

Verantwoording bij de Lijst van toegestane instrumenten voor de verwijzing naar LWOO/PrO voor het schooljaar 2009-2010.

1. Inleiding

De certificeringscommissie van de landelijke RVC's ontleent zijn bestaansrecht aan de "Regeling regionale verwijzingscommissies en zorgbudget Voortgezet onderwijs". In deze regeling wordt in de Toelichting gesteld dat jaarlijks een ministeriële regeling opgesteld wordt waarin de voor het leerling-dossier toegestane screeningsinstrumenten of toetsen en tests zijn opgesomd. Hiertoe adviseren de RVC's jