief leerlingrapport, figuur 3.7, en alternatief groepsprofiel, figuur 3.8, kunnen de

vaardigheidsscores naar keuze omgezet worden naar bovengenoemde typen scores, maar ook naar functioneringsniveaus.

Alternatief leerlingrapport

Leerling: Luna van der Veer

Toets: Rekenen-Wiskunde 2010



Afnamedatum	Leeftijd jr-mnd	Taak	Toetsscore/ Vaardigheidsscore	Functionerings- niveau
26-06-2007	7-08	Toets E3	33/31	M3/E3
13-01-2008	8-03	Toets M4	28/38	E3/M4
18-01-2009	9-03	Toets M4	37/48	M4
11-06-2009	9-08	Toets E4	32/51	M4/E4
21-01-2010	10-03	Toets M5	24/53	M4/E4

Figuur 3.7a: Alternatief leerlingrapport

De horizontale lijnen in het alternatief leerlingrapport, figuur 3.7, markeren de functioneringsniveaus.

Bij Luna die in groep 5 een eigen leerlijn voor Rekenen-Wiskunde volgt, wordt halverwege het schooljaar de toets Rekenen-Wiskunde M4 afgenomen -zie de kolom 'Taak' in de tabel onder de grafiek. Luna haalt een vaardigheidsscore van 48 op deze toets. Deze score valt precies op de lijn behorend bij functioneringsniveau M4. Dit betekent dat Luna op dit moment een niveau heeft dat overeenkomt met dat van een gemiddelde leerling medio groep 4.

Onder de grafiek staan in een tabel de scores van de leerling weergegeven. In de kolom 'Toetsscore/Vaardigheidsscore' staan respectievelijk het aantal goede antwoorden en de vaardigheidsscore. In de laatste kolom staat het 'Functioneringsniveau'.

Indien bij 'niveau codering' gekozen wordt voor A t/m E of I t/m V, is in de laatste kolom deze codering te zien (zie figuur 3.7b). Wanneer de leerling een toets maakt die afwijkt van het toetsmoment dat bij de leeftijdsgroep van de leerling past, bijvoorbeeld een leerling die midden groep 5 de toets M4 maakt, geeft de computer door middel van een pijl het gecorrigeerde niveau weer. Het gecorrigeerde niveau is het niveau dat de leerling zou hebben behaald als zijn vaardigheidsscore wordt vergeleken met de normeringsgegevens van de leeftijdsgroep, in dit geval dus met de normering medio groep 5. Het niveau voor de pijl geeft dus de normscore op het toetsmoment behorend bij de toets, in het voorbeeld dus M4. Het niveau achter de pijl geeft de normscore in vergelijking met de leeftijdsgroep weer, in dit geval M5.

Afnamedatum	Leeftijd jr-mnd	Taak	Toetsscore/ Vaardigheidsscore	Functionerings- niveau
26-06-2007	7-08	Toets E3	33/31	IV → IV
13-01-2008	8-03	Toets M4	28/38	IV → IV
18-01-2009	9-03	Toets M4	37/48	III → V
11-06-2009	9-08	Toets E4	32/51	IV → V
21-01-2010	10-03	Toets M5	24/53	V → V

Figuur 3.7b: Gecorrigeerde scores I t/m V

In het alternatief groepsprofiel, figuur 3.8, ziet een leerkracht, zeker wanneer de resultaten van hoog naar laag gesorteerd worden, in een oogopslag welke leerlingen qua niveau om extra aandacht vragen.

Op de horizontale as staan de vaardigheidsscores. Voor elke leerling is de vaardigheidsscore ingetekend. Hoe hoger de vaardigheidsscore, hoe verder het kruisje naar rechts staat, oftewel hoe langer de lijn vanaf de naam van de leerling. Bij het kruisje staat steeds aangegeven welke toets de leerling heeft gemaakt, bijvoorbeeld E4, welke niveauaanduiding de leerling heeft, bijvoorbeeld D, in welke maand de toets is afgenomen, bijvoorbeeld juni, en tot slot tussen haakjes welke vaardigheidsscore de leerling heeft gehaald, bijvoorbeeld 114.

De verticale lijnen markeren de (landelijk) gemiddelde scores per toetsmoment. Wanneer een kruisje bijvoorbeeld op de verticale lijn staat waarboven 'E3' staat, betekent dit dat het niveau van de leerling overeenkomt met dat van de gemiddelde leerling eind groep 3. Staat het kruisje tussen de verticale lijn M5 en E5? Dan is het niveau van de leerling vergelijkbaar met het niveau van een leerling tussen midden en eind groep 5.

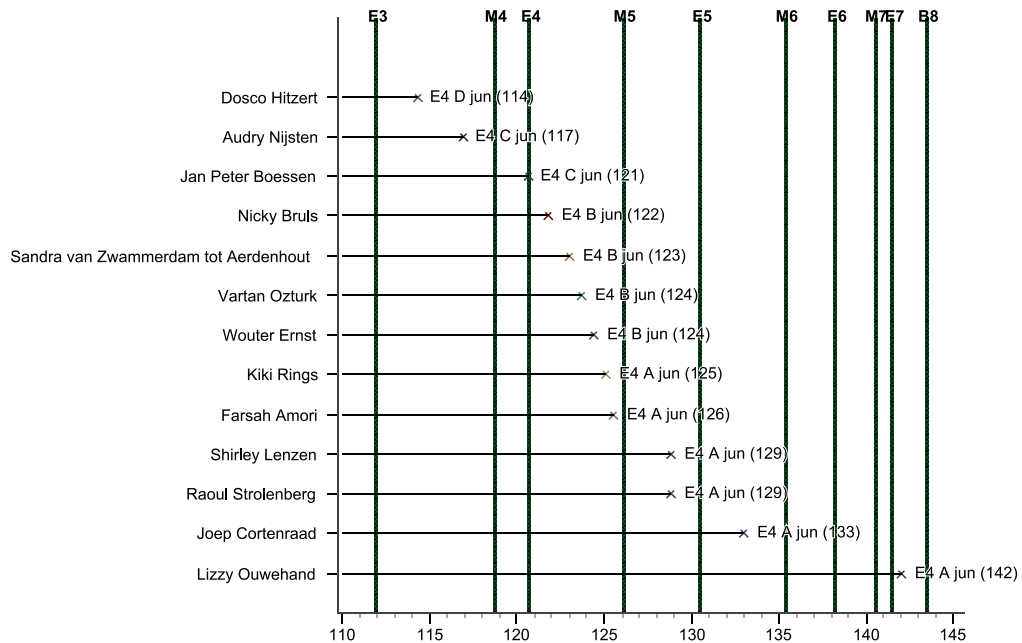
Figuur 3.8 laat zien dat Dosco en Audrey op de toets Spelling E4 onder het niveau van de gemiddelde leerling halverwege groep 4 functioneren. Zeker als Dosco en Audrey ook nog stagneren of terugvallen in hun ontwikkeling, is het dan zaak dat er een nadere analyse van het resultaat wordt gemaakt. Shirley, Raoul, Joep en Lizzy scoren relatief ver boven het niveau van de gemiddelde leerling eind groep 4 en hebben misschien wel extra uitdaging nodig.

Alternatief groepsprofiel

Groep: 4 - 4A (13 lln)

Vergelijking: Alle leerlingen

Toets: Spelling 2010 - 2009-2010 Eind (Apr-Jul, gem. score 125,2)



Figuur 3.8: Alternatief groepsprofiel

Welke informatie geeft dit voor het handelen?

Door te kijken naar het functioneringsniveau beschikt de leerkracht over een indicatie van het beheersingsniveau van de groep en van individuele leerlingen.

Op groepsniveau: Door te kijken naar het niveau van de groep als geheel kan worden vastgesteld op welk niveau de groepsinstructie kan worden gegeven. Met behulp van het alternatief groepsprofiel kunnen bijvoorbeeld eenvoudig niveaugroepjes in een klas worden samengesteld. Bijvoorbeeld, alle leerlingen met een score beneden de fictieve lijn E3 vormen een niveaugroepje.

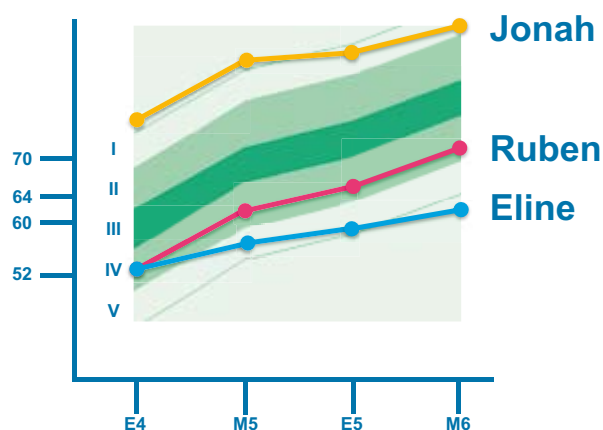
Op leerlingniveau: Door te kijken naar het niveau per leerling, kan ingespeeld worden op individuele verschillen. Leerlingen waarbij het niveau sterk afwijkt van het niveau van de groep, vragen om extra aandacht voor het betreffende leergebied. Bijvoorbeeld een (zeer) intensief arrangement, conform het Onderwijscontinuüm zoals besproken in hoofdstuk 1. Door te kijken naar het functioneringsniveau beschikt de leerkracht over handvatten over op welk niveau de leerling vlot instructie tot zich kan nemen.

Zie voor meer informatie over het werken met het alternatief leerlingrapport en het alternatief groepsprofiel de kwaliteitskaart 'Werken met functioneringsniveaus' op www.schoolaanzet.nl > kwaliteitskaarten.

3.2.2 Kijken naar groei

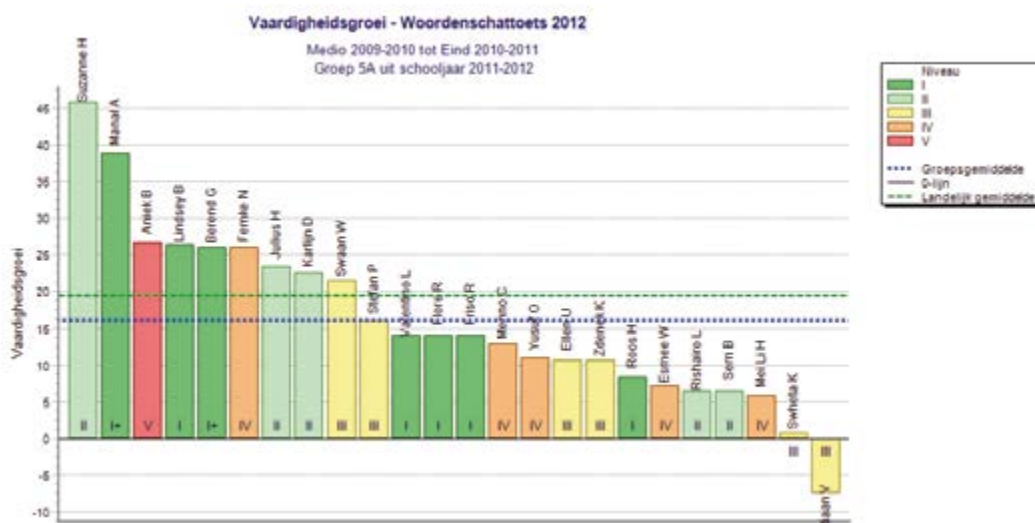
Figuur 3.9 laat zien dat Eline op het moment E4 een IV-score heeft. Op het volgende toetsmoment en de toetsmomenten daarna behaalt ze vaardigheidsniveau V. Op toetsmoment M6 behoort zij zelfs tot de 10% leerlingen met de laagste scores. Door te kijken naar vaardigheids-

scores is al tijdig zichtbaar dat Annet minder groei doormaakt dan gemiddeld en dat zij dus afzakt naar lagere vaardigheidsniveaus. Alleen kijken naar de aanduidingen I t/m V en A t/m E heeft als gevaar dat een minder gemiddelde groei pas gesignaleerd wordt als de leerling een andere niveauaanduiding krijgt. Wanneer gekeken wordt naar vaardigheidsscores is de verminderde groei zichtbaar.



Figuur 3.9: Leerlingrapport van 3 leerlingen

Kijken naar groei kan met de leerlingrapportages, bijvoorbeeld door te kijken naar 'lage I's' versus 'hoge I's'. Een handigere manier is door gebruik te maken van de rapportage vaardigheidsgroei in het Computerprogramma LOVS van Cito (figuur 3.10).



Figuur 3.10: Rapportage vaardigheidsgroei

De rapportage kan tussen twee verschillende toetsmomenten naar keuze worden opgevraagd. In het voorbeeld wordt gekeken hoeveel vaardigheidspunten de leerlingen tussen het M-moment en het E-moment in groep 4 gegroeid zijn voor Rekenen-Wiskunde. Per moment en per inhoudsgebied is de te verwachten groei verschillend. In dit voorbeeld is de landelijk gemiddelde groei tussen deze twee momenten 9 punten. Dit wordt weergegeven door de gestreepte lijn. De gestippelde lijn geeft de gemiddelde groei van de groep weer. De gestippelde lijn ligt in dit geval boven de gestreepte lijn, oftewel de groei van de groep is hoger dan de groei van het landelijk gemiddelde.

Na het bekijken van de groei op groepsniveau door de gestreepte lijn te vergelijken met de gestippelde lijn, kan de groei op leerlingniveau worden bekeken. Elk balkje in de figuur is een leerling. Hoe hoger het balkje, hoe groter de groei van de leerling. Komt het balkje boven de gestreepte lijn uit, dan betekent dit dat de groei van de leerling groter is dan het landelijk gemiddelde. Komt het balkje boven de gestippelde lijn uit, dan is de groei van de leerling groter dan het groepsgemiddelde.

Door te kijken naar welke leerlingen onder de gestreepte en/of gestippelde lijn scoren, heeft de leerkracht een snel beeld welke leerlingen een benedengemiddelde groei hebben doorgemaakt. Let op: een benedengemiddelde groei betekent niet automatisch een benedengemiddeld niveau en vice versa. Met de kleurcoderingen en de vaardigheidsniveaus in de balkjes ziet u snel of de leerling met een benedengemiddelde groei ook een benedengemiddeld niveau geeft. In figuur 3.10 is zichtbaar dat Roos, met vaardigheidsniveau I, een groei doormaakt die beneden het landelijk en groepsgemiddelde ligt. Aniek, met vaardigheidsniveau V, maakt met haar benedengemiddelde niveau daarentegen juist een bovengemiddelde groei door.

Door op het balkje te klikken verschijnen de toetsscores van de leerling van de afgelopen toetsmomenten in beeld en krijgt u dus meer informatie over het verloop in de vaardigheidsniveaus.

Welke informatie geeft dit voor het handelen?

Op groepsniveau: Indien de gemiddelde groei van de groep lager ligt dan de landelijk gemiddelde groei, weet een school dat extra aandacht voor dit vakgebied wenselijk is.

Op leerlingniveau: Bij leerlingen die een beperkte groei doormaken, is nadere analyse noodzakelijk om vast te stellen of de leerling extra instructie en/of oefening op zijn niveau nodig heeft om verder te komen. Bij leerlingen met een (boven)gemiddelde groei weet de leerkracht dat de huidige aanpak werkt en dus voortgezet kan worden.

3.2.3 Kijken naar specifieke aandachtsgebieden

Een leerling kan een gemiddeld niveau hebben, een gemiddelde groei doormaken, en toch opvallen. Dit kan doordat de leerling op de vaardigheid als geheel (boven)gemiddeld scoort, maar op onderdelen zwakker scoort. Het is van belang dit tijdig te signaleren omdat dergelijke disharmonische resultaten vaak voorafgaan aan een beperktere groei.

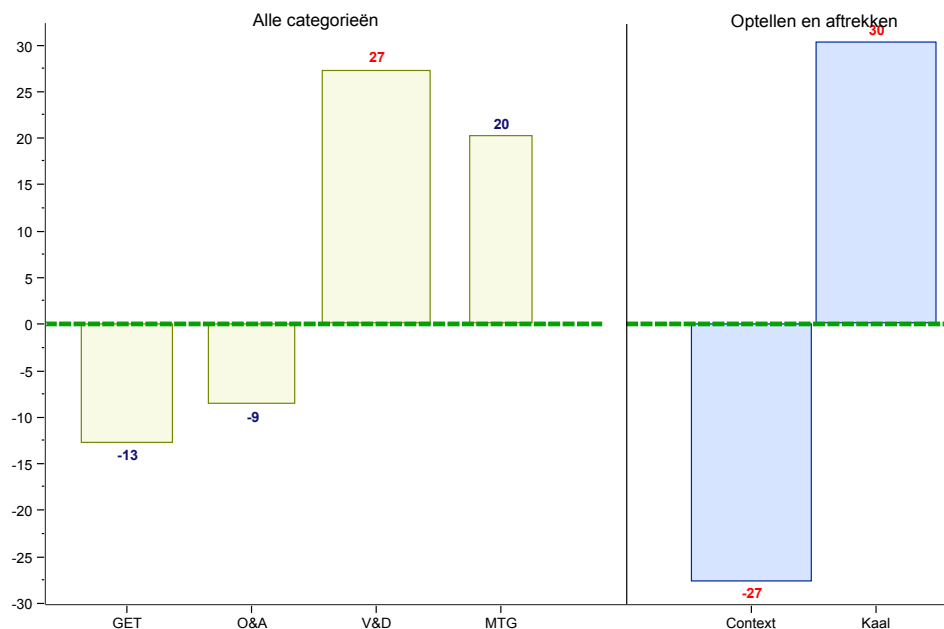
Om vast te stellen of de score van de leerling op de verschillende onderdelen van een leergebied ongeveer overeenkomt, kunnen verschillende analyses uitgevoerd worden. Hier zullen twee soorten analyses besproken worden: een foutenanalyse bij de LVS-toets Spelling en een categorieënanalyse bij de LVS-toets Rekenen-Wiskunde. Een categorieënanalyse is ook mogelijk bij andere toetsen zoals de toetsen Rekenen voor kleuters en Taal voor kleuters.

Categorieënanalyse Rekenen-Wiskunde

Als de gegevens van de LVS-toets Rekenen-Wiskunde zijn ingevoerd in het Computerprogramma LOVS, kan een categorieënanalyse worden gemaakt. De categorieënanalyse vergelijkt de scores van een leerling op verschillende deelgebieden met de totaalscore van een leerling. Op basis van de totale vaardigheidsscore voorspelt het computerprogramma hoeveel opgaven een leerling per categorie goed zou moeten maken, de verwachte score per categorie. Dit wordt vergeleken met de daadwerkelijk behaalde score, de geobserveerde score per categorie. Bij de meeste leerlingen komt de daadwerkelijk behaalde score behoorlijk overeen met de verwachte score. Sommige leerlingen hebben echter bij een specifieke categorie meer opgaven goed of fout dan verwacht. Zo kan een leerling relatief eenvoudige opgaven over optellen en aftrekken uit de toets fout maken, terwijl de leerling relatief moeilijkere opgaven rondom getallen en getalrelaties goed beantwoordt. In figuur 3.11 is een voorbeeld opgenomen van een categorieënanalyse uit de toets Rekenen-Wiskunde 3.0.

Categorieënanalyse leerling

Leerling: **Rick Asgaraly** (3 - 3A)
 Toets - taak: **Rekenen-Wiskunde 3.0 - E3-digi**



Afnamedatum: 25-06-2014

Score / Vaard. Niv.: 30 / 128 **IV**

Alle categorieën: 0 2 Niet opvallend

Optellen en aftrekken: 0 2 Zeer opvallend

Categorie label	Categorie	Aantal opgaven	% Score geobserveerd	% Score verwacht	% Score afwijking
GET	Getallen en getalrelaties	16	48	61	-13
O&A	Optellen en aftrekken	20	49	58	-9
V&D	Vermenigvuldigen en delen	8	75	48	+27
MTG	Metten, tijd en geld	8	75	55	+20
Context	Context	10	20	47	-27
Kaal	Kaal	10	81	51	+30

Figuur 3.11: Voorbeeld van een categorieënanalyse Rekenen-Wiskunde

Bij de toetsen die voor 2012 zijn uitgegeven, is een analyse van de rekendomeinen mogelijk. Bij toetsen Rekenen-Wiskunde 3.0 groep 3 t/m groep 6 is naast een analyse van rekendomeinen ook de mogelijkheid om verschillen in prestaties tussen kale en contextopgaven in kaart te brengen.

Een balkje naar beneden betekent dat Rick bij dat betreffende onderdeel minder opgaven goed maakte dan verwacht werd op basis van zijn totale score op de toets. In het voorbeeld was de verwachting dat hij 47% van de contextopgaven goed zou maken, zie kolom '% Score verwacht'. Hij heeft echter maar 20% van deze vragen goed beantwoord, zie kolom '% Score geobserveerd'. Bij de kale opgaven heeft hij daarentegen juist meer opgaven goed dan verwacht.

Onder de grafiek staat of het profiel niet opvallend of (zeer) opvallend is. Het profiel van Rick is op het gebied van kaal-context zeer opvallend. Dit betekent dat de scores bij kale en context-opgaven sterk afwijken van wat de verwachting is. Er is sprake van een significant, dat wil zeggen een niet door toeval veroorzaakt verschil, en nadere actie is gewenst. Bijvoorbeeld door het aanbieden van hulp op maat na het verder in kaart brengen van de problematiek aan de hand van een diagnostisch gesprek. Rick scoort bij de rekendomeinen niet opvallend. Hij scoort op de verschillende onderdelen weliswaar enigszins anders dan verwacht werd op basis van zijn totale vaardigheidsscore, maar het gaat niet om significante verschillen. Er zijn met andere woorden geen specifieke rekendomeinen die Rick meer of minder beheerst dan andere. Kortom, alle inhouden vragen ongeveer om evenveel aandacht bij het eventueel aanbieden van extra ondersteuning. De extra ondersteuning kan zich wel sterker richten op contextopgaven dan op kale opgaven.

Zoals in het linkerdeel van figuur 3.11 zichtbaar is, is ook bij een niet opvallend profiel sprake van balkjes naar boven en naar beneden. Geen enkele leerling scoort immers precies zoals verwacht. De computer berekent of de afwijking van de verwachting significant afwijkend is. In een klas van 20 leerlingen hebben gemiddeld 2 tot 4 leerlingen een (zeer) opvallend profiel. Bij het bepalen of er sprake is van een significant verschil spelen allerlei factoren mee, onder andere het aantal opgaven in een categorie, maar ook statistische gegevens zoals de mate waarin een opgave onderscheid maakt tussen vaardige en minder vaardige leerlingen. Hoe dikker het balkje, hoe sterker deze categorie meetelt bij het bepalen of de leerling een (zeer) opvallend profiel heeft. Bij de profielen staat bij het signaal een staafje. Hoe verder het staafje is ingekleurd, hoe dichter het profiel van een leerling tegen een opvallend profiel aan ligt. Zowel leerlingen met een hoge als met een lage vaardigheidsscore kunnen een opvallend profiel hebben. Het is daarom aan te bevelen om voor alle leerlingen een categorieënanalyse uit te voeren.

Met de informatie uit de categorieënanalyse kan de leerkracht vaststellen of het nodig is aan specifieke inhouden extra aandacht te besteden. Het is belangrijk om te realiseren dat wanneer een balkje naar boven staat, dit niet betekent dat de leerling per definitie zeer vaardig is in het betreffende onderdeel. Een balkje naar boven betekent alleen dat de leerling op het betreffende onderdeel meer opgaven goed heeft gemaakt dan werd verwacht op basis van zijn vaardigheidsscore op de totale toets. Ook minder vaardige leerlingen kunnen dus balkjes naar boven hebben, net zoals meer vaardige leerlingen balkjes naar beneden kunnen hebben.

De categorieënanalyse kan ook op groepsniveau gemaakt worden. In een overzicht is dan snel zichtbaar hoeveel leerlingen bij elke categorie hoger of lager scoren dan verwacht en wanneer dit opvallend is. Zo ziet een leerkracht in een oogopslag aan welke categorie(ën) op groepsniveau aandacht besteed kan worden.

Foutenanalyse Spelling

Wanneer de leerkracht na afname van de LVS-toets Spelling meer inzicht wil in het soort fouten die de zwakke leerlingen maakten, kan voor die leerlingen een foutenanalyse worden gemaakt. Zie voor een voorbeeld figuur 3.12. Natuurlijk kan er ook voor gekozen worden de toetsen van de hele groep nader te analyseren. Dit is met name aan te bevelen wanneer het vermoeden bestaat dat sommige spellingcategorieën voor meer leerlingen in een groep lastig zijn.

Foutenanalyse

Groep: **5 - 5A**Toets taak: **Spelling 2010 - E3 start + 1**

	Categorie	Cat 1		Cat 2		Cat 3		Cat 4		Cat 5		Cat 6		Cat 7		Tot%	
		c	a	c	a	c	a	c	a	c	a	c	a	c	a	c	a
Jonny Hummel		40	0	0	0	50	0	0	13	0	0	0	17	25	25	18	8
Gemiddelde% (1 lln)		40	0	0	0	50	0	0	13	0	0	0	17	25	25	18	8

Categorie 1: mkm-woorden (5 opgaven)

Categorie 2: mmkm-woorden en mkmm-woorden (6 opgaven)

Categorie 3: mmkmm-woorden (10 opgaven)

Categorie 4: éénlettergrepige woorden met niet geschreven tussenklank (8 opgaven)

Categorie 5: woorden met meer dan twee medeklinkers na elkaar (7 opgaven)

Categorie 6: woorden met sch- of schr- (6 opgaven)

Categorie 7: woorden met -ng of -nk (8 opgaven)

c = categorie fout (percentage)

a = andere fout (percentage)

Figuur 3.12: Voorbeeld van een foutenanalyse Spelling

In de foutenanalyse staat per categorie het percentage foute antwoorden. Als 50% of meer van de woorden uit een categorie fout zijn gespeld, kleurt het getal rood. Dit is een signaal dat de leerling deze categorie in elk geval nog onvoldoende beheerst.

Maak bij het uitvoeren van een foutenanalyse altijd een vergelijking met het onderwijsaanbod. Bij het ontwikkelen van de toetsen houdt Cito zoveel mogelijk rekening met de volgorde waarin verschillende categorieën in de methodes aan bod komen. Desondanks kan het voorkomen dat een categorie op een moment al wel in de toets zit, maar nog niet aan bod is geweest in de methode. In deze gevallen hoeft de leerkracht zich, totdat de categorie met de leerling(en) is geoefend, nog geen zorgen te maken.

Bij het maken van de foutenanalyse maken we onderscheid tussen categoriefouten (de kolom met de 'c' erboven) en andere fouten (de kolom met de a erboven). Bij een categoriefout zit de fout van de leerling in de categorie waarvan het woord in de toets opgenomen is. Wanneer bijvoorbeeld 'flipperkast' wordt opgenomen om te bepalen of de leerling de regel van de gesloten lettergreep beheerst, wordt het schrijven van 'flipperkast' gerekend als categoriefout. De leerling kan echter ook een andere fout maken, denk aan 'vlipperkast'. Deze fout telt dan mee als andere fout. Als uit de foutenanalyse blijkt dat een leerling vooral moeite heeft met de schrijfwijze van woorden uit één of enkele specifieke spellingcategorieën, is het aan te bevelen om de daarbij behorende controledictees af te nemen. Deze controledictees zijn te vinden in de Hulpboeken Spelling bij de LVS-toetsen van Cito. Deze dictees zijn bedoeld om na te gaan of er wel echt sprake is van een probleem in de betreffende categorie. Nadat met de controledictees de categorieën zijn vastgesteld die voor een leerling moeilijkheden opleveren, stelt de leerkracht een hulplan op. Daarvoor kan gebruik worden gemaakt van de didactische aanwijzingen, oefenvormen en verwijzingen in het Hulpboek. Vervolgens voert de leerkracht het plan uit. Na afloop kan de leerkracht met de betreffende controledictees nagaan of de genomen maatregelen effect hebben gehad.

Welke informatie geeft dit voor het handelen?

Op groepsniveau kan de leerkracht kijken of er categorieën zijn waarop de hele groep, of in elk geval een groot deel van de leerlingen, uitvalt. In het geval van Spelling controleert de leerkracht in dit geval of deze categorieën al in de methode aan bod zijn geweest. Zo niet, dan volgt de leerkracht de eigen methode en houdt de categorie bij een volgende toetsafname goed in de gaten. Zo ja, dan vraagt dit om herhaling van de betreffende categorie op groepsniveau.

In het geval van Rekenen-Wiskunde kan de leerkracht, indien veel leerlingen op dezelfde categorie uitvallen, hier extra nadruk op leggen bij de instructie.

Ook op leerlingniveau kan de leerkracht zo bepalen welke categorieën om extra aandacht vragen. De leerlingen kunnen zo gerichte instructie en materialen aangeboden krijgen.

Aanvullende informatie verzamelen

Om het onderwijsaanbod af te stemmen op de specifieke behoeften van leerlingen, kunnen de gegevens over niveau, groei en specifieke aandachtsgebieden aangevuld worden met bevindingen uit diagnostische toetsen en/of een diagnostisch gesprek. Zo geeft een opvallend profiel bij Rekenen van een leerling met niveau M5, een iets benedengemiddelde groei en uitval op getallen en bewerkingen, aan dat extra instructie en oefening op het gebied van getallen en bewerkingen op ongeveer het niveau M5 raadzaam is. Hoe deze instructie het best aangeboden kan worden, kan via een diagnostisch gesprek worden nagegaan. Heeft deze leerling bijvoorbeeld de structuur in eenheden, tientallen en honderdtallen door? En kan de leerling gebruik maken van dubbeln? Een diagnostisch gesprek levert dergelijke specifieke informatie waardoor het onderwijs nog gericht op de leerling kan worden afgestemd. Dit kan met materiaal uit de methode, maar kan ook aan de hand van de handreikingen in de hulpboeken bij de toetsen. Hierin staan per deelgebied suggesties over hoe hier meer informatie over te verzamelen is. Bij het verzamelen van aanvullende informatie is het van belang om verder te kijken dan alleen naar het toetsresultaat. Alle waarnemingen uit de klas, aangevuld met diagnostische informatie, kunnen samen met de toetsresultaten geïntegreerd worden om tot een goed beeld van de leerling te komen.

Er zijn, met name voor diagnosticeren van rekenproblemen, speciale uitgaven op de markt om resultaten van leerlingen verder te analyseren. Denk aan diagnostische gesprekken van Maatwerk of aan de map Diagnosticeren en plannen van Cito. Maatwerk is gericht op cruciale kennis, zoals weet een leerling dat 50% de helft is? Diagnosticeren en plannen is gericht op de verschillende strategieën die een leerling inzet bij het oplossen van een rekenopgave. Telt de leerling op z'n vingers, heeft de leerling de structuur in tientallen en eenheden door?

3.3 Toetsscores als basis voor ontwikkelingsperspectief

Voor leerlingen die, bijvoorbeeld door specifieke onderwijsbehoeften en/of een lage intelligentie, niet met het reguliere programma mee kunnen doen, is vergelijken met het landelijk gemiddelde veelal niet reëel en weinig zinvol. Om het optimale uit de leerlingen te halen, is het echter wel van belang om doelen te stellen om met deze leerlingen naar toe te werken. Om deze reden is het goed om voor leerlingen die voor één of meer vakken op een eigen niveau werken een ontwikkelingsperspectief op papier te zetten.

Scholen voor (voortgezet) speciaal onderwijs moesten al vanaf augustus 2013 voor alle leerlingen een ontwikkelingsperspectief vaststellen. Vanaf augustus 2014 is een ontwikkelingsperspectief ook verplicht voor leerlingen in het praktijkonderwijs, in het speciaal basisonderwijs en leerlingen uit het reguliere onderwijs die extra ondersteuning vanuit het samenwerkingsverband ontvangen.

Vaststellen van een ontwikkelingsperspectief

Het ontwikkelingsperspectief bestaat onder andere uit een geplande vaardigheidsgroei (ook wel leerrendementsverwachting genoemd) en een uitstroombestemming. Een geplande vaardigheidsgroei wordt per vak of vakoverstijgend domein opgesteld. Per gebied wordt gekeken welk niveau de leerling naar verwachting aan het einde van het basisonderwijs zal halen. Een uitstroombestemming is een gewogen besluit waarbij de geplande vaardigheidsgroei van verschillende leergebieden in relatie tot kenmerken van de leerling worden bekeken. Behaalde toetsresultaten vormen belangrijke input voor het opstellen van de geplande vaardigheidsgroei. Toetsresultaten worden beïnvloed door leerlingkenmerken (zoals intelligentie en aanwezigheid van stoornissen), door omgevingskenmerken (zoals hoog-opgeleide ouders, thuis veel voorlezen, niet-Nederlands sprekende ouders) en schoolkenmerken (zoals veel individuele tijd per leerling, rustig klaslokaal). Deze factoren zijn los van elkaar moeilijk meetbaar te maken en dus in een model te vatten. Door naar toetsresultaten te kijken worden al deze factoren indirect meegewogen. Op basis van minimaal drie toetsmomenten kan een school plannen waar een leerling het volgende toetsmoment, en eventueel het toetsmoment erna, uit kan komen.

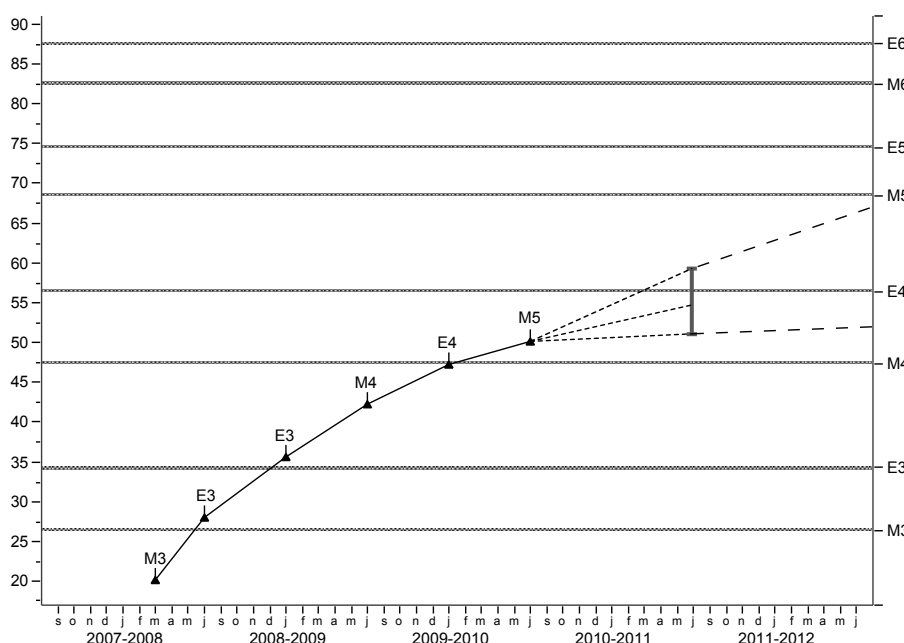
Uitsluitend Test - ONGEREGISTREERD voor schoolgebruik! Arnhem



Alternatief leerlingrapport

Leerling: **Kristi Haynes (7 - 7A)**
Toets: **Rek-Wisk 2012**

Vergelijking: **Alle leerlingen**



Figuur 3.13: Model van ontwikkelingsperspectief op basis van toetsresultaten

Verschillende administratiesystemen bieden de mogelijkheid om een geplande vaardigheidsgroei, ook wel leerrendementsverwachting genoemd, uit te draaien. De berekeningen hierachter verschillen per administratiepakket. Zo zijn er systemen die zich baseren op DLE's waarbij het beginniveau van de leerling niet wordt meegenomen. Het is daarom raadzaam de geplande vaardigheidsgroei altijd te leggen naast de eigen verwachtingen van de leerlingen. Niets is zo vervelend als dat wanneer een leerling over- of onderschat wordt en daardoor het verkeerde aanbod krijgt.

Welke methode voor voorspellingen ook gehanteerd wordt, het is van belang de planning van de vaardigheidsgroei altijd als een marge te zien. Elke meting, maar zeker ook elke voorspelling/planning heeft immers een onnauwkeurigheid. In figuur 3.13 staat een voorbeeld van hoe de geplande vaardigheidsgroei er in de toekomst in het Computerprogramma LOVS ongeveer uit zal zien. De behaalde scores van de leerling zijn verbonden door een lijn. Bij de voorspelling van de scores is een marge weergegeven. In de figuur is zichtbaar dat hoe verder de voorspelling/planning is, hoe groter de marge. Het systeem 'voorspelt' maximaal twee toetsmomenten vooruit. Verder vooruit voorspellen levert immers een te grote marge op en zegt dus weinig meer. De voorspelling is gebaseerd op een grote hoeveelheid data van leerlingen. Hierbij is gekeken naar patronen in het scoreverloop van leerlingen. Welke scores behalen bijvoorbeeld leerlingen op het toetsmoment E4 die op M3 een score 53 halen, op E3 een vaardigheidsscore 60 en op M4 een vaardigheidsscore 68? En welke scores behalen leerlingen E4 bij een scoreverloop van 45 naar 50 naar 60? Deze manier van voorspellen van de scores bleek uit het landelijke project 'Leerwinst en toegevoegde waarde' de meest betrouwbare resultaten op te leveren en dus een goede manier om te kunnen plannen.

Zoals gezegd zijn computermodellen een hulpmiddel, en moeten deze altijd gebruikt worden in combinatie met de kennis van de leerkracht over de leerling en de mogelijkheden van het onderwijs. Bedenk bij het plannen waar u met een leerling naartoe wilt dat hoge verwachtingen samengaan met hoge resultaten. Stel doelen dus ambitieus maar realistisch.

Enkele misverstanden bij het opstellen van een ontwikkelingsperspectief

Voor het bepalen van een ontwikkelingsperspectief waren tot voor kort nog weinig hulpmiddelen beschikbaar. Hieronder staan enkele misverstanden uit de praktijk bij het vaststellen van een ontwikkelingsperspectief. Deze misverstanden maken duidelijk dat bij het vaststellen van een ontwikkelingsperspectief zowel naar het huidige niveau, als naar de groei in de afgelopen jaren en naar de groei van de gemiddelde leerling gekeken moet worden.

Alleen kijken naar het huidige niveau: In sommige modellen voor het vaststellen van ontwikkelingsperspectief wordt alleen het huidige niveau als uitgangspunt genomen bij het bepalen van het ontwikkelingsperspectief. De groei in de afgelopen jaren wordt in de voorspelling niet meegenomen. Heeft de leerling nu een score op II-niveau? Dan gaat men er bij het bepalen van het ontwikkelingsperspectief soms vanuit dat de leerling een score op dit niveau zal behouden. Hierbij wordt dus niet gekeken naar groei: Een leerling die eerder op I-niveau scoorde en nu op niveau II scoort, zal immers een ander perspectief hebben dan een leerling die voorheen op III-niveau scoorde en nu een score op II-niveau behaalt.

De vaardigheidsgroei in een rechte lijn doortrekken: Om te bepalen wat de vaardigheidsscore van een leerling aan het einde van het PO is, wordt soms uitgegaan van een lineaire groei in vaardigheidsscore. Is de leerling de afgelopen 2 jaar gemiddeld 8 punten op de vaardigheidsschaal gegroeid? Dan gaan sommige modellen er vanuit dat de leerling ook de jaren tot aan het einde van het PO 8 punten per jaar zal groeien. Dit is geen reële verwachting aangezien de groei van de gemiddelde leerling niet lineair verloopt. Bij veel leergebieden maakt de leerling in de lagere groepen een forse groei door en neemt deze groei in de bovenbouw, wanneer inzicht een grotere rol gaat spelen, enigszins af. Dit is zichtbaar in de groeicurves op de leerlingrapporten van de LVS-toetsen. Modellen die uitgaan van een lineaire groei leiden dus vaak tot een te ambitieus ontwikkelingsperspectief.

Volgen op basis van ontwikkelingsperspectief

Wanneer de geplande vaardigheids groei is vastgesteld, kan bij een volgend toetsmoment worden gekeken in hoeverre de toetsresultaten passen binnen het gestelde perspectief. Indien op een toetsmoment de toetsresultaten lager uitvallen dan op basis van de geplande vaardigheids groei werd verwacht, kan de school verschillende opties nagaan:

- Past het gegeven onderwijsaanbod bij de leerling? Indien met bijvoorbeeld een intensivering van het onderwijsaanbod het gestelde perspectief bereikt kan worden, kan het gestelde perspectief worden gehandhaafd.
- Waren er specifieke omstandigheden die een verklaring kunnen bieden voor de tegenvallende resultaten? Wanneer de leerling bijvoorbeeld een slechte dag heeft gehad, of wanneer de leerling een moeilijke periode achter de rug heeft, is het niet altijd noodzakelijk de gestelde doelen bij te stellen.
- Zijn de doelen (nog) realistisch? Mogelijk zijn er specifieke leerling- en/of omgevings-kenmerken die ertoe leiden dat het perspectief naar beneden bijgesteld moet worden. Indien de leerling bijvoorbeeld van een vrij stabiele naar een meer onrustige thuissituatie is gegaan, kan dit aanleiding zijn om het gestelde perspectief bij te stellen. Het verdient echter aanbeveling om in eerste instantie te kijken of het gestelde perspectief door aanpassing van het onderwijsaanbod alsnog behaald kan worden. Bijstellen van het perspectief is alleen raadzaam wanneer er duidelijk aanwijsbare (blijvende) factoren zijn die de tegenvallende resultaten kunnen verklaren.

4 De kwaliteit van toetsen in het primair onderwijs

4 De kwaliteit van toetsen in het primair onderwijs

Piet Sanders

Gebruikers van toetsen moeten kunnen vertrouwen op de kwaliteit van toetsen, zeker als het toetsen betreft op basis waarvan belangrijke beslissingen genomen worden. De kwaliteit van toetsen in het primair onderwijs wordt beoordeeld door de Commissie Testaangelegenheden Nederland (COTAN) en de Expertgroep Toetsen Primair Onderwijs (Expertgroep).

Alle toetsen en tests worden beoordeeld met het 'COTAN Beoordelingssysteem voor de kwaliteit van tests' (Evers e.a., 2010) en de eindtoetsen worden beoordeeld door de Expertgroep (zie <http://expertgroepoetsenpo.nl>). Voor de beoordeling van de onderwijskundige en organisatorische aspecten van de eindtoetsen wordt een door de Expertgroep ontwikkeld beoordelingskader gebruikt en de psychometrische aspecten worden met het beoordelingssysteem van de COTAN (zie http://expertgroepoetsenpo.nl/beoordelingskader_eindtoets) beoordeeld. De beoordeling van de onderwijskundige en psychometrische aspecten van de LOVS-toetsen vindt plaats met een beoordelingskader (zie http://expertgroepoetsenpo.nl/beoordelingskader_lovstoetsen) dat door de Expertgroep ontwikkeld is.

Paragraaf 1 van dit hoofdstuk bevat een beschrijving van het COTAN beoordelingssysteem plus een voorbeeld van een COTAN beoordeling van een toets. Paragraaf 2 en 3 zijn respectievelijk gewijd aan de beoordelingskaders van de Expertgroep voor eindtoetsen en LOVS-toetsen en voorbeelden van beoordelingen van die toetsen. Het hoofdstuk sluit af met conclusies en discussie in paragraaf 4.

4.1 COTAN Beoordelingssysteem voor de kwaliteit van tests

Het doel van de COTAN, een commissie van het Nederlands Instituut van Psychologen (NIP), is het bevorderen van de kwaliteit van tests en testgebruik in Nederland en het informeren van testgebruikers over de stand van zaken op testgebied. Voor uitgebreide informatie over de COTAN zie www.psynip.nl > Tests/COTAN. De COTAN beoordeelt tests en toetsen sinds 1969 met behulp van een beoordelingssysteem dat in de loop der jaren steeds verder ontwikkeld is. In 2009 is de laatste versie van het beoordelingssysteem getiteld 'COTAN Beoordelingssysteem voor de kwaliteit van tests' verschenen. In 2010 verscheen een herziene druk en met ingang van 1 juli 2015 zijn twee addenda aan het beoordelingssysteem toegevoegd.

De COTAN beoordeelt de kwaliteit van tests op basis van zeven criteria:

- 1 Uitgangspunten van de testconstructie
- 2 Kwaliteit van het testmateriaal
- 3 Kwaliteit van de handleiding
- 4 Normen
- 5 Betrouwbaarheid
- 6 Begripsvaliditeit
- 7 Criteriumvaliditeit

Twee toetsdeskundigen beoordelen de toetsen onafhankelijk van elkaar. Is er sprake van discrepanties, dan bespreekt een onafhankelijke coördinator die met beide beoordelaars. Indien nodig schakelt de coördinator een derde deskundige in. Nadat de eindbeoordeling naar de auteur(s) van de test gestuurd is, kan deze binnen een maand op de beoordeling reageren waarna de COTAN de eindbeoordeling al of niet aanpast. De definitieve eindbeoordelingen van toetsen worden opgeslagen in een database die men kan raadplegen door een abonnement te nemen, zie www.cotandocumentatie.nl. Van veel toetsen die in het onderwijs gebruikt worden, kan men de beoordelingen ook vinden op www.toetswijzer.nl > Toetsgids. Toetsen kan men ter beoordeling aan de COTAN aanbieden maar toetsen worden ook ongevraagd door de COTAN beoordeeld.

Hierna worden net als in hoofdstuk 10 van Toetsen op School (Sanders e.a., 2013) de zeven kwaliteitscriteria van de COTAN gepresenteerd en besproken. De presentatie bestaat uit een letterlijke weergave van de belangrijkste criteria uit het beoordelingssysteem. De lezer die kennis wil nemen van het volledige beoordelingssysteem verwijzen we naar www.psynip.nl > Tests / COTAN > COTAN Beoordeling > Beoordelingssysteem. De bespreking hier bestaat uit het geven van literatuurverwijzingen en toelichtingen bij de COTAN-criteria.

1 Uitgangspunten van de testconstructie

Bij dit criterium worden drie vragen gesteld:

- 1.1 (Basisvraag) Is er aangegeven wat het gebruiksdoel is van de test?
 - a Is er aangegeven welke construct(en) de test beoogt te meten?
 - b Is er aangegeven wat de doelgroep(en) is (zijn) van de test?
 - c Is er aangegeven wat de functie is van de test?
- 1.2 Is de herkomst van het constructie-idee beschreven en/of worden de te meten constructen gedefinieerd?
- 1.3 Wordt de relevantie van de testinhoud voor de te meten construct(en) aannemelijk gemaakt?

In het beoordelingssysteem staan bij elke vraag aanwijzingen voor de beoordelaar.

Bij vraag 1.1a moet de beoordelaar nagaan of het duidelijk is welke constructen de test beoogt te meten. Voorbeelden van constructen zijn intelligentie, leesvaardigheid, prestatiemotivatie of ADHD. Vraag 1.1b spreekt voor zich terwijl vraag 1.1c uitgebreid aan de orde is geweest in hoofdstuk 1.

Bij vraag 1.2 moet de beoordelaar nagaan of de test aansluit bij een bestaande theorie over bijvoorbeeld leesvaardigheid of intelligentie en of die theorie voldoende beschreven is.

Bij vraag 1.3 moet de beoordelaar nagaan of er een zodanige omschrijving van het itemdomein beschikbaar is dat duidelijk is of een willekeurig item uit de test wel of niet tot het construct behoort.

Hoe de COTAN-criteria gescoord/beoordeeld worden, lichten we hier toe voor criterium 1.

Alle (sub)vragen worden door een beoordelaar als onvoldoende (1 scorepunt), voldoende (2 scorepunten) of goed (3 scorepunten) beoordeeld. Vraag 1.1 is een basisvraag wat betekent dat als een van de subvragen, 1.1a, 1.1b of 1.1c, als onvoldoende beoordeeld is, criterium 1 als onvoldoende beoordeeld wordt en dat men de rest van de vragen van dit criterium kan overslaan. Indien dit laatste niet het geval is, gebeurt de vaststelling van het eindoordeel voor criterium 1 op basis van het aantal scorepunten op de (sub)vragen die resulteert in een onvoldoende, voldoende of goede beoordeling van dit criterium. Zo is het eindoordeel voor criterium 1 'goed' als de som van de beoordelingen op de subvragen van vraag 1.1, 8 of 9 scorepunten bedraagt en als op de vragen 1.2 en 1.3 minstens 2 scorepunten toegekend zijn. Het eindoordeel voor criterium 1 is 'voldoende' als de som van de beoordelingen op de subvragen van vraag 1.1, 6 of 7 scorepunten bedraagt, aan geen van de subvragen van 1.1, 1

scorepunt toegekend is en als op de vragen 1.2 en 1.3 minstens 2 scorepunten toegekend zijn. Is in het laatste geval aan een van de vragen 1.2 of 1.3, 1 scorepunt toegekend dan is het eindoordeel 'onvoldoende'.

Wat betreft de beoordeling en scoring van de andere criteria verwijzen we de lezer naar de COTAN-publicatie van het beoordelingssysteem op www.psynip.nl.

2 Kwaliteit van het testmateriaal

Bij de beoordeling van dit criterium wordt onderscheid gemaakt tussen tests die schriftelijk of met behulp van de computer worden afgenomen.

Bij dit criterium worden voor beide afnamewijzen drie basisvragen gesteld:

- 2.1 Zijn de testopgaven gestandaardiseerd?
- 2.2
 - a Is er sprake van een objectief scoringssysteem? en/of
 - b Als de scoring door observatoren gebeurt, is dan het beoordelings- of observatiesysteem volledig en duidelijk?
- 2.3 Zijn de items vrij van racistische, etnocentrische, seksistische en voor bepaalde bevolkingsgroepen kwetsende inhoud?

Bij vraag 2.1 moet de beoordelaar nagaan of de items/opdrachten wat betreft vorm, inhoud en volgorde voor iedereen hetzelfde zijn. Standaardiseren is belangrijk als men de toetsscores van verschillende personen/leerlingen wil interpreteren en vergelijken. De (on)mogelijkheden om te standaardiseren verschillen nogal per toetsvorm. Toetsen bestaande uit meerkeuzevragen zijn eenvoudiger te standaardiseren dan toetsen bestaande uit open vragen en die zijn weer eenvoudiger te standaardiseren dan toetsen bestaande uit praktische opdrachten.

Bij vraag 2.2a moet de beoordelaar nagaan of de scores die aan alle mogelijke antwoorden van personen worden toegekend bij voorbaat zodanig vastliggen, dat elke testleider/docent, afgezien van administratieve fouten die bij de scoring gemaakt kunnen worden, tot dezelfde score zal komen.

Bij vraag 2.2b moet de beoordelaar nagaan of de richtlijnen voor de beoordeling/scoring, waaronder modelantwoorden, modelgedragingen, schaalankers en dergelijke, de objectiviteit van de beoordeling zo goed mogelijk waarborgen.

Voor een toelichting bij vraag 2.3 verwijzen we naar de publicatie van Bügel en Sanders (1998) getiteld 'Richtlijnen voor de ontwikkeling van onpartijdige toetsen', die men kan downloaden op de website van Cito, www.cito.nl > Onderzoek en wetenschap > Psychometrie > Richtlijnen ontwikkeling toetsen > Richtlijnen ontwikkelen onpartijdige toetsen.

Behalve de drie basisvragen worden bij criterium 2 voor beide afnamewijzen, schriftelijk en computer, nog twaalf andere vragen gesteld. Het betreft vragen naar de kwaliteit van de instructie voor de geteste, de correcte formulering van de items, de kwaliteit van het testmateriaal, etc. Voor meer informatie hierover verwijzen we naar het COTAN beoordelings-systeem en de hoofdstukken 6, 7 en 8 uit Toetsen op School (Sanders e.a., 2013).

3 Kwaliteit van de handleiding

Bij dit criterium wordt gevraagd naar de volledigheid van de informatie die de handleiding biedt voor de gebruiker. Hierbij gaat het enerzijds om praktische aanwijzingen voor de afname, scoring en interpretatie (gebruikershandleiding) en anderzijds om informatie over onderzoek dat met de test is verricht (wetenschappelijke of technische verantwoording). Beide soorten informatie zijn voor de gebruiker van belang om te kunnen beoordelen welke conclusies er aan een testscore kunnen worden verbonden.

Bij dit criterium worden de volgende vragen gesteld:

- 3.1 (Basisvraag) Is er een handleiding beschikbaar?
- 3.2 Zijn de aanwijzingen voor de testleider volledig en duidelijk?
- 3.3 Wordt er informatie gegeven over de gebruiksmogelijkheden en beperkingen van de test?
- 3.4 Wordt in de handleiding een samenvatting van de onderzoeksresultaten gegeven?
- 3.5 Wordt met behulp van voorbeelden aangegeven hoe testscores kunnen worden geïnterpreteerd?
- 3.6 Wordt er gewezen op informatie die bij de interpretatie van belang kunnen zijn?
- 3.7 Wordt de mate van deskundigheid vermeld die vereist is voor afname en interpretatie van de test?

Bij vraag 3.1 moet de beoordelaar nagaan of er een handleiding is. Indien er geen handleiding is, kan men de rest van de vragen van dit criterium overslaan.

Bij vraag 3.2 moet de beoordelaar nagaan of de aanwijzingen voor de testleider zodanig zijn dat de afname gestandaardiseerd plaatsvindt. Er moet zoveel mogelijk letterlijk zijn voorgeschreven wat de testleider wel en niet mag zeggen en welke handelingen de testleider moet verrichten. Ook moet voorgeschreven worden hoe de testleider op vragen moet ingaan (er kunnen bijvoorbeeld standaardteksten worden gegeven voor antwoorden op veel voorkomende vragen), welke mate van ondersteuning mag worden geboden en welke hulpmiddelen de geteste mag gebruiken.

Bij vraag 3.3 moet de beoordelaar nagaan of de handleiding volledig, nauwkeurig en duidelijk is over de gebruiksmogelijkheden en beperkingen van de test. Het moet voor de testgebruiker duidelijk zijn welke constructen met behulp van de test worden gemeten, voor welke doelgroep de toets bedoeld is en wat het doel van de toets is. Verder moeten de beperkingen van de test zijn beschreven. Is er bijvoorbeeld bij classificatiebeslissingen in het onderwijs aangegeven dat de beslissing niet op een enkele toets mag worden gebaseerd? Wijst men bij het gebruik van voortgangstoetsen op de relatie tussen de toetsscore en het verdere onderwijsleerproces?

Bij vraag 3.4 moet de beoordelaar nagaan of de samenvatting van de resultaten van normerings-, betrouwbaarheids- en validiteitsonderzoek zodanig is dat een gebruiker van de test zich een oordeel kan vormen of de test voor zijn doeleinden geschikt is en de vereiste kwaliteit heeft. Het antwoord op deze vraag zal afhankelijk zijn van de expertise die een gebruiker heeft of die van een gebruiker verwacht mag worden. De afgelopen jaren is een aanvang gemaakt met het opleiden en certificeren van personen die bij construeren, beoordelen en afnemen van toetsen betrokken zijn. Diverse organisaties bieden opleidingen aan voor toetsconstruëer, toetsvaststeller en assessor.

Bij vraag 3.5 moet de beoordelaar nagaan of voorbeelden gegeven worden van hoe de toetsscore wel en hoe niet geïnterpreteerd mag worden. Een voorbeeld van de nauwkeurigheid van toetsscores en cijfers hebben we in hoofdstuk 3 van Toetsen op School (Sanders e.a., 2013) bij de bespreking van de standaardmeetfout gepresenteerd.

Bij vraag 3.6 moet de beoordelaar nagaan of er bijvoorbeeld informatie gegeven wordt over de mogelijke invloed van achtergrondvariabelen en (test)ervaring op de scores of dat aangegeven wordt welke andere variabelen aan bijvoorbeeld de voorspelling van de Eindtoets Basisonderwijs bijdragen. In het laatste geval zijn de oordelen van docenten en van ouders een belangrijke bron van informatie.

Bij vraag 3.7 moet de beoordelaar nagaan over welke mate van deskundigheid de gebruiker dient te beschikken. Welke professionals zijn gezien hun opleiding en werkervaring geschikt? Welke kennis en vaardigheden zijn voor de afname en interpretatie van de test noodzakelijk? Zo wordt een docent wel geschikt geacht om de score behaald op een eindtoets te interpreteren maar niet de score behaald op een intelligentietest.

4 Normen

Bij dit criterium worden eerst twee algemene basisvragen gesteld:

- 4.1 Worden er normen verstrekt?
- 4.2 Zijn de normen actueel?

Indien de beoordelaar bij vraag 4.1 constateert dat er geen normen beschikbaar zijn, kan men de rest van de vragen van dit criterium overslaan.

Bij vraag 4.2 moet de beoordelaar nagaan of de normen nog actueel zijn. Het standpunt van de COTAN is dat normen na twintig jaar niet meer bruikbaar zijn en dus niet voldoen aan dit criterium.

Na de twee algemene basisvragen volgen er (basis)vragen afhankelijk van of er sprake is van normgerichte, domeingerichte of criteriumgerichte interpretatie van de scores. Bij normgerichte interpretatie wordt de testscore van een persoon vergeleken met de testscores van andere personen. Dit type normen wordt ook wel relatieve normen genoemd. Bij toetsen die in het primair onderwijs gebruikt worden, is er bijna altijd sprake van normgerichte interpretatie. Voor de volledigheid bepreken we hier ook kort domeingerichte en criteriumgerichte interpretatie.

Indien er sprake is van normgerichte interpretatie worden de volgende vragen gesteld:

- 4.3 (Basisvraag) Wat is de kwaliteit van de verstrekte normgroepen?
 - a Zijn de normgroepen groot genoeg?
 - b Zijn de normgroepen representatief?
- 4.4 Worden de betekenis en de beperkingen van de normschaal duidelijk gemaakt voor de gebruiker en is het type normschaal in overeenstemming met het doel van de test?
- 4.5 Worden er gemiddelden, standaardafwijkingen en gegevens over de scoreverdeling vermeld?
- 4.6 Worden er gegevens verstrekt over mogelijke verschillen tussen subgroepen (bijvoorbeeld allochtonen-autochtonen, mannen-vrouwen)?
- 4.7 Worden er gegevens verstrekt over de nauwkeurigheid van de meting en de daarbij behorende intervallen?
 - a standaardmeetfout
 - b standaardschattingsfout
 - c testinformatiefunctie

Bij vraag 4.3a heeft de COTAN de volgende beoordelingsregels opgesteld:

- Bij tests voor belangrijke beslissingen op individueel niveau (bijvoorbeeld personeelsselectie, verwijzing naar speciaal onderwijs of certificering) wordt een normgroep van 400 of meer personen als goed, een groep van 300 of meer tot 400 personen als voldoende en een groep van minder dan 300 personen als onvoldoende beoordeeld.
- Bij tests voor relatief minder belangrijke beslissingen op individueel niveau (bijvoorbeeld voortgangscntrole) wordt een normgroep van 300 of meer personen als goed, een groep van 200 of meer tot 300 personen als voldoende en een groep van minder dan 200 personen als onvoldoende beoordeeld.

Met de representativiteit van normgroepen bij vraag 4.3b wordt bedoeld of de samenstelling van de onderzoeksgroep, bijvoorbeeld de leerlingen van groep 8, qua regio, etniciteit, urbanisatiegraad en sekse overeenkomt met de populatie leerlingen van groep 8.

Bij vraag 4.4 moet de beoordelaar nagaan of de gebruikte normschaal in overeenstemming is met het doel van de test. Het is bijvoorbeeld niet zinvol een percentielschaal te gebruiken als de minimumscore op een test 0 en de maximumscore 20 bedraagt omdat dan veel scorepunten dezelfde percentielscore zullen hebben.

Wat betreft vraag 4.7 wordt de lezer voor een toelichting bij 4.7a verwezen naar hoofdstuk 3 van Toetsen op School (Sanders e.a., 2013). Voor de begrippen standaardschattingfout en testinformatiefunctie wordt verwezen naar Eggen en Sanders (1993).

Bij domeingerichte interpretatie wordt de testscore van een persoon vergeleken met een of meer standaarden, cesuren of grensscores. De standaarden worden op de een of andere wijze door beoordelaars bepaald. Dit type normen wordt ook wel absolute normen genoemd.

Indien er sprake is van domeingerichte interpretatie worden de volgende vragen gesteld:

- 4.8 (Basisvraag) Is er voldoende overeenstemming tussen de beoordelaars?
- 4.9 Zijn de procedures op grond waarvan de grensscores zijn bepaald correct?
- 4.10 Zijn de beoordelaars naar behoren geselecteerd en getraind?

Bij vraag 4.8 moet de beoordelaar nagaan of de overeenstemming tussen de experts of beoordelaars voldoende hoog is om de standaard te kunnen legitimeren.

Bij vraag 4.9 moet de beoordelaar nagaan of de gevolgde stappen en beslissingen in overeenstemming zijn met de gebruikte standaardbepalingsmethode.

Bij vraag 4.10 moet de beoordelaar nagaan welke procedure gevolgd is bij de selectie van de experts en of de experts getraind zijn in het toepassen van de standaardbepalingsmethode.

Bij criteriumgerichte interpretatie wordt de standaard, cesuur of grensscore aan onderzoeksgegevens ontleend. Bij deze wijze van normeren moeten er naast de gegevens op de test ook gegevens over het criterium zijn verzameld. Zo kan bijvoorbeeld bij het toekennen van licenties of diploma's de grensscore of cesuur bepaald worden door na te gaan bij welke testscore de gunstigste verhouding bestaat tussen deelnemers die in de praktijk succesvol en onsuccesvol blijken te zijn.

Indien er sprake is van criteriumgerichte interpretatie worden de volgende vragen gesteld:

- 4.11 (Basisvraag) Rechtvaardigen de onderzoeksresultaten het gebruik van grensscores?
- 4.12 Is de onderzoeksgroep in overeenstemming met het bedoelde gebruik?
- 4.13 Is de onderzoeksgroep groot genoeg?

Bij vraag 4.11 moet de beoordelaar nagaan of de testauteur voldoende bewijs levert voor de bruikbaarheid van de gekozen grensscore.

Bij vraag 4.12 moet de beoordelaar nagaan of het onderzoek waarin de grensscore bepaald is, betrekking heeft op de populatie waarvoor de test gebruikt is.

Bij vraag 4.13 heeft de COTAN de volgende beslissingsregel opgesteld. Ervan uitgaand dat grensscores alleen worden bepaald in situaties waarin het gaat om belangrijke beslissingen op individueel niveau (zie vraag 4.3a), wordt een onderzoeksgroep bestaande uit minstens 300 personen als goed, een groep bestaande uit minstens 200 personen als voldoende en een onderzoeksgroep bestaande uit minder dan 200 personen als onvoldoende beoordeeld.

5 Betrouwbaarheid

Bij dit criterium worden drie (sub)vragen gesteld:

- 5.1 (Basisvraag) Worden er gegevens over de betrouwbaarheid verstrekt?
- 5.2 Zijn de resultaten voldoende, gelet op het beoogde type beslissingen dat met behulp van de test wordt genomen?
- 5.3 Wat is de kwaliteit van het onderzoek naar de betrouwbaarheid?

Bij vraag 5.1 moet de beoordelaar nagaan of er informatie over de betrouwbaarheid van de testscores verstrekt is. Indien er geen informatie beschikbaar is, kan men de rest van de vragen van dit criterium overslaan.

Bij vraag 5.2 heeft de COTAN de volgende beoordelingsregels opgesteld:

- Bij tests voor belangrijke beslissingen op individueel niveau (bijvoorbeeld het zakken of slagen van leerlingen, doorverwijzing naar speciaal onderwijs) wordt een betrouwbaarheid van 0,90 of hoger als goed, van 0,80 of hoger en kleiner dan 0,90 als voldoende en kleiner dan 0,80 als onvoldoende beoordeeld.
- Bij tests voor minder belangrijke beslissingen op individueel niveau (bijvoorbeeld voortgangscntrole) wordt een betrouwbaarheid van 0,80 of hoger als goed, van 0,70 of hoger en kleiner dan 0,80 als voldoende en kleiner dan 0,70 als onvoldoende beoordeeld.
- Bij tests voor onderzoek op groepsniveau (bijvoorbeeld klimaat in de klas) wordt een betrouwbaarheid van 0,70 of hoger als goed, van 0,60 of hoger en kleiner dan 0,70 als voldoende en kleiner dan 0,60 als onvoldoende beoordeeld.

Hoofdstuk 3 van *Toetsen op School* (Sanders e.a., 2013) bespreekt het gebruik van de betrouwbaarheidscoëfficiënt en andere maten zoals de standaardmeetfout en het percentage misclassificaties om de betrouwbaarheid van een toets te kwantificeren. Hoewel genoemde maten met elkaar samenhangen, wordt hier voorgesteld om niet alleen de betrouwbaarheidscoëfficiënt maar ook de andere genoemde maten naast of in plaats van de betrouwbaarheidscoëfficiënt te gebruiken.

De beoordeling van validiteit volgens de COTAN

Voorafgaande aan de presentatie van de criteria die COTAN hanteert bij de beoordeling van de validiteit van een toets, dient opgemerkt te worden dat er in de testliteratuur verschillende opvattingen over validiteit bestaan. Volgens hoofdstuk 4 van *Toetsen op School* (Sanders e.a., 2013), heeft validiteit betrekking op de interpretatie en het gebruik van toetsscores en bestaat het valideren van toetsscores uit bewijzen die de interpretatie en het gebruik ondersteunen. De COTAN gaat uit van de klassieke driedeling van validiteit: inhoudsvaliditeit, begripsvaliditeit en criteriumvaliditeit. Van deze drie wordt validiteitsinformatie die betrekking heeft op de relevantie van de inhoud van een test (inhoudsvaliditeit) en op de betekenis van een testscore (begripsvaliditeit) voor alle typen tests van belang geacht, ongeacht het doel van de toets. Dit geldt echter niet voor informatie over de voorspellende waarde van testcores (criteriumvaliditeit): voor tests die geen voorspellende pretentie hebben, is dit type informatie niet vereist. De drie soorten validiteit die de COTAN onderscheidt, komen ook in hoofdstuk 4 van *Toetsen op School* (Sanders e.a., 2013) aan de orde. Inhoudsvaliditeit wordt door de COTAN echter opgevat als onderdeel van het testontwikkelingsproces en dus van criterium 1 en dan met name criterium 1.3. De criteria begripsvaliditeit en criteriumvaliditeit worden hierna besproken.

6 Begripsvaliditeit

Bij dit criterium worden de volgende vragen gesteld:

- 6.1 (Basisvraag) Worden er gegevens over de begripsvaliditeit verstrekt?
- 6.2 Maken de resultaten voldoende aannemelijk dat het begrip zoals bedoeld, wordt gemeten?
- 6.3
 - a Zijn de procedures op basis waarvan de begripsvaliditeitsgegevens zijn berekend correct?
 - b Komen de steekproeven die in het begripsvalideringsonderzoek zijn gebruikt, overeen met groepen waarvoor de test is bedoeld?
 - c Wat is de kwaliteit van de andere maten die in het begripsvalideringsonderzoek zijn gebruikt?
 - d Is de kwaliteit van het onderzoek, zoals beoordeeld in de vragen 6.3a tot en met 6.3c, zodanig dat de beoordeling van de begripsvaliditeit in vraag 6.2, kan worden bevestigd?

Indien de beoordelaar bij vraag 6.1 constateert dat er geen gegevens over de begripsvaliditeit

verstrekt zijn, kan men de rest van de vragen van dit criterium overslaan.

Voor toelichting bij vraag 6.2 wordt de lezer verwezen naar hoofdstuk 4 van Toetsen op School (Sanders e.a., 2013), met name de paragrafen die over de verschillende soorten bewijzen gaan. Bij vraag 6.3a tot en met 6.3d moet de beoordelaar nagaan of het onderzoek naar de begripsvaliditeit goed uitgevoerd is (6.3a en 6.3b) en dat de toetsen of andere meetinstrumenten die in het onderzoek gebruikt zijn, voldoende betrouwbaar zijn (6.3c).

7 Criteriumvaliditeit

Bij dit criterium worden de volgende vragen gesteld:

- 7.1 (Basisvraag) Worden er gegevens verstrekt over het verband test-criterium?
- 7.2 Zijn de resultaten voldoende gelet op het type beslissingen dat met de test genomen moet worden?
- 7.3
 - a Zijn de procedures op grond waarvan de criteriumvaliditeitsgegevens zijn berekend correct?
 - b Zijn de steekproeven op grond waarvan de criteriumvaliditeitsgegevens zijn berekend in overeenstemming met het beoogde testgebruik?
 - c Wat is de kwaliteit van de criteriummaten?
 - d Is de kwaliteit van het onderzoek, zoals beoordeeld in de vragen 7.3a tot en met 7.3c, zodanig dat de beoordeling van de criteriumvaliditeit in vraag 7.2, kan worden bevestigd?

Indien de beoordelaar bij vraag 7.1 constateert dat er geen gegevens over de criteriumvaliditeit verstrekt zijn, kan men de rest van de vragen van dit criterium overslaan. Zoals hiervoor reeds opgemerkt werd, is de criteriumvaliditeit niet voor elke test relevant.

Voor toelichting bij vraag 7.2 wordt de lezer verwezen naar hoofdstuk 4 van Toetsen op School (Sanders e.a., 2013), met name naar de paragraaf waar de voorspellende waarde of criteriumvaliditeit van de Centrale Eindtoets besproken wordt.

Bij vraag 7.3a tot en met 7.3d moet de beoordelaar nagaan of het onderzoek naar de criteriumvaliditeit goed uitgevoerd is (7.3a en 7.3b) en dat de toetsen of andere meetinstrumenten die in het onderzoek als criterium gebruikt worden, voldoende betrouwbaar zijn (7.3c).

De COTAN heeft in 2015 besloten het beoordelingssysteem uit te breiden met twee addenda die per 1 juli 2015 in werking treden. Het aantal criteria blijft echter zeven.

Het eerste addendum betreft 'unproctored gegevensverzameling' en valt onder basisvraag 4.3b 'Zijn de normgroepen representatief?', aangezien deze basisvraag gaat over de manier van gegevensverzameling. Tot nu toe werd in het beoordelingssysteem steeds uitgegaan van proctored gegevensverzameling, d.w.z. gegevensverzameling onder toezicht van de testleider. Met de komst van computertests heeft unproctored gegevensverzameling, d.w.z. gegevensverzameling zonder toezicht van een testleider, zijn intrede gedaan. Naast voordelen zoals dat kandidaten op verschillende locaties de test kunnen afleggen, is een groot nadeel dat de identiteit van de kandidaat nooit met zekerheid kan worden vastgesteld wat tot fraude aanleiding kan geven. Vandaar dat bij tests met unproctored gegevensverzameling een volledige beschrijving dient te zijn opgenomen van hoe fraude bij de testafname te voorkomen valt.

Het tweede addendum is 'fairness' of onpartijdigheid en betreft de geschiktheid van (psychologische) tests bij verschillende groepen, onafhankelijk van hoe die groepen gedefinieerd zijn. Fairness is volgens de COTAN geen geïsoleerd concept en kan meerdere aspecten van een test betreffen. Of er al of niet sprake is van fairness wordt onderzocht aan de hand van een zogenaamde 'fairness matrix'. Uit het onderzoek kan bijvoorbeeld blijken dat er geen of onvoldoende rekening gehouden is met personen die dyslexie hebben.

Naar aanleiding van de twee addenda zijn de tabellen voor de vaststelling van het eindoordeel

van een criterium van het COTAN Beoordelingssysteem aangepast.

4.1.1 Voorbeeld van een COTAN-beoordeling

COTAN-beoordelingen van een groot aantal toetsen staan in Toetsgids op de website Toetswijzer die door Cito beheerd wordt, zie www.toetswijzer.nl > Toetsgids. Hieronder staat een download van de ADHD-Vragenlijst (AVL) uit Toetsgids. De toetsbeoordeling informeert de potentiële gebruiker over verschillende aspecten van de toets en eindigt met de COTAN beoordeling op de zeven beoordelingscriteria. Op basis van deze informatie kan de potentiële gebruiker beslissen of de toets al of niet geschikt is voor het doel waarvoor de toets ingezet gaat worden. Is die beslissing positief dan kan de potentiële gebruiker verdere informatie inwinnen bij de uitgever van de test.

ADHD-Vragenlijst (AVL)

Omschrijving

De AVL is een gedragsbeoordelingslijst waarmee kan worden nagegaan of, en in welke mate, een kind de gedragsymptomen van ADHD (Attention Deficit Hyperactivity Disorder) vertoont. De vragenlijst dient te worden ingevuld door beoordelaars die goed op de hoogte zijn van het gedrag van het kind.

Informatie en bestellen

Uitgever Bohn Stafleu van Loghum [Info] [Site] [Meer van...]
Jaar van uitgave 2005

Inhoud en gebruik

De vragenlijst bestaat uit 18 items zoals 'wordt gemakkelijk afgeleid' en 'praat aan een stuk door'. Op vijf puntsschalen van 'niet' tot 'zeer vaak' wordt door ouders, leerkrachten, groepsopvoeders of andere hulpverleners (die het gedrag van de jeugdige goed kennen) aangegeven in welke mate de jeugdige de afgelopen maanden het gedrag vertoonde. De vragenlijst omvat drie subschalen: Aandachtstekort (snel afgeleid en weinig taakgericht), Hyperactiviteit (overbeweeglijkheid en motorische onrust), Impulsiviteit (ongecontroleerd en ongeremd gedrag).

De AVL wordt gebruikt om ADHD-gedrag thuis, op school of in de (semi)residentiële leefgroep te inventariseren en vormt een onderdeel van de ADHD-diagnostiek. Bij periodieke afname kan de ontwikkeling van ADHD-symptomatologie worden gevolgd en kan de behandeling worden in- of bijgesteld.

Doelgroep

Normaal jeugdigen en klinische jeugdigen van 4 t/m 18 jaar.

Meetpretentie

Gedragsymptomen van ADHD-problematiek.

Doel

- Screening binnen onderwijs.
- Diagnose binnen speciaal onderwijs en jeugdzorg.
- Evalueren van voortgang en effectiviteit van behandelingen.
- Te gebruiken bij indicatiestelling voor het speciaal onderwijs en leerlinggebonden financiering.

Voorbeelditem

Er is geen voorbeelditem van deze toets.

Opmerking

De AVL kan ook via de pc worden afgenomen en gescoord met behulp van Testweb.

Kernpublicatie

Scholte, E.M., & Ploeg, J.D. van der (2005). ADHD Vragenlijst: Handleiding. Houten: Bohn Stafleu van Loghum.

Weblinks

- Informatie over AVL op Testweb van Bohn Stafleu van Loghum.

COTAN-beoordeling

	Criteria	Beoordeling
UTC	Uitgangspunten bij de testconstructie [Toelichting]	Goed
KTM	Kwaliteit van het testmateriaal [Toelichting]	Goed
KHL	Kwaliteit van de handleiding [Toelichting]	Goed
NRM	Normen [Toelichting]	Voldoende
BTR	Betrouwbaarheid [Toelichting]	Goed
BVA	Begripsvaliditeit [Toelichting]	Goed
CVA	Criteriumvaliditeit [Toelichting]	Voldoende

4.2 Beoordelingskader Eindtoets

Met ingang van het schooljaar 2014-2015 moeten alle schoolverlaters in het reguliere basisonderwijs een eindtoets maken die hun taal- en rekenvaardigheden meet. Daarnaast kunnen scholen er vrijwillig voor kiezen om ook een toets wereldoriëntatie af te nemen. In 2016 konden scholen uit het basisonderwijs kiezen tussen de Centrale Eindtoets en twee andere eindtoetsen. De eindtoetsen worden afgenomen tussen 15 april en 15 mei, en niet meer in februari. Hiermee wordt aangegeven dat niet het resultaat op de eindtoets maar het schooladvies leidend is bij de toelating tot het voortgezet onderwijs. De Centrale Eindtoets heeft een waarborgfunctie en moet dus altijd beschikbaar zijn ongeacht de mate waarin sprake is van andere eindtoetsen waaruit kan worden gekozen. Mocht een school namelijk de eindtoets van zijn keuze onverhoopt niet kunnen afnemen dan kan altijd een beroep gedaan worden op de waarborgfunctie van de Centrale Eindtoets. Voor inhoudelijke informatie over de Centrale Eindtoets en de andere eindtoetsen verwijzen we naar hoofdstuk 1 en voor wettelijke informatie naar het Toetsbesluit PO, www.rijksoverheid.nl > Documenten > Toetsbesluit PO.

Aan welke inhoudelijke en procedurele kwaliteitseisen de Centrale Eindtoets moet voldoen, staat beschreven in een verantwoordingsdocument van het CvTE (2015). Die eisen zijn vergelijkbaar met de eisen die voor de andere eindtoetsen gelden. Een beoordeling van de kwaliteit van de Centrale Eindtoets door een externe organisatie is op het moment van verschijnen van deze publicatie nog niet beschikbaar. Het is de bedoeling dat er in 2017-2018 een adaptieve versie van de Centrale Eindtoets aangeboden wordt, zie voor meer informatie www.centraleeindtoetspo.nl > Voor schoolbesturen en directies > Algemeen > Adaptieve Centrale Eindtoets. Extern toezicht op de Centrale Eindtoets zal plaatsvinden door de Inspectie en via audits door externe instanties.

De andere eindtoetsen moeten voldoen aan kwaliteitseisen die vastgelegd zijn in een beoordelingskader. Op basis van de beoordeling van een eindtoets met het beoordelingskader beslist de Expertgroep Toetsen PO of een eindtoets wel of niet toegelaten wordt. Voordat we het beoordelingskader eindtoetsen presenteren, vermelden we eerst de taken van de Expertgroep. We sluiten de paragraaf af met een bespreking van de beoordeling van eindtoetsen die wel en niet toegelaten zijn.

Volgens het 'Instellingsbesluit Expertgroep toetsen PO' is de Expertgroep een onafhankelijke commissie die tot taak heeft:

- a de minister te adviseren over: 1. De toelating van andere eindtoetsen; 2. De intrekking van de andere eindtoetsen;
- b het geven van een eigenstandig oordeel over de inhoudelijke validiteit, betrouwbaarheid en deugdelijke normering van een tussentijdse toets of reeks van tussentijdse toetsen.

Op de website van de Expertgroep, www.expertgroepoetsenpo.nl > Beoordelingskader, staan dat de Expertgroep het Beoordelingskader Eindtoets gebruikt. Dit beoordelingskader beschrijft:

- de aspecten waarop eindtoetsen worden beoordeeld;
- de wijze waarop de eindtoetsen worden beoordeeld;
- de wijze waarop de Expertgroep tot een advies komt.

De voorwaarden waaraan een eindtoets moet voldoen, staat beschreven in het Toetsbesluit PO. Voor een eindtoets, zie de website van de Expertgroep, gelden de volgende onderwijskundige/organisatorische en psychometrische kwaliteitscriteria:

- wordt jaarlijks aangeboden voor afname in het tijdvak 15 april tot en met 15 mei;
- meet ten minste de kennis en vaardigheden van de leerling op het terrein van Nederlandse taal en rekenen en wiskunde;
- heeft een mogelijkheid voor een tweede afnamemoment voor het geval een leerling verhinderd is bij de eerste afname;
- leidt tot een eenduidig advies aan de leerling omtrent het te volgen onderwijs;
- bestaat jaarlijks uit nieuwe opgaven voor taal en rekenen;
- kan ook gemaakt worden door leerlingen met een specifieke ondersteuningsbehoefte;
- biedt inzicht in de beheersing van de referentieniveaus taal en rekenen;
- beantwoordt aan de kerneisen van psychometrische kwaliteit: inhoudelijke kwaliteit, betrouwbaarheid en deugdelijke normering.

Het Beoordelingskader Eindtoets bestaat uit twee onderdelen:

- 1 Onderwijskundige en organisatorische aspecten die door onderwijsdeskundigen beoordeeld worden. Dit onderdeel van het beoordelingskader is in samenwerking met het expertisecentrum Leerplanontwikkeling (SLO) ontwikkeld. In de adviezen van de Expertgroep wordt dit onderdeel met 'onderwijsinhoud' aangeduid;
- 2 Psychometrische aspecten die door de COTAN beoordeeld worden. Voor de beoordeling gebruikt de COTAN haar beoordelingssysteem, aangevuld met kwaliteitsaspecten die betrekking hebben op de normering van referentieniveaus en adaptief testen.

De onderwijskundige/organisatorische aspecten van het beoordelingskader hebben betrekking op vijf thema's: inhoud, leerlingrapport, leerlingen, afname en organisatie toets. Bij elk thema worden kwaliteitseisen, operationaliseringsvragen (met mogelijke antwoorden) en beslisregels geformuleerd.

Voor bijvoorbeeld het thema 'Afname eindtoets' betekent toepassing van het beoordelingskader het volgende:

Kwaliteitseisen

- A1.1: De eindtoets wordt jaarlijks afgenomen in het tijdvak tussen 15 april en 15 mei.
- A1.2: De eindtoets heeft een mogelijkheid voor een tweede afnamemoment (voor leerlingen die de eerste afname zijn verhinderd).
- A1.3: Het bij de eindtoets horende toetsreglement (TR) regelt de wijze van afnemen, zoals beschreven in het Toetsbesluit, artikel 7.

Optimaliseringsvragen zoals

- A1.1 Is het eerste afnamemoment voorzien in het tijdvak 15 april – 15 mei? (ja/nee)
- A1.2 Is er een mogelijkheid voor een tweede afnamemoment? (ja/nee)
- A1.3.1 Staat in het TR/de handleiding vermeldt hoe de directeur de leerlingen voor de toets moet aanmelden? (ja/nee)
- A1.3.6 Worden in het TR/de handleiding de voorwaarden aan de (wijze van) waarop toezicht op de leerlingen wordt gehouden beschreven? (ja/nee)

Beslisregel afname eindtoets

Om dit thema met een voldoende te kunnen afsluiten dienen alle vragen met JA te worden beantwoord.

Op basis van de beoordelingen van de onderwijskundige/organisatorische en psychometrische aspecten stelt de Expertgroep een definitief advies op. Dit advies gaat naar de Minister van Onderwijs, Cultuur en Wetenschappen die, op basis van het ontvangen advies, een definitieve beslissing neemt over het al dan niet toelaten van de betreffende eindtoets.

De psychometrische aspecten van het beoordelingskader betreffen het 'COTAN Beoordelings-systeem voor de kwaliteit van tests' wat in paragraaf 1 uitgebreid besproken is. Ten behoeve van de beoordeling van eindtoetsen is aan het beoordelingssysteem een onderdeel over de normering van referentieniveaus en een onderdeel over adaptief toetsen toegevoegd.

4.2.1 Voorbeelden van beoordelingen andere eindtoetsen

Op de website van de Expertgroep staat dat de Expertgroep in 2014 twee adviezen en in 2015 drie adviezen heeft uitgebracht. Over de drie toetsen uit 2015, Dia-Eindtoets van Dia taal, CESAN van SM&C Internet Services en AMN Eindtoets van AMN, werd negatief geadviseerd. Over de twee toetsen uit 2014, de IEP Eindtoets van Bureau ICE en ROUTE 8 Van A-VISION werd positief geadviseerd.

Hieronder staan van de Dia-Eindtoets 2015 de beoordelingen van de onderwijskundige/organisatorische aspecten en van de psychometrische aspecten en het bijbehorende advies over de toelating.

De beoordeling van de onderwijskundige/organisatorische aspecten staat in onderstaande verzamelstaat.

Thema's	Oordeel
Inhoud	voldoende
• Validiteit toets	voldoende
• Handleiding	voldoende
• <i>Taal</i>	voldoende
• Lezen	voldoende
• Taalverzorging	voldoende
• <i>Rekenen</i>	voldoende
• Verdeling toetsopgaven over domeinen	voldoende
• Verdeling toetsopgaven over onderdelen	voldoende
• Opgaven met en zonder context	voldoende
• Gebruik kladpapier	voldoende
• Gebruik rekenmachine	voldoende
• Niveaus S en F (nog uitwerken)	voldoende
Leerlingrapport	voldoende
Leerlingen	voldoende
Afname eindtoets	voldoende
Organisatie eindtoets	voldoende

De beoordeling van de psychometrische aspecten van de Dia-Eindtoets betreft de COTAN-beoordeling uit 2015 en staat in onderstaande verzamelstaat.

Criteria	Beoordeling
Uitgangspunten bij de testconstructie	goed
Kwaliteit van het testmateriaal	goed
Kwaliteit van de handleiding	goed
Normen	onvoldoende
Betrouwbaarheid	onvoldoende
Begripsvaliditeit	onvoldoende
Criteriumvaliditeit	onvoldoende

De beoordeling van de psychometrische aspecten wordt uitvoerig toegelicht en in die toelichting staat dat de reden voor de onvoldoende voor 'normen' is 'onvoldoende informatie over het afnamedesign en de gebruikte steekproeven, onvoldoende bewijs voor equivalentie papieren en digitale versie en onvoldoende bewijs voor de gehanteerde grensscores.' De Dia-Eindtoets is aanvullend beoordeeld op het aspect 'normering referentieniveaus' en dit aspect is als 'onvoldoende' beoordeeld.

Op basis van bovenstaande beoordelingen heeft de Expertgroep haar advies opgesteld. Wat betreft de onderwijskundige/organisatorische aspecten constateert de Expertgroep dat hoewel op alle aspecten een voldoende gegeven is er bij een aantal aspecten wel kanttekeningen te plaatsen zijn. Het betreft de aspecten 'Validiteit toets', 'Taal-lezen', 'Rekenen-opgaven met en zonder context', 'Leerlingrapport' en 'Leerlingen'. Ondanks die kanttekeningen concludeert de Expertgroep dat de onderwijskundige en organisatorische aspecten voldoende zijn. Wat betreft de psychometrische aspecten neemt de Expertgroep de COTAN-beoordeling op normering, normering referentieniveaus en betrouwbaarheid over. Dit betekent dat de psychometrische aspecten van de toets als onvoldoende beoordeeld worden.

Op basis van voorgaande beoordelingen adviseert de Expertgroep de Dia-Eindtoets 2016 niet toe te laten en dit advies is door de staatssecretaris overgenomen. Over de twee eindtoetsen uit

2014, Route 8 van A-VISION en de IEP van Bureau ICE, is positief geadviseerd. Deze twee eindtoetsen zijn in 2015 opnieuw beoordeeld. De onderwijskundige/ organisatorische aspecten van zowel de IEP als ROUTE 8 zijn als voldoende beoordeeld. De psychometrische of COTAN-beoordeling uit 2015 van de IEP Eindtoets 2016 op voornoemde zeven criteria is: goed, goed, voldoende, onvoldoende, onvoldoende, onvoldoende, onvoldoende en het aspect 'normering referentieniveaus' is ook als onvoldoende beoordeeld. De COTAN-beoordeling is voor de Expertgroep geen reden om de IEP Eindtoets 2016 in te trekken. Wel wijst zij de toetsaanbieder op een aantal verbeterpunten, met name met betrekking tot de verslaglegging, waarmee deze bij de toekomstige jaarlijkse check van de toets rekening mee moet houden.

De COTAN-beoordeling uit 2015 van ROUTE 8 op voornoemde zeven criteria is: goed, voldoende, goed, onvoldoende, onvoldoende, onvoldoende, onvoldoende en de aspecten 'normering referentieniveaus' en 'computer adaptief toetsen' zijn ook als onvoldoende beoordeeld. Ook de COTAN-beoordeling van deze toets is voor de Expertgroep geen reden om de toets in te trekken en ook hier geeft zij de toetsaanbieder een aantal aanwijzingen voor verbetering mee

Voor 2016-2017 zijn de eindtoetsen die eerder afgewezen waren, de Dia-Eindtoets, de CESAN en de AMN Eindtoets, ook toegelaten. Het voorgaande betekent dat scholen nu naast de Centrale Eindtoets uit vijf andere eindtoetsen kunnen kiezen. Voor meer informatie over de jaarlijkse beoordelingen van de eindtoetsen, wordt verwezen naar www.expertgroepoetsenpo.nl > Adviezen en oordelen > Jaarlijkse beoordelingen Eindtoetsen.

4.3 Beoordelingskader LOVS-toetsen

Voor de LOVS-toetsen, zie de website van de Expertgroep, gelden de volgende onderwijskundige/ organisatorische en psychometrische kwaliteitscriteria:

- a dient inhoudelijk valide te zijn, betrouwbaar te zijn en een deugdelijke normering te hebben;
- b meet de kennis en vaardigheden van de leerling op het terrein van Nederlandse taal, rekenen en wiskunde, en/of de emotionele en verstandelijke ontwikkeling, de ontwikkeling van de creativiteit, de ontwikkeling in het verwerven van noodzakelijke kennis en/of sociale, culturele en lichamelijke vaardigheden;
- c de toetsresultaten worden zodanig weergegeven dat de leervorderingen van leerlingen voldoende inzichtelijk zijn voor ouders, voogden of verzorgers.

LOVS-toetsen worden beoordeeld met het 'Beoordelingskader voor (reeksen van) toetsen uit leerlingvolgsystemen (LOVS).' Dit beoordelingskader is ontwikkeld door een aantal psychometrische experts, onderwijsdeskundigen van de SLO en de leden van de Expertgroep. De beoordeling van de psychometrische aspecten richt zich op het genoemde in punt a. De beoordeling van de onderwijskundige aspecten richt zich ook op de inhoudelijke validiteit en het genoemde in de punten b en c.

Het beoordelingskader voor (reeksen van) toetsen uit leerlingvolgsystemen (LOVS) bestaat uit zes criteria die, vergelijkbaar met het COTAN beoordelingssysteem, 'bevroegd' worden en waarvan de vragen van een toelichting zijn voorzien. We beperken ons hier tot het vermelden van de criteria met bijbehorende vragen en een toelichting bij een aantal vragen. Voor het volledige beoordelingskader wordt verwezen naar de website van de Expertgroep, www.expertgroepoetsenpo.nl > Toetsingscriteria.

1 De kwaliteit van de dataverzameling

S1.1 Is de steekproef representatief?

S1.2 Is in geval van een onvolledig dataverzamelingsdesign het design adequaat?

Toelichting vraag S1.2:

Bij LOVS-toetsen worden data vaak verzameld in een onvolledig design waarbij niet alle leerlingen alle items maken. Vanwege het grote aantal items dat geanalyseerd moet worden maken leerlingen verschillende zogenaamde boekjes met items. Toepassing van modellen uit de itemresponsentheorie maakt het mogelijk om leerlingen en items op eenzelfde schaal af te beelden, zie de vaardigheidsschaal in figuur 5.1 in hoofdstuk 1. Daarvoor dient het onvolledig design wel voldoende verbonden te zijn wat wil zeggen dat er voldoende overlap bestaat tussen items en boekjes. Een eenvoudig voorbeeld is dat 200 leerlingen de items 1 t/m 40 maken en dat 200 andere leerlingen de items 26 t/m 65 maken. De overlap of het anker bestaat in dit geval uit 15 items en met dit aantal kan het doel om leerlingen en items op eenzelfde schaal af te beelden gerealiseerd worden. In het kader van deze vraag dient ook de passing van het gehanteerde itemresponsmodel aangetoond te worden.

2 Normering

N1.1 Absolute normen

N1.1.1 Is de standaardbepalingsmethode gemotiveerd en op de juiste wijze uitgevoerd?

N1.1.2 Zijn de beoordelaars/vakdeskundigen/experts naar behoren geselecteerd en getraind?

N1.1.3 Is er voldoende overeenstemming tussen de beoordelaars?

N1.2 Relatieve normen

N1.2.1 Zijn de normgroepen groot genoeg?

N1.2.2 Zijn de normgroepen representatief?

3 Betrouwbaarheid

B1.1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

B1.2 Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

Toelichting vraag B1.2:

Bij LOVS-toetsen worden leerlingen vaak gecategoriseerd. Bij de te verstrekken betrouwbaarheidsgegevens hoort dan ook het percentage verwachte misclassificaties. Dit betekent dat als de leerlingen gecategoriseerd worden in categorieën A, B, C, D en E (of 1, 2, 3, 4 en 5), aangegeven moet worden wat de kans is dat ze ook werkelijk in de genoemde categorie zitten en niet in een andere categorie.

4 Validiteit

V1.1 Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items).

5 Het volg-aspect

Va1.1 Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Va1.2 Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden? Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Toelichting vraag Va1.1:

Er dient betekenisvolle informatie verstrekt te worden over de samenhang tussen twee of meer tijdstippen/meetmomenten. Essentieel is dat de schaal waarop de groei wordt weergegeven grondig is onderbouwd.

Toelichting vraag Va1.2:

De handleiding moet een beschrijving bevatten van hoe de gebruiker (leerkrachten, ouders e.a.) de gegevens met betrekking tot de (stagnatie van de) groei van een leerling inhoudelijk en relatief ten opzichte van een referentiepopulatie dient te interpreteren. Daarbij is het ook van belang om een indicatie van de betrouwbaarheid van die interpretatie te geven. Maakt men gebruik van met een itemresponsmodel geschatte vaardigheidsscores dan dient de schattingsfout van het verschil in vaardigheid op twee tijdstippen vermeld te worden. Maakt men gebruik van percentielscores dan dient de schattingsfout van het verschil in percentielscores op twee tijdstippen vermeld te worden.

6 Inzicht in de leervorderingen

- 11.1 Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders, verzorgers, voogden en docenten begrijpelijk is?

4.3.1 Voorbeeld van beoordeling LOVS-toets

Op de website van de Expertgroep, www.expertgroepoetsenpo.nl > Kwaliteitsoordelen Lovs-Toetsen, staan kwaliteitsoordelen en uitgebreide beoordelingen van een groot aantal toetsen van Cito en een toets van uitgeverij Boom. Het voorbeeld hier heeft betrekking op de toets voor groep 4 uit de reeks toetsen van Cito Begrijpend Luisteren groep 3 tot en met 8.

De toets Begrijpend Luisteren groep 4 is een onderdeel van het Cito Volgsysteem primair en speciaal onderwijs en is bedoeld voor leerlingen in groep 4 van het primair onderwijs. De toets vormt samen met de LOVS toetsen Begrijpend Luisteren voor groep 3 tot en met 8 een systeem om vast te stellen hoe leerlingen met begrip kunnen luisteren en hoe hun luistervaardigheid zich in de basisschoolperiode ontwikkelt. Doel van de toets Begrijpend Luisteren groep 4 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van begrijpend luisteren. Hiervoor wordt de door de leerling behaalde vaardigheidsscore normgericht geïnterpreteerd op basis van de vaardigheidsverdeling in een landelijke referentiegroep. De vaardigheidsscore wordt uitgedrukt in de symmetrische niveau-indeling I t/m V en in de asymmetrische niveau-indeling A t/m E. De toets 4 bestaat uit 32 opgaven en wordt M(edio)4, halverwege groep 4, en E(ind)4, eind groep 4, afgenomen. De toets 4 maakt het mogelijk om:

- De vaardigheid begrijpend luisteren van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) te beoordelen via een vergelijking van de behaalde scores met de scores van een landelijke referentiegroep. We spreken hier van niveaubepaling.
- De ontwikkeling van de vaardigheid begrijpend luisteren van zowel individuele leerlingen als groepen leerlingen te beoordelen. We spreken hier van progressiebepaling.

Hieronder worden de bevindingen en oordelen van de door de Expertgroep ingezette deskundigen, ontleend aan de uitgebreide toelichting, op de zes criteria kort samengevat.

1 De kwaliteit van de dataverzameling

- S1.1 Is de steekproef representatief?

De steekproef is representatief met betrekking tot regio, urbanisatiegraad, schooltype en geslacht.

Oordeel: voldoende.

- S1.2 Is in geval van een onvolledig dataverzamelingsdesign het design adequaat?

Voor het proefonderzoek is een adequaat onvolledig design gebruikt. Voor de toets M4/E4 is een adequaat volledig design gebruikt. De data die met beide designs verzameld zijn, zijn met het itemresponsmodel OPLM geanalyseerd. Het blijkt dat zowel de individuele items als de

gehele toets bij het model passen. Dit betekent dat er sprake is van een vaardigheidsschaal waar zowel items als leerlingen op afgebeeld kunnen worden.

Oordeel: voldoende.

2 Normering

N1.2 Relatieve normen

N1.2.1 Zijn de normgroepen groot genoeg?

Voor M4 zijn de toetsresultaten van 1785 leerlingen en voor E4 van 1961 leerlingen gebruikt. Voor beide momenten (M4 en E4) worden vaardigheidsverdelingen gepresenteerd. Dit betreft de gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75, P80 en P90. Van hieruit kunnen de beide niveau-indelingen, A t/m E en I t/m V, worden bepaald.

Oordeel voldoende.

N1.2.2 Zijn de normgroepen representatief?

De representativiteit is bij vraag S1.1 al voldoende bevonden.

Oordeel: voldoende.

3 Betrouwbaarheid

B1.1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

De betreffende coëfficiënt MAcc (Measurement Accuracy coëfficiënt) als indicator voor de betrouwbaarheid van de toets is correct berekend.

Oordeel: voldoende.

B1.2 Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

De klassieke betrouwbaarheidscoëfficiënt van M4 en E4 bedraagt meer dan 0,70 wat volgens de criteria van de COTAN voldoende is als de toets geen zware consequenties voor de leerlingen heeft. Daarnaast zijn indices voor de nauwkeurigheid van de classificaties voor de twee niveau-indelingen berekend. Uit de hoogte van de indices blijkt dat de laagst en hoogst scorende leerlingen accuraat te classificeren zijn maar dat tussen leerlingen in de niveaugroepen B, C en D, respectievelijk II, III en IV, minder duidelijk onderscheid te maken is.

Oordeel: voldoende.

4 Validiteit

V1.1 Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items).

Naar de validiteit van de toetsen is inhoudelijk en psychometrisch onderzoek verricht. Hoewel het inhoudelijk onderzoek in het algemeen positief is, wordt wel opgemerkt dat als gevolg van het gebruik van meerkeuzevragen niet alle aspecten van begrijpend luisteren aan de orde komen. Uit de resultaten van het psychometrisch onderzoek mag men afleiden dat de items eenzelfde vaardigheid, opgevat als 'begrijpend luisteren', meten. De samenhang tussen de afzonderlijke items en de totaalscore is hoog. De correlaties met toetsen met dezelfde meetpretentie is duidelijk hoger dan met toetsen met een andere meetpretentie. Er is geen sprake van differentieel itemfunctioneren (dif) met betrekking tot sekse wat betekent dat de items bij de jongens hetzelfde meten als bij de meisjes. Wel behalen de meisjes hogere scores dan de jongens.

Oordeel: voldoende.

5 Het volg-aspect

Va1.1 Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

De resultaten van de psychometrische analyse laten zien dat de items van de toets op een eendimensionale vaardigheidsschaal afgebeeld kunnen worden. Dit betekent dat afhankelijk van het aantal items dat een leerling goed maakt aan hem/haar een bepaalde vaardigheids-

score toegekend kan worden. Tussen M4 en E4 is er sprake van een relatief grote groei terwijl er in de leerjaren 5 en 6 sprake is van een bescheiden groei.

Oordeel: voldoende

Va1.2 Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden?

Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Het blijkt dat de gemiddelde toename van groei steeds aanmerkelijk kleiner is dan de spreiding in vaardigheid binnen de groep op enig afnamemoment. Dat impliceert dat het meerdere keren vaststellen en in die zin volgen van leerlingen weinig zin heeft en dat dus met een enkel afnamemoment per jaar volstaan kan worden. In de wetenschappelijke verantwoording wordt aangegeven welke vergelijkingen gemaakt kunnen worden om leerlingen in hun ontwikkeling te volgen. Voor alle vergelijkingen geldt dat uitspraken over de ontwikkeling van leerlingen gerelativeerd moeten worden vanwege de (on)betrouwbaarheid van de toetsen. Door betrokkenen bij de toetsen Begrijpend Luisteren moet beseft worden dat vaardigheidsgroei zich langzaam in de tijd voltrekt.

Oordeel: voldoende.

6 Inzicht in de leervorderingen

11.1 Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders, verzorgers, voogden en docenten begrijpelijk is?

In de handleiding zijn registratieformulieren opgenomen voor een leerlingrapport en een groepsrapport die door de leerkracht handmatig of digitaal kunnen worden ingevuld. De wijze waarop de formulieren zijn vormgegeven met de uitleg erbij geven inzichtelijk het niveau en de groei weer. Ook wordt in de handleiding beschreven hoe er met de ouders over de toetsresultaten gecommuniceerd kan worden.

Oordeel: voldoende.

Het kwaliteitsoordeel van de Expertgroep over deze toets luidt als volgt:

‘Alles overwegende en gezien de beoordelingen van de inhoudelijk deskundigen, oordeelt de Expertgroep dat de beoordeelde toets Begrijpend Luisteren groep 4 van voldoende kwaliteit is om in het primair onderwijs te worden gebruikt voor het systematisch meten van de leervorderingen van de leerlingen.’

4.4 Conclusies en discussie

Op basis van voorgaande bespreking van beoordelingssystemen die in het primair onderwijs gehanteerd worden, kunnen we concluderen dat in geen enkele sector van het Nederlandse onderwijs de kwaliteit van toetsen zo grondig gecontroleerd wordt als in het primair onderwijs. De bespreking laat ook zien dat die controle succes heeft. Van de eindtoetsen en LOVS-toetsen die door de COTAN en de Expertgroep positief beoordeeld zijn, mag de gebruiker aannemen dat de toets meet wat deze beoogt te meten (validiteit) en dat men staat kan maken op de resultaten die de toets oplevert (betrouwbaarheid).

Scholen kunnen nu kiezen uit zes eindtoetsen. Het zal moeten blijken of het primair onderwijs wel behoefte heeft aan een dergelijk aanbod van eindtoetsen.

Op de website van de Expertgroep staat ook dat de staatssecretaris de Expertgroep gevraagd heeft de vergelijkbaarheid van de eindtoetsen te onderzoeken. De vraag is wat het betekent als uit dat onderzoek mocht blijken dat verschillende eindtoetsen tot verschillende schooladviezen voor schoolkeuze leiden? Volgens de website van de Expertgroep zal in 2017 over dit onderzoek gerapporteerd worden.

In dit hoofdstuk zagen we dat er verschillende beoordelingssystemen in gebruik zijn. Het COTAN beoordelingssysteem is oorspronkelijk ontwikkeld voor de beoordeling van psychologische tests en blijkt dan ook minder geschikt te zijn om de toenemende verscheidenheid van studietoetsen te beoordelen. Vandaar dat hier gepleit wordt om in plaats van het COTAN beoordelingssysteem het RCEC beoordelingssysteem (Sanders e.a., 2016) in te zetten dat speciaal ontwikkeld is voor de beoordeling van huidige en nog nieuw te ontwikkelen studietoetsen en examens.

5 Gebruikers van toetsresultaten

5 Gebruikers van toetsresultaten

Judith Hollenberg en Marleen van der Lubbe

Hoewel een school verantwoordelijk is voor het verzamelen van toetsresultaten, is de school niet de enige gebruiker van de resultaten. Ouders willen weten wat de toetsresultaten van hun kind zijn. Leerlingen willen vaak ook zelf graag weten hoe zij gepresteerd hebben. In paragraaf 5.1. staat beschreven op welke wijze richting ouders en leerlingen over toetsresultaten gecommuniceerd kan worden. In paragraaf 5.2 wordt aangegeven op welke manieren de school toetsresultaten kan gebruiken, terwijl in paragraaf 5.3 beschreven wordt hoe schoolbesturen onderwijsresultaten kunnen benutten. Tot slot staat in paragraaf 5.4 op welke wijze de Inspectie van het onderwijs en indicatiecommissies toetsresultaten gebruiken.

5.1 Ouders en leerlingen

Toetsresultaten worden op de meeste scholen gecommuniceerd naar de ouders, en op veel scholen ook richting de leerlingen. Ouders willen uiteraard weten hoe hun kind ervoor staat. Ook leerlingen willen na een toets graag weten hoe zij de toets hebben gemaakt. De leerlingen krijgen daarom een paar keer per jaar een rapport mee en als aanvulling daarop zijn er voor ouders vaak 10-minutengesprekken.

Toetsresultaten, zowel van methodegebonden als methodeonafhankelijke toetsen, zijn een prima middel om ouders een goed beeld te geven van hoe hun zoon of dochter functioneert op verschillende vakgebieden. De toetsresultaten moeten dan echter wel vergezeld gaan van een duidelijke toelichting. Onderstaande voorbeelden uit de praktijk zijn veelvoorkomende misverstanden rondom de interpretatie van LVS-toetsen die duidelijk maken hoe belangrijk een goede voorlichting over de interpretatie van de toetsresultaten is.

Voorbeeld 1: *Martijn heeft bij de M3-toets van Rekenen-Wiskunde een IV-score behaald. Op de E3-toets, de M4-toets en de E4-toets heeft hij eveneens een IV-score behaald. De leerkracht heeft toegelicht dat Martijn binnen zijn niveau een goede groei doormaakt. Wanneer de leerkracht in groep 5 spreekt over Martijn als een ‘zwakke rekenaar’ zijn de ouders verbaasd. Hij heeft toch immers een goede groei doorgemaakt?*

Dit voorbeeld maakt duidelijk dat ouders moeten weten dat bij toetsresultaten zowel gekeken moet worden naar het (huidige) niveau van een leerling als naar de groei die een leerling heeft doorgemaakt. In het geval van Martijn is er sprake van een goede groei en gaan de ouders er vanuit dat een goede groei automatisch een goed niveau betekent. De leerkracht dient daarom duidelijk uit te leggen dat groei en niveau twee verschillende zaken zijn en dat een leerling die voldoende groeit wellicht niet direct op specifieke gebieden extra ondersteuning behoeft maar wel over de gehele linie van het betreffende vakgebied een zwakke leerling kan zijn.

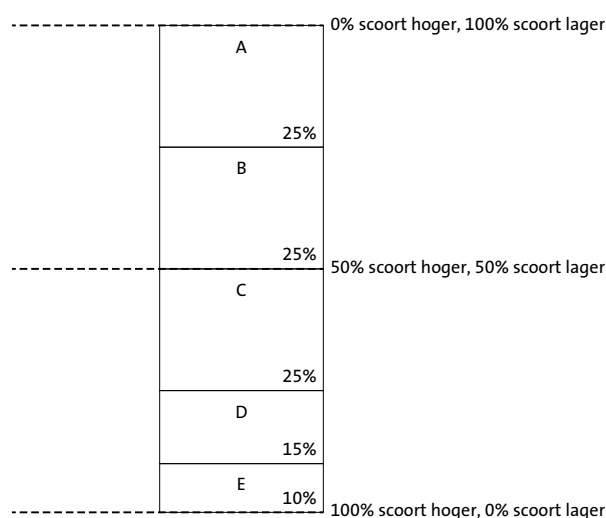
Voorbeeld 2: Farid heeft bij Spelling een II-score gehaald. Bij de volgende toets haalt hij wederom een II-score. De ouders maken zich zorgen: heeft hij dan niets geleerd?

Het geval van Farid laat zien dat de betekenis van de niveaus I t/m V of A t/m E om een duidelijke toelichting naar de ouders vraagt. De niveaus zijn een handig hulpmiddel om leerlingen te vergelijken met zijn leeftijdsgenoten. In figuur 5.1 staat een voorbeeld van een toelichting op de scores A t/m E voor ouders. Dit kan bijvoorbeeld bij een rapport toegevoegd worden. Daarnaast is het aan te bevelen om de betekenis van toetsscores tijdens een ouderavond of 10-minutengesprek toe te lichten om verwarring te voorkomen.

Bij uw kind is een volgtoets van Cito afgenomen. Uit deze toets komt een getal, ook wel vaardigheidsscore genoemd. U kunt aan deze score niet meteen zien of uw kind op, onder of boven het gemiddelde heeft gescoord. Daarom wordt er ook een niveau bij de score vermeld en dit is opgenomen op het rapport van uw kind. Dit niveau (A, B, C, D of E) geeft aan hoe uw kind scoort in vergelijking met leeftijdgenoten.

‘A’ betekent dat uw kind landelijk gezien behoort tot de 25% hoogst scorende kinderen.

‘B’ betekent dat ten minste 25% van de kinderen hoger scoort dan uw kind (namelijk de kinderen die op A-niveau scoren) en 50% van de kinderen een lagere score haalt (de kinderen die op C, D en E-niveau scoren). Als uw kind niveau A of B behaalt, is het niveau dus boven het gemiddelde. Zie ook in onderstaande figuur.



Bij het kijken naar toetsscores, is het belangrijk om niet alleen maar naar het niveau A t/m E te kijken. Dat zegt alleen iets over de score van uw kind op dit moment en in vergelijking met leeftijdgenoten. Het is belangrijk om daarnaast ook te kijken naar de groei of ontwikkeling van uw kind door de score van nu te vergelijken met de score op de vorige momenten. Wanneer een kind een achterstand heeft ten opzichte van zijn leeftijdgenoten, maar deze achterstand tijdens de vorige afname nog veel groter was en nu aanzienlijk is verminderd, kunt u tevreden zijn. Vergelijk dit met de groeicurve op het consultatiebureau: Indien uw kind relatief klein is, bent u tevreden wanneer uw kind goed gegroeid is en zijn groeiachterstand enigszins inloopt, ook al is uw kind nog steeds kleiner dan gemiddeld. Net zoals groeicurves van lengte op het consultatiebureau, zijn er op school groeicurves met daarin de toetsscores.

Figuur 5.1: Voorbeeld toelichting niveau A t/m E

Gebruik van het (alternatief) leerlingrapport

Een handig hulpmiddel om zowel de groei als het niveau voor ouders inzichtelijk te maken, en daarnaast de scores I t/m V of A t/m E te kunnen toelichten, is het (alternatief) leerlingrapport. Bijvoorbeeld door de minigrafieken van verschillende leergebieden op een pagina te printen, kan een school aan de ouders in een oogopslag laten zien hoe een leerling er op verschillende leergebieden voor staat.

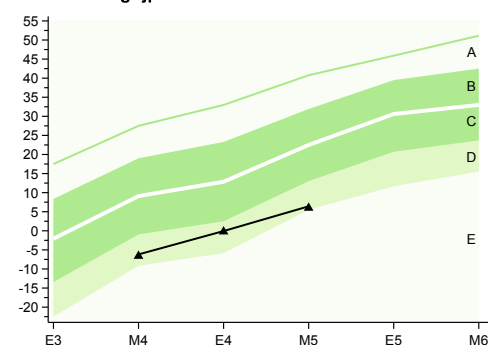


Leerlingrapport

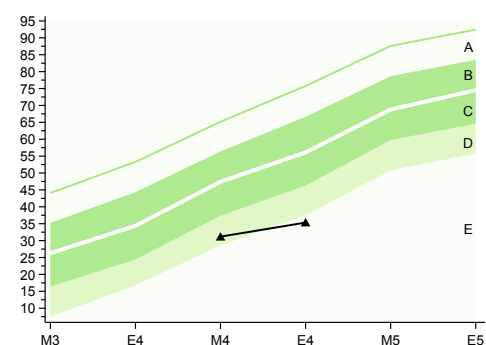
Leerling: **Selma Azizallah (5 - 5A)**

Vergelijking: **Alle leerlingen**

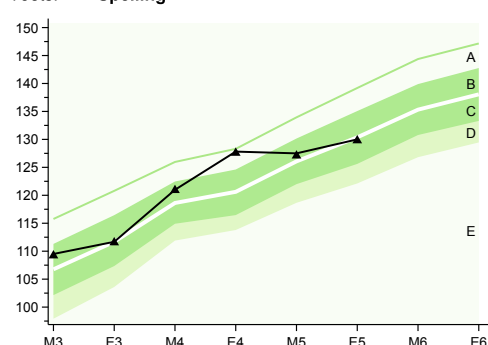
Toets: **Begrijpend lezen**



Toets: **Rek-Wisk**



Toets: **Spelling**



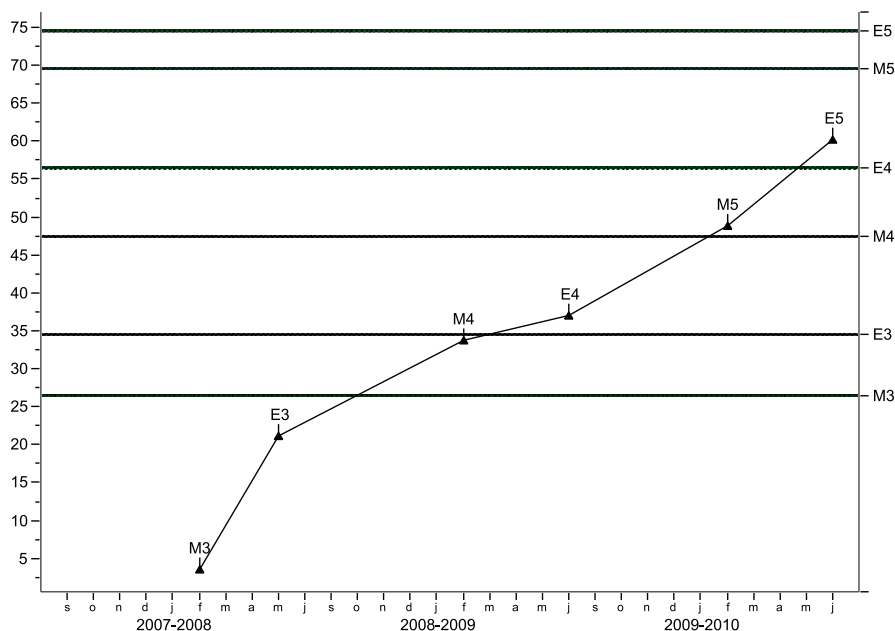
Figuur 5.2: Regulier leerlingrapport voor ouders

- **Kijken naar niveau:** In het reguliere leerlingrapport (figuur 5.2) wijst de leerkracht op de indeling I-V of A-E en laat zo zien hoe de score van de leerling zich verhoudt tot de score van leeftijdgenoten. Wat betreft Begrijpend lezen heeft Selma bijvoorbeeld een D-score, wat betekent dat zij tot de 25% leerlingen met de laagste score behoort.
- **Kijken naar groei:** In het leerlingrapport is door de curve die – als het goed is – omhoog loopt, te zien dat de vaardigheid van de leerling toeneemt. Door te wijzen op de groei binnen een niveau - hoe verloopt de gemiddelde groei binnen bijvoorbeeld niveau C? - kan de leerkracht tonen hoe de groei van de leerling zich verhoudt ten opzichte van het gemiddelde. Hierbij kan toegelicht worden dat wanneer een leerling van een 'hoge B' naar een 'lage B' gaat, de groei lager is dan verwacht en dat er wat moet gebeuren om deze trend niet door te laten zetten. In het geval van Selma is zichtbaar dat bij Spelling de score aan het afzakken is van een B- naar een C-score.

Alternatief leerlingrapport

Leerling: **Dyllon Jansen (5 - 5A)**
Toets: **Rek-Wisk**

Vergelijking: **Alle leerlingen**



Afname datum	Leeftijd jr-mnd	Jaar groep	Taak	Toetsscore / Vaard.score	Score interval	Func. niveau
10-02-2008	6-10	3	M3	17 / 4	0:8	< M3
27-05-2008	7-01	3	E3	24 / 21	17:25	< M3
15-02-2009	7-10	4	M4	22 / 34	30:37	E3
01-06-2009	8-02	4	E4	19 / 37	33:41	E3
01-02-2010	8-10	5	M5	20 / 49	45:53	M4
21-06-2010	9-02	5	E5	34 / 60	57:63	E4/M5

Figuur 5.3: Alternatief leerlingrapport

Kijken naar groei en niveau voor leerlingen die werken op hun eigen niveau, kan met behulp van het alternatief leerlingrapport. Voor een voorbeeld zie figuur 5.3. Ook van het alternatief leerlingrapport kunnen minigrafieken worden uitgedraaid om zo alle leergebieden aan ouders te kunnen presenteren. Wanneer een leerling een D- of E-score heeft, zegt ouders dit vaak niet zoveel. Door te laten zien op welk niveau de leerling functioneert, middels het functioneringsniveau wordt de grootte van de achterstand van de leerling duidelijker. Zo functioneert Dyllon in juni groep 5 op het niveau E4/M5 bij Rekenen-Wiskunde. Dit maakt duidelijk dat hij een kleine achterstand heeft. Dit kan voor ouders confronterend zijn, maar voorkomt onaangename verrassingen wanneer in groep 8 een schoolkeuzeadvies wordt gegeven. Met het alternatief leerlingrapport wordt zichtbaar dat de leerling een achterstand heeft en hoe groot deze ongeveer is door het functioneringsniveau te vergelijken met de groep waarin de leerling zit. Ook is de groei zichtbaar door de opgaande lijn. Daarnaast is duidelijk te illustreren hoe de groei zich verhoudt tot het landelijk gemiddelde. Een leerling die van functioneringsniveau E3 in midden groep 4, naar functioneringsniveau E4 midden groep 5 gaat, heeft weliswaar een achterstand, maar maakt wel een gemiddelde groei door. De achterstand neemt dus niet toe. Ook is het mogelijk dat een leerling een beperktere groei doormaakt wanneer bijvoorbeeld in 2 jaar tijd het functioneringsniveau met 1 jaar is toegenomen. Dit is duidelijk zichtbaar voor ouders door het alternatief leerlingrapport met hen te bespreken.

Voor leerlingen waarvoor een ontwikkelingsperspectief (zie hoofdstuk 3) is vastgesteld, kan de school eventueel tijdens een oudergesprek in het alternatief leerlingrapport het verwachte perspectief intekenen. Op deze wijze kan een leerkracht niet alleen toelichten hoe de groei van de leerling zich verhoudt tot de gemiddelde leerling, maar kan een leerkracht ook duidelijk maken hoe de groei van de leerling verloopt ten opzichte van het gestelde ontwikkelingsperspectief. Op deze wijze weten ouders tijdig welk niveau hun leerling aan het einde van het basisonderwijs ongeveer zal behalen.

Toetsresultaten versus rapport

Wat vormt de basis voor het rapport dat een leerling meerdere malen per jaar mee naar huis krijgt? De ene school kiest ervoor om gegevens op methodeonafhankelijke toetsen te rapporteren zoals in figuur 5.4, de andere school doet dit niet.

Groep 3		Naam leerling: Jorik Frederiks	
Leerkracht: Juf Sandra		Datum: 18-6-2010	Schooljaar: 2009-2010
		rapport 3	Opmerkingen
Spelling			
Algemeen		O	
Cito spelling		D	
Taal			
Methodetoets		O/V	
Mondeling taalgebruik		G	
Woord- en zinsbouw		V	
Lezen			
Letters		O	Rick heeft M3 instructieniveau gehaald
Woorden		O	
Cito begrijpend lezen		D	
Rekenen			
Optellen/Aftrekken		V	
Inzicht		O	
Cito rekenen		C	
Schrijven			
Algemeen		V	
Wereldoriëntatie			
Belangstelling		G	

Figuur 5.4: Voorbeeld van een rapport met resultaten van LVS-toetsen (O=Onvoldoende, V=Voldoende, G=Goed)

Zowel methodegebonden toetsen, methodeonafhankelijke toetsen, als motivatie/inzet kunnen als input gebruikt worden. Onderstaand voorbeeld uit de praktijk maakt duidelijk dat ook een goede communicatie over hoe een rapport tot stand komt van belang is.

Voorbeeld 3: Mariëlle behaalt bij methodeonafhankelijke toetsen meestal een III-score. Methodegebonden toetsen maakt ze meestal prima: hierop heeft ze hooguit enkele fouten. Op het rapport staan de cijfers op basis van de methodegebonden toetsen. Wanneer de ouders in de bovenbouw de toetsresultaten van de LVS-toetsen zien, schrikken ze. Ze waren er vanuit gegaan dat hun dochter tot de betere leerlingen behoort, nu blijkt ze een 'gemiddelde leerling' te zijn.

In veel gevallen worden de resultaten op methodegebonden toetsen gebruikt als input voor een rapport. Op veel scholen zijn afspraken gemaakt, bijvoorbeeld dat elke fout bij een dictee tot een halve punt aftrek van een 10 leidt. Drie fouten betekent dan een 8,5. Op deze wijze wordt per dictee een cijfer gegeven en wordt het gemiddelde berekend om tot een cijfer voor op het rapport te komen. Een aandachtspunt hierbij is dat veel methodegebonden toetsen zodanig zijn geconstrueerd dat een groot deel van de leerlingen deze (bijna) foutloos zou moeten kunnen maken. Het zijn immers beheersingstoetsen die nagaan of de in de afgelopen periode aangeboden stof beheerst wordt. Een groot deel van de leerlingen zal dus naar verwachting een goed of het cijfer 10 op de toets halen, dus zowel de leerlingen die de leerstof net beheersen, als ook de leerlingen die wat extra's aankunnen. Een 10 betekent dus dat de leerling de leerstof prima beheerst, maar niet automatisch dat het een excellente leerling betreft die uiteindelijk bijvoorbeeld naar het vwo zal doorstromen. Ouders zullen deze conclusie vaak echter niet zelf trekken waardoor dit op een later moment, bijvoorbeeld bij het schoolkeuzeadvies, tot onaangename verrassingen leidt.

Methodeonafhankelijke toetsen kunnen als aanvulling op de methodegebonden toetsen worden gebruikt. Methodeonafhankelijke toetsen maken immers wel onderscheid tussen leerlingen die de leerstof net beheersen en leerlingen die wat extra's aankunnen. Wanneer een leerling de methodegebonden toetsen prima maakt, maar een III-score heeft op de LVS-toets, maakt dit duidelijk dat de leerling weliswaar goed mee kan komen, maar dat de leerling voor dit vakgebied een gemiddelde leerling is en uiteindelijk waarschijnlijk niet door zal stromen naar het gymnasium.

Resultaten op methodeonafhankelijke toetsen zijn niet bedoeld om rechtstreeks te vertalen naar een cijfer of de aanduiding voldoende/onvoldoende. Een IV-score is weliswaar benedengemiddeld, maar dit betekent niet automatisch dat een leerling een 'zwak' of 'onvoldoende' op zijn rapport verdient. Opties om resultaten op de volgtoetsen naar ouders te communiceren zijn bijvoorbeeld:

- De resultaten op methodeonafhankelijke toetsen in een oudergesprek toelichten aan de hand van het leerlingrapport. Zo kan de school duidelijk maken dat gekeken moet worden naar zowel het niveau als de groei van het kind.
- Bij het rapport op een invoegblad de resultaten op de methodeonafhankelijke toetsen vermelden. De leerling kan trots met zijn rapport, met of zonder zonder invoegblad, naar opa en oma. Tegelijkertijd zijn de ouders zijn wel op de hoogte van hoe hun zoon/dochter functioneert ten opzichte van medeleerlingen.
- Combineren van resultaten op methodegebonden toetsen en methodeonafhankelijke toetsen door, zoals in figuur 5.4, deze beide op het rapport te vermelden waarbij bij de volgtoetsen de niveau-aanduiding I t/m V of A t/m E wordt vermeld. De aanduiding goed-voldoende-onvoldoende is gebaseerd op de scores op de methodegebonden toetsen.
- Op het rapport het niveau I t/m V of A t/m E op de volgtoetsen vermelden. Daarnaast door middel van een aanduiding goed-voldoende-onvoldoende een waardering geven aan de groei die een leerling ten opzichte van het vorige toetsmoment heeft doorgemaakt. Een leerling met een D-score die een vorige keer een E-score had, kan zo de aanduiding 'G' (goed) verdienen.

Voorbeeld 4: Marvin behaalt bij de veel toetsen een V- of een IV-score. Op zijn rapporten geven leerkrachten hem wegens zijn goede inzet voor de meeste vakken al jaren een voldoende of hoger. Wanneer in groep 8 blijkt dat Marvin geen havo/vwo-advies krijgt, reageren de ouders boos. Zijn rapporten waren toch altijd goed?

Motivatie kan worden meegewogen bij het bepalen van een cijfer. Wanneer een leerling met dyslexie ontzettend zijn best doet, werkt het weinig motiverend wanneer hij na deze enorme inspanning een rapport krijgt met daarop een onvoldoende voor Spelling en Technisch lezen. Om deze reden geven scholen bij leerlingen met leerproblemen dikwijls een cijfer waarin motivatie wordt meegenomen. Let er wel op om duidelijk te vermelden waar de beoordeling op gebaseerd is om verwarring zoals in het voorbeeld van Marvin te voorkomen.

Een aanvulling op bovengenoemde opties om resultaten op methodeonafhankelijke toetsen te communiceren naar ouders, is daarom de volgende:

- Op het rapport zowel het niveau I t/m V of A t/m E als een score gebaseerd op groei en/of methodegebonden toetsen – goed, voldoende, onvoldoende – vermelden. Dit kan aangevuld worden met een aparte beoordeling van de inzet – goed, voldoende, onvoldoende – van de leerling. Bij deze drie beoordelingen – niveau, groei, inzet – dient duidelijk vermeld te worden welk aspect wordt beoordeeld. Dit voorkomt de verwarring dat een ‘goed’ op basis van inzet bij de ouders de verwachting schept dat het niveau of de groei van de leerling ‘goed’ is.

Let op dat sommige scholen op het rapport cijfers geven gerelateerd aan ‘het niveau van de leerling’. Denk aan formuleringen als ‘Op zijn eigen niveau scoort deze leerling een 7,5 voor Begrijpend lezen’. Deze formulering is feitelijk onjuist: op zijn eigen niveau scoort een leerling altijd goed, anders was het immers niet zijn eigen niveau. Oplossingen om de leerling een beoordeling te geven die past bij zijn inzet en vermogens zijn bijvoorbeeld om duidelijk te vermelden dat het cijfer mede tot stand komt op basis van de motivatie van de leerling. Een andere oplossing is om in de bovenbouw bij het geven van beoordelingen te differentiëren naar het verwachte schooltype VO, ook hier weer met een duidelijke vermelding dat het oordeel gekoppeld is aan het specifieke schooltype. Een leerling die naar verwachting naar vmbo-t gaat, krijgt dan met hetzelfde aantal goede antwoorden op zijn rapport een hogere beoordeling dan een vwo-leerling. De beoordeling zal passen bij het niveau van de leerling, maar is toch anders dan in het voorbeeld ‘Op zijn eigen niveau scoort deze leerling een 7,5 voor Begrijpend lezen’ doordat bij ‘het niveau van een vmbo-t-leerling’ mensen eenzelfde beeld kunnen vormen over wat een leerling beheerst en het bij ‘een eigen niveau’ onduidelijk blijft wat dat niveau precies inhoudt.

Welke wijze van beoordelen een school ook kiest, er moet voor gezorgd worden dat de afspraken/regels duidelijk vastliggen. Dus wanneer krijgt een leerling een ‘ruim voldoende’ en welke regels gelden er voor beoordelingen gekoppeld aan VO-schooltypen? Hierdoor wordt voorkomen dat een leerling bij vergelijkbare prestaties, bij de overgang van de ene naar de andere leerkracht opeens een slechter of beter rapport heeft.

Communicatie naar leerlingen

Een rapport wordt niet alleen bekeken door ouders, maar ook door de leerlingen zelf. Bij het bespreken van het rapport en/of toetsresultaten met leerlingen, is het goed om te focussen op inzet en eventueel op groei. Voor jongere leerlingen is een focus op inzet voldoende. Voor oudere leerlingen kan het daarnaast motiverend zijn om, met een hele duidelijke toelichting, te laten zien hoe zij zich ontwikkelen. Leerlingen met een I-score die weinig inzet hebben ‘omdat ze het toch al kunnen’, kan duidelijk gemaakt worden dat ze hoog scoren, maar misschien minder dan gemiddeld groeien. Leerlingen met een V-score die veel faalervaringen hebben, kan duidelijk gemaakt worden dat ze een goede groei doormaken en dus prima op deze weg door kunnen gaan. Daarbij kan eventueel de koppeling worden gelegd met de opgestelde leerrendementsverwachting zoals opgenomen in het ontwikkelingsperspectief.

De wijze waarop toetsresultaten met leerlingen besproken kunnen worden, is sterk afhankelijk van zowel de leeftijd als van de sociaal-emotionele ontwikkeling van een leerling. Een faalangstige leerling vraagt om een andere benadering dan een leerling die zeker van zichzelf is. Een leerkracht kan het beste beoordelen wat een goede wijze is om toetsresultaten met

leerlingen te bespreken. Het is aan te bevelen om, net als over de wijze waarop beoordelingen tot stand komen, ook over de communicatie van toetsresultaten naar leerlingen op school-niveau afspraken te maken. Dit betekent niet dat de wijze van communicatie voor alle leerlingen hetzelfde hoeft te zijn, maar wel dat een leerling bij verschillende leerkrachten op ongeveer dezelfde wijze informatie over toetsresultaten krijgt. Wanneer de leerkracht in groep 5 focust op groei, terwijl de leerkracht in groep 6 meer nadruk legt op het niveau, kan dit voor leerlingen immers verwarrend werken.

5.2 Schoolteam en schooldirectie

Op schoolniveau: Doelen stellen en evalueren

Op een gemiddelde school heeft 20% van de leerlingen bij Rekenen-Wiskunde een I-score, heeft 20% een II-score etcetera. Midden groep 5 is de gemiddelde vaardigheidsscore bij Spelling 126,2. Het voorgaande betekent niet dat een school automatisch moet streven naar 20% I-leerlingen of naar een vaardigheidsscore 126,2 bij M5-Spelling. Een school in een achterstandswijk met veel allochtone leerlingen mag meer dan tevreden zijn wanneer 20% van de leerlingen een I-score behaalt. Een school met veel leerlingen van hoog opgeleide ouders en een bovengemiddelde intelligentie kan hogere eisen stellen, denk aan 50% leerlingen met een I- of II-score. Hoge verwachtingen van leerkrachten gaan samen met hogere resultaten en het is daarom aan te bevelen om ambitieuze doelen te stellen. Deze doelen moeten echter wel realistisch zijn. Dit kan door beredeneerde keuzes te maken op basis van de schoolpopulatie en op basis van de gegevens van de afgelopen schooljaren. Dwarsdoorsneden en trendanalyses (zie hoofdstuk 1) zijn een goed hulpmiddel om een schooleigen trend uit af te leiden. Als bijvoorbeeld over een periode van bijvoorbeeld 3-4 jaar zichtbaar is dat de jaargroepen 3 ruim boven het landelijk gemiddelde scores, kan als doelstelling een score op dat niveau worden geformuleerd. Wanneer besloten wordt verbeteracties in te zetten, kan als doelstelling een score boven het huidige niveau gekozen worden. Na het uitvoeren van de verbeteractie is in de trendanalyse zichtbaar of de actie tot het gewenste resultaat heeft geleid. De trendanalyse is zo een goede basis voor het werken aan kwaliteitsverbetering.

Het is raadzaam om doelen niet alleen in termen van gemiddelde vaardigheidsscores te formuleren, maar ook als het percentage leerlingen dat bijvoorbeeld een A/I-score haalt. Hierdoor is in een oogopslag duidelijk hoe de ambitie van de school zich verhoudt tot het landelijk gemiddelde, en hoe de resultaten/ambities van de verschillende groepen zich tot elkaar verhouden. Zie voor meer informatie de kwaliteitskaart 'Doelen stellen' via www.schoolaanzet.nl > thematische inhoud, zoeken op 'opbrengstgericht werken, gebruik data'.

Op schoolniveau: Verbeteracties kiezen en evalueren

Schoolspecifieke doelen vormen de basis voor de keuze van verbeteracties. Voldoen de resultaten op een leergebied niet aan de ambities van de school, dan kan dit betekenen dat de ambities te hoog gesteld zijn. In veel gevallen kunnen de doelen echter wel behaald worden met behulp van een verbeteractie. Wanneer bijvoorbeeld blijkt dat de resultaten bij Woordenschat lager zijn dan verwacht en gewenst, kan besloten worden om meer tijd in te ruimen voor het geven van onderwijs in Woordenschat. Indien de resultaten bij groep 5 bij verschillende leergebieden achterblijven, kan gekeken worden welke extra ondersteuning/bijscholing de betreffende leerkracht kan gebruiken.

Na het uitvoeren van een verbeteractie is het effect veelal zichtbaar in de leerresultaten. Na een jaar uitvoeren van een nieuwe aanpak van de instructie op het gebied van Spelling is bijvoorbeeld de verwachting dat de resultaten omhoog zijn gegaan. Is dit inderdaad het geval, dan betekent dit dat de verbeteractie het gewenste resultaat heeft gehad en gecontinueerd kan worden. Vallen de resultaten tegen? Dan kan gekeken worden of de aanpak juist is

geïmplementeerd of dat er gezocht moet worden naar een andere manier om de opbrengsten te verhogen.

Let op dat het invoeren van een nieuwe methode doorgaans op korte termijn geen verhoging biedt van de resultaten. Leerkrachten hebben tijd nodig om de methode goed te leren kennen en in de vingers te krijgen. In veel gevallen is na de invoering in eerste instantie zelfs een terugval in resultaten te zien. Wanneer de leerkrachten meer ervaring met de methode hebben, wordt het positieve effect op de resultaten zichtbaar.

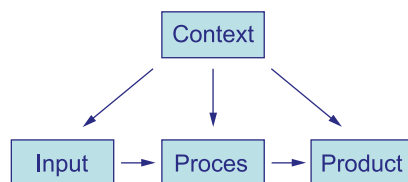
Bij het bestuderen van opbrengsten op schoolniveau, is het van belang de relatie te leggen tussen resultaten (product), onderwijsaanbod (proces) en aanwezige voorwaarden op school (input).

- Als leerlingen met onvoldoende letterkennis op school instromen (input), zal het moeilijk zijn om direct met technisch lezen van woorden te starten en de gestelde doelen op dit gebied te behalen.
- Als leerlingen voldoende letterkennis hebben, maar de beschikbare methode sterk verouderd is en pas op een laat moment aanvangt met technisch lezen (proces), zal het eveneens lastig zijn om gemiddelde resultaten bij technisch lezen te halen.
- Als de leerlingen voldoende letterkennis hebben en er een goede methode beschikbaar is, maar de doelen te hoog gesteld zijn (product), zal het resultaat ook tegenvallen.

Bestuderen van opbrengsten betekent dus niet alleen kijken of doelen zijn gehaald en dus of deze te hoog of te laag zijn gesteld, maar betekent juist ook kijken naar hoe deze resultaten tot stand zijn gekomen. Een goede analyse van factoren die van invloed zijn op de opbrengsten helpt bij het maken van een effectief verbeterplan. Wanneer de resultaten tegenvallen, kunnen de doelen te hoog gesteld zijn, maar kan ook gekeken worden of het proces wel goed genoeg was om het gestelde doel te bereiken. Zo nee, was het plan betreffende het onderwijsleerproces wel toereikend en ontbrak het aan de uitvoering ervan, of moet het plan aangepast worden? Een andere vraag is of de doelen wel realistisch zijn gezien de beschikbare middelen en aanwezige voorkennis (input).

Leerresultaten kunnen een trigger zijn om het proces en inputfactoren nader te bestuderen. Het CIPP-model en het model van Stake (1967) zijn hierbij handige hulpmiddelen. Het CIPP-model heeft als uitgangspunt dat een context, bestaande uit onder andere de schoolomgeving, de achtergrondkenmerken en thuissituatie van leerlingen, zowel van invloed is op:

- Input-factoren: denk aan aanwezige voorkennis, beschikbare leertijd en materialen.
- Proces: denk aan de didactische aanpak, de gebruikte methode.
- Product: in dit geval de leerresultaten.



Figuur 5.5: CIPP-model

Stake kijkt zowel Input, Proces als Product op vier manieren. Namelijk door te kijken naar:

- Intenties: Wat is het plan?
- Waarnemingen: Hoe is de situatie?
- Standaarden: Wanneer zijn we tevreden? Dus wanneer zijn de intenties naar behoren uitgevoerd?
- Oordelen: Is aan de standaarden voldaan?

Combinatie van het CIPP-model en de vier punten van Stake leidt tot onderstaande analysematrix.

Intenties	Waarnemingen		Standaarden	Oordelen
• extra rekentijd • voorkennis (groep 3 en 4) vergroten	• 1 uur extra rekentijd per groep per week heeft plaatsgevonden • cursus 'Met sprongen vooruit' is gevolgd	Input (voorwaarden)	• in alle groepen is 1 uur extra rekentijd gerealiseerd • cursus is gevolgd	• voldaan • voldaan
• meer gedifferentieerde instructie • meer aandacht voor automatiseren	• in alle groepen heeft instructie in 3 groepen plaatsgevonden • elke les is gestart met rekenstarters	Proces (onderwijsaanbod)	• in alle groepen wordt gewerkt volgens het directe instructiemodel (geobserveerd door IB-er) • elke les is gestart met een rekenstarter van 5 minuten	• voldaan • voldaan
• verbeteren resultaten rekenen-wiskunde groep 3 t/m 8	• toename resultaten in alle groepen	Product (leerresultaten)	• 60% A/B-leerlingen in alle groepen	• voldaan, behalve voor groep 6

Figuur 5.6: Analysematrix voor het kijken naar Input, Proces en Product

Op groepsniveau: Onderwijs afstemmen op behoeften van (groepen) leerlingen

In hoofdstuk 1 en 3 kwam aan de orde op welke wijze een leerkracht aan de hand van toetsresultaten het niveau van (groepen) leerlingen kan bepalen. De groepsrapportages vormen een basis voor het opstellen van een groepsplan:

- Op welk niveau moet de instructie aangeboden worden?
- Welke leerlingen kunnen samen in een niveaugroepje geplaatst worden?

Maar ook bieden sommige toetsen mogelijkheden om meer specifieke informatie te verzamelen:

- Aan welke categorieën van Rekenen-Wiskunde moet extra aandacht besteed worden?
- En welke spellingscategorieën vragen om extra aandacht?

In hoofdstuk 1 en 3 kwamen deze en andere vragen uitgebreid aan de orde.

Op leerlingniveau: Onderwijs, al dan niet op basis van opgesteld ontwikkelingsperspectief, afstemmen op behoeften van individuele leerlingen en leerlingen volgen

In hoofdstuk 1 en hoofdstuk 3 kwam aan de orde op welke wijze toetsresultaten ingezet kunnen worden om het onderwijs af te stemmen op individuele leerlingen, bijvoorbeeld aan de hand van de informatie die een categorieën- of foutenanalyse oplevert. In hoofdstuk 3 kwam ook aan de orde hoe toetsresultaten op leerlingniveau een basis kunnen vormen voor het opstellen van een leerrendementsverwachting als onderdeel van een ontwikkelingsperspectief.

5.3 Schoolbestuur

Veel scholen maken deel uit van een schoolbestuur dat bestaat uit meerdere scholen. Scholen uit een bestuur kunnen van elkaar leren door ervaringen uit te wisselen over bijvoorbeeld hun aanpak om de resultaten op een bepaald leergebied te verbeteren. Door de leeropbrengsten van scholen uit een bestuur met elkaar te vergelijken, ontstaat er zicht op de sterke en zwakke punten van een school. Op de gebieden waar een school goed presteert, kunnen de factoren die daaraan bijdragen in kaart gebracht worden zodat andere scholen daarvan kunnen profiteren. Andersom geldt dat de ervaringen van andere scholen benut kunnen worden om de prestaties op achterblijvende leergebieden te verbeteren.

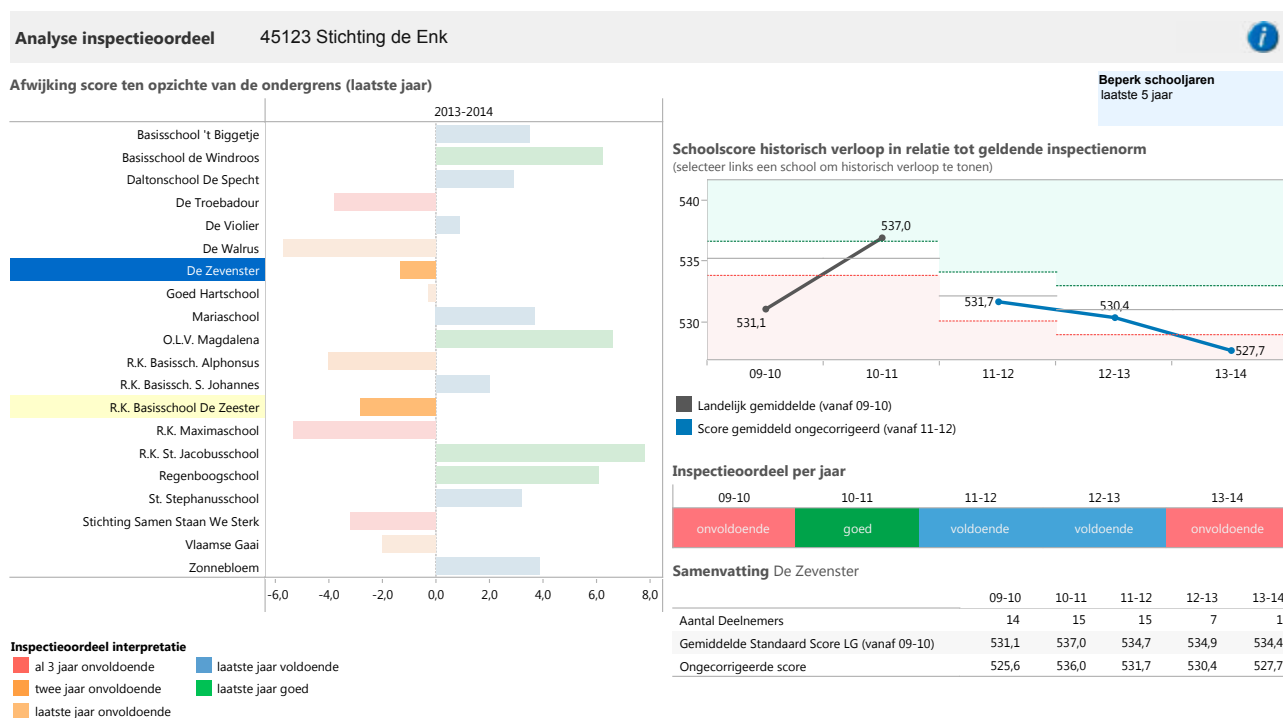
Is het gemiddelde bij Rekenen-Wiskunde op een school in een achterstandswijk hoger dan het gemiddelde bij Rekenen-Wiskunde op een school met een meer gevarieerde leerlingpopulatie? Dan kan dat bijvoorbeeld een aanleiding zijn om de aanpak van de eerste school te analyseren om zo elementen ervan toe te passen in de tweede school.

Ook bij verbeteracties die bestuursbreed, dus op meerdere scholen, plaatsvinden, is het waardevol om resultaten van de verschillende scholen met elkaar te vergelijken. Door eventuele verschillen in leerlingpopulatie op de scholen zullen de scores wellicht niet gelijk zijn, maar bij eenzelfde aanpak kan op de diverse scholen wel eenzelfde groei in scores verwacht worden. Waarom heeft de aanpak op de ene school meer effect dan op de andere? Door de verschillen in kaart te brengen kan het onderwijs op een individuele school profiteren van de aanpak bij omliggende scholen.

Een bovenschoolse module bij een leerlingvolgsysteem maakt het mogelijk om resultaten van verschillende scholen met elkaar te kunnen vergelijken. Dergelijke modules zijn op dit moment volop in ontwikkeling. In figuur 5.7 staat een voorbeeld van een rapportage van Infotopics (www.infotopics.nl). Een onderdeel van het concept dat Infotopics geeft besturen inzicht in de (verwachte) inspectieoordelen en de onderliggende schoolscores.

Voorbeelden van vragen die door middel van deze visualisaties worden beantwoord zijn:

- Hoe verhouden de schoolscores zich t.o.v. de geldende inspectienormen?
- Hoe ontwikkelen de schoolscores zich in de tijd?
- Wat is het verschil tussen de ondergrens en de schoolscore?
- Hoe verhouden de schoolscores zich ten opzichte van andere scholen binnen het bestuur?
- Hoeveel leerlingen hebben deelgenomen aan de Cito eindtoets?



Figuur 5.7: Rapportage uit Bovenschoolse module van Infotopics¹⁰ (www.infotopics.nl)

In figuur 5.7 staan aan de linkerkant alle scholen uit een bestuur. Met gekleurde staafjes staat aangegeven hoeveel punten het afgelopen jaar de gemiddelde score van de school afwijkt ten opzichte van de landelijke ondergrens zoals gehanteerd door de inspectie.

- Een balkje naar rechts betekent dat het gemiddelde van de school boven de ondergrens ligt; de scholen met een balkje naar rechts hebben voor dat jaar dus allemaal het oordeel 'voldoende' of 'goed' wat betreft hun eindopbrengsten.
- Een balkje naar links betekent dat het gemiddelde van de school beneden de ondergrens van de inspectie ligt; scholen met een balkje naar links krijgen voor dat jaar dus het oordeel 'onvoldoende' wat betreft hun eindopbrengsten.

De kleuren van de balkjes geven extra informatie: Een blauw balkje betekent dat de eindresultaten het laatste jaar volgens de norm voldoende zijn. Een groene balk betekent 'goed'. Bij de balkjes naar links geven de kleuren aan of de resultaten de voorgaande jaren ook al onvoldoende waren. Bij lichtoranje had de school alleen het laatste jaar een onvoldoende score. Bij oranje heeft de school voor het tweede jaar een onvoldoende score. Bij rood is dit het derde jaar waarop de school onvoldoende wat betreft de norm voor eindopbrengsten scoort. Door op de naam van een school te klikken ziet u meer informatie over het scoreverloop van de betreffende school. In het voorbeeld is school 'De Zevenster' aangeklikt. Rechtsboven staan t/m 2010-2011 de scores de gecorrigeerde scores met bijbehorende inspectiegrenzen weergegeven. Vanaf 2011-2012 zijn, conform de gegevens waar de inspectie naar kijkt bij de beoordeling, de ongecorrigeerde scores weergegeven. Bij de Zevenster ligt de gemiddelde score in 2013-2014 in het rode vlak, oftewel onder de ondergrens van de inspectie. De twee jaren ervoor valt de score van de school in het witte vlak, de school scoort dan dus voldoende. In 2010-2011 lag de score zelfs in het blauwe vlak, wat past bij inspectieoordeel 'goed'. In 2009-2010 was het oordeel net als in 2013-2014 onvoldoende. Door die twee onvoldoendes de afgelopen vijf jaar kleurt het

¹⁰ Aangezien de inspectienorm van 2006 bij Infotopics onbekend was, kon in de rapportage geen oordeel voor 2006 gegeven worden. Omdat voor een verwacht inspectieoordeel van de afgelopen drie jaar toetsgegevens vergeleken moeten worden met de inspectienorm, staat bij 2006 t/m 2008 'onbekend'

balkje van de Zevenster in de linker figuur oranje. De inspectieoordelen en de gegevens zoals weergegeven in de grafiek van de school staan rechtsonder weergegeven.

In de bovenschoolse module van het Computerprogramma LOVS kunnen bestuurders eveneens de resultaten op de centrale eindtoets van verschillende scholen over verschillende jaren met elkaar vergelijken. Ook kunnen de gegevens uit de LVS-toetsen op bestuursniveau geanalyseerd worden. Hierbij bestaat de mogelijkheid om scores zowel af te zetten tegen de gemiddelde vaardigheidsscore als tegen de door de Inspectie gehanteerde ondergrenzen.

Ook andere systemen bieden mogelijkheden voor bovenschoolse modules aan, denk aan bovenschoolse modules van ParnasSys, Dotcomschool en ESIS. Zie ook de brochure 'Vroegsignalering' van de PO-raad op www.poraad.nl > publicaties.

5.4 Onderwijsinspectie

Leerlingen en hun ouders moeten erop kunnen vertrouwen dat een school onderwijs van voldoende kwaliteit realiseert. Daarom houdt de Inspectie van het onderwijs toezicht op de kwaliteit van het onderwijs. Dit doet de Inspectie aan de hand van zeven indicatoren:

- 1.1 De resultaten van de leerlingen aan het eind van de basisschool liggen ten minste op het niveau dat op grond van de kenmerken van de leerlingenpopulatie mag worden verwacht.
- 1.2 De resultaten van de leerlingen voor Nederlandse taal en voor rekenen en wiskunde tijdens de schoolperiode liggen ten minste op het niveau dat op grond van de kenmerken van de leerlingenpopulatie mag worden verwacht.
- 1.3 De leerlingen doorlopen in beginsel de school binnen de verwachte periode van 8 jaar.
- 1.4 Leerlingen met specifieke onderwijsbehoeften ontwikkelen zich naar hun mogelijkheden.
- 1.5 De sociale competenties van de leerlingen liggen op een niveau dat mag worden verwacht.
- 1.6 De adviezen van de leerlingen voor het vervolgonderwijs zijn in overeenstemming met de verwachtingen op grond van de kenmerken van de leerlingenpopulatie.
- 1.7 De leerlingen functioneren naar verwachting in het vervolgonderwijs.

Met ingang van 2016 beoordeelt de Inspectie de tussenresultaten (indicator 1.2) niet meer.

De inspectie blijft het belangrijk vinden dat scholen de leerlingen goed volgen en zo het onderwijs op de leerlingen afstemmen. Ook vinden ze het belangrijk dat de leerlingen resultaten halen die passen bij hun mogelijkheden. Het bestuur en de scholen zijn echter verantwoordelijk voor deze kwaliteit. De Inspectie kijkt in het toezicht naar de wijze waarop besturen en scholen zorgen voor deze kwaliteit, niet meer naar het behalen van specifieke resultaten.

De inspectie kijkt nog wel naar de behaalde eindopbrengsten (indicator 1.1). In figuur 5.8 staat aangegeven in welke gevallen de eindresultaten als goed, voldoende of onvoldoende worden beoordeeld.

Beslisregel leerresultaten Indicator 1.1	De leerresultaten van de leerlingen aan het eind van de basisschool liggen ten minste op het niveau dat op grond van de kenmerken van de leerlingenpopulatie mag worden verwacht.
	De leerresultaten van de afgelopen drie schooljaren zijn:
goed ('4'):	drie keer goed.
voldoende ('3'):	- twee of drie keer voldoende; - één keer voldoende (aantal in de beoordeling betrokken leerlingen 10 of meer); - één keer voldoende (aantal in de beoordeling betrokken leerlingen minder dan 10) én voldoende leerresultaten in leerjaar 7; of - één keer voldoende (aantal in de beoordeling betrokken leerlingen minder dan 10) én onvoldoende leerresultaten in leerjaar 7 én voldoende leerresultaten in een vierde leerjaar 8.
onvoldoende ('2'):	- drie keer onvoldoende; of - één keer voldoende (aantal in de beoordeling betrokken leerlingen minder dan 10), onvoldoende leerresultaten in leerjaar 7 en onvoldoende leerresultaten in een vierde leerjaar 8.
niet te beoordelen ('5'):	bijzondere omstandigheden (zie bijlage D van deze regeling).

Figuur 5.8: Beslisregel beoordeling leerresultaten (<http://wetten.overheid.nl/BWBR0035918/2016-01-01#BijlageE>)

Behalve op de kwaliteit van het onderwijs houdt de Inspectie ook toezicht op de naleving van wet- en regelgeving door scholen en op hun financiën. Bij dit toezicht worden jaarlijks een aantal prioriteiten gedefinieerd die in dat jaar als uitgangspunt worden genomen voor het toezicht. Als een school niet voldoet aan de wet- en regelgeving, krijgt de school een zogenaamd aangepast toezicht voor naleving.

Op het moment dat de Inspectie risico's of mogelijke tekortkomingen op een school constateert, voeren zij vervolgonderzoek uit. Er zijn hierbij vier scenario's mogelijk:

1. Kwaliteitszorg op orde en geen tekortkomingen bij scholen en/of het bestuur: vertrouwen – geen vervolgonderzoek nodig.
2. Kwaliteitszorg op orde, maar tekortkomingen: afspraken met bestuur over eigen rol bij kwaliteits- en/of herstelonderzoek.
3. Kwaliteitszorg niet op orde: in alle gevallen voert de inspectie kwaliteits- en/of herstelonderzoek uit.
4. Financieel beheer niet op orde: combinatie van scenario's.

Zie voor de meest recente versie van het document "Analyse en waardering van opbrengsten" op www.onderwijsinspectie.nl > primair onderwijs > analyse en waardering van opbrengsten.

Geraadpleegde literatuur

Gebruikte literatuur

Berkel, van S., & Hilte, M. (2009). *Leerling- en onderwijsvolgsysteem Woordenschat*. Handleiding. Arnhem: Cito.

Berkel, van S., Krom, R., Heesters, K., Schoot, van der F. & Hemker, B. (2008). *Balans van het leesonderwijs halverwege de basisschool 4. Uitkomsten van de vierde peiling in 2005*. Arnhem: Cito.

Borich, G.D. (1971). *Expanding the Stake Model to increase information yield about new educational products*. Educational Technology, 11, 12, 21-23.

Bügel K., & Sanders, P.F. (1998). *Richtlijnen voor de constructie van onpartijdige toetsen*. Arnhem: Cito.
Online: www.cito.nl/onderzoek

Cijvat, A., & Bloemendaal, T. (2010). *Handreiking OGW. Doelen stellen op groeps- en schoolniveau met Cito-toetsen. Hoe doe je dat?* Utrecht: Projectenbureau Kwaliteit.

Cijvat, A., & Bloemendaal, T. (2010). *Handreiking OGW. Overzichten van toetsresultaten: LOVS Cito*. Utrecht: Projectenbureau Kwaliteit.

Cijvat, A., & Bloemendaal, T. (2010). *Handreiking OGW. Overzichten van toetsresultaten: Dotcomschool*. Utrecht: Projectenbureau Kwaliteit.

Cijvat, A., & Bloemendaal, T. (2010). *Handreiking OGW. Overzichten van toetsresultaten: ESIS Webbased*. Utrecht: Projectenbureau Kwaliteit.

Cijvat, A., & Bloemendaal, T. (2010). *Handreiking OGW. Overzichten van toetsresultaten: ParnasSys*. Utrecht: Projectenbureau Kwaliteit.

CvTE (red.) (2015). *Verantwoording Centrale Eindtoets PO*. Utrecht: CvTE.

Evers, A., & Resing, W. (2007). *Het drijfzand van didactische leeftijdsequivalenten*. De Psycholoog, p. 466-472.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
Online: www.psynip.nl/COTAN

Evers, A., Vliet-Mulder, J.C. van, Resing, W.C.M., Starren, J.C.M.G. & Bostel, H. van. (2002). *Cotan Testboek voor het onderwijs*. Amsterdam: NDC/Boom.

Inspectie van het onderwijs (2009). *Toezichtskader PO/VO*. Utrecht.

Inspectie van het onderwijs (2010). *Analyse en waarderings van opbrengsten*. Primair onderwijs. Utrecht.

Jansen, J., Scheltens, F., & Kraemer, J.M. (2007). *Leerling- en onderwijsvolgsysteem Rekenen-Wiskunde*. Handleiding. Arnhem: Cito.

Sanders, P.F. (red.) (2013). *Toetsen op School*. Arnhem: Cito.
Online: www.toetsenopschool.nl

Sanders, P., Dijk, P. van, Eggen, T., Otter, D. den, & Veldkamp, B. (2016). *RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens*. Enschede: RCEC.
Online: www.rcec.nl/Beoordelingssysteem

Scholte, E.M., & Ploeg, J.D. van der (2005). *ADHD Vragenlijst: Handleiding*. Houten: Bohn Stafleu van Loghum.

Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen*. Tilburg: TUP.

Stufflebeam, D.L. (1971). *The Relevance of the CIPP Evaluation Model for Educational Accountability*. Paper Annual meeting of the American of School Administrators. Atlantic City, N.J.

Wijs, de A., Krom, R., & Berkel, S. van (2006). *Leerling- en onderwijsvolgsysteem Spelling groep 4*. Handleiding. Arnhem: Cito.

Index

Index

A

absolute normering 60, 61
adaptief toetsen 7, 17
adviescore (A-score) 18
alternatief groepsprofiel 69, 71, 72
alternatief leerlingrapport 64, 69, 70, 72, 107, 108
analysematrix 113
analyseren 21, 52, 54, 60, 71, 74, 75, 76, 78

B

basisaanbod 18
basisarrangement 18
begripsvaliditeit 90
beheersingsniveau 65, 72
beheersingstoets 62
beoordelingsmodellen 8
beoordelingsschaal 8
betrouwbaarheid 8, 42, 55, 89
betrouwbaarheidsinterval 28, 30
bovenschoolse module 115, 116

C

categorieënanalyse 54, 74, 75, 76
categoriefout 77
centrale eindtoets 12, 60, 68, 93
certificeren 10
CIPP-model 112, 113
Cito Portal 27
Cito Volgsysteem 7
classificeren 10, 12, 54
Commissie van Onderzoek 11
Commissie voor de Indicatiestelling 11
Computerprogramma LOVS 22, 58, 60, 74
COTAN 84
Criteriumvaliditeit 91

D

diagnosticeren en plannen 78
diagnostische toets 78
diagnostisch gesprek 78
didactische leeftijd (DL) 66
didactische leeftijdsequivalenten (DLE) 43, 65, 66, 68, 79
differentiëren 18, 19
digitale toetsen 7

doelen stellen 111
doorgaande lijn 64, 65
Dotcomschool 65, 66
dwarsdoorsnede 30, 31, 111

E

eindtoets 10, 12, 43, 54, 93
Entreetoets 27
Esis B 22
evalueren 21
Expertgroep Toetsen PO 55, 58, 84

F

fairness 91
foutenanalyse 74, 76, 77
frustratieniveau 65
functioneringsniveau 53, 64, 70, 72, 107
fundamentele niveau 45, 47

G

gecorrigeerde scores 68, 71
geleide afname 7
geplande vaardigheidsgroei 60, 79, 81
gesloten vragen 7, 56
groei 42, 63, 66, 68, 72, 73, 74, 78, 80
groepsniveau 72, 74, 76, 78, 113
groepsrapport 24, 69

H

handelen 6, 21, 68, 72, 74, 78
handleiding 57, 58, 60, 86

I

IEP Eindtoets 12, 14
indicatieaanvraag 65
inhoudelijke interpretatie 61, 65
inhoudsbeschrijving 52, 53
inhoudsverantwoording 58, 69
instructieniveau 65
intelligentie 68
intelligentietest 43, 60
intensieve arrangement 18

K

kerndoelen 42, 45, 48, 49, 52
kwaliteitskaarten 72

L

leerdoelen 49, 52
leerlijn 42, 48, 49, 52, 53, 54, 70
leerlingdossier 6
leerlinggewicht 68
leerlingkenmerken 68
leerlingniveau 72, 74, 78, 113
leerlingrapport 22, 64, 69, 70, 72, 73
leerlingrapporten 21
leerrendementsverwachting 66, 79, 110, 113
LVS-toetsen 7, 43, 44, 53, 54, 55, 60, 61, 64, 66, 74, 76, 80, 97, 104, 108, 116
LWOO 65

M

methode 6, 52, 53, 54, 61, 62, 65, 69, 77, 78, 112
methodegebonden toetsen 6, 8, 19, 42, 52, 54, 61, 62, 63, 108, 109, 110
methodeonafhankelijke toetsen 6, 19, 42, 43, 52, 53, 54, 62, 104, 108, 109, 110
model van Stake 112
moeilijkheidsgraad 46, 54, 57, 58, 60
mondelinge toetsen 7
motivatie 108, 109, 110

N

niveaugroep 23, 42
niveauwaarden 67, 68
normen 60, 61, 88
normeringsonderzoek 57, 58
normgericht 21, 61

O

observatiepunten 8
observatiesystemen 6, 7, 8
onderwijsaanbod 42, 43, 48, 49, 53, 60, 61, 68, 77, 78, 81
onderwijscontinuüm 18, 72
ontwikkelingsperspectief 42, 60, 78, 79, 80, 81, 108, 110, 113
open vragen 6, 7, 55
ouderavond 105

P

ParnasSys 22, 67
Peil. onderwijs 40
pen-en-papiertoetsen 7
percentage goed 61
percentielen 36, 67

PIRLS

PIRLS 40
plaatsen 18
plannen 21
poppetjes-grafiek 13
PPON 35
Praktijkonderwijs (Pro) 65
proefafname 57

R

referentieniveau 14, 15, 42, 45, 46, 47, 48
referentiesets 48
reguliere leerlingrapport 106
relatieve normering 61
ROUTE 8 12, 17
ruwe score 60

S

samenwerkingsverband 11
schooladvies 12, 107, 109
schooladviesscore 15
schoolbestuur 114
schoolniveau 111, 112
SchoolOAS 22
schoolrapport 27
schoolzelfevaluatie 26
schriftelijke toetsen 7
selecteren 8, 10, 54, 57
self-referenced 21
signaleren 21, 74
slijten van de normering 58
sociale competentie 43
speciaal onderwijs 7, 42, 48
standaard 6, 55
standaardisatie 57
standaardscore 13, 15
streefniveaus 45, 47

T

Taal voor kleuters 43, 44, 74
testconstructie 85
testmateriaal 86
TIMSS 40
toelaatbaarheidsverklaring 11
toelatings- en doorstroomonderzoek 13
toetsdoel 54
Toetsen speciale leerlingen 43
Toetsgids 7
toetsscore 20, 47, 60, 61, 63, 65
toetsspecificaties 55
toetsvorm 7, 8, 54, 56
Toetswijzer 44

trendanalyse 30, 32, 33, 34, 111
tussendoelen 42, 48, 49, 52
tussenresultaten 43

U

uitstroombestemming 79

V

vaardigheidsgroei 73, 80
vaardigheidsniveau 21
vaardigheidsschaal 60, 63, 64, 80
vaardigheidsscore 20, 25, 26, 47, 58, 61, 63,
64, 65, 66, 68, 70, 71, 74, 76, 80
validiteit 8, 42, 90
verbeteractie 111, 114
verdiepte arrangement 18
volgtoetsen 43, 61, 63

W

waarderingskader opbrengsten PO 43
wetenschappelijke verantwoording 58

Z

zeer intensieve arrangement 19

Toetsen op School

Primair onderwijs

Judith Hollenberg en Marleen van der Lubbe

Piet Sanders (redactie)

Cito

Amsterdamseweg 13

Postbus 1034

6801 MG Arnhem

T (026) 352 11 11

klantenservice@cito.nl

www.cito.nl

Fotografie: Ron Steemers

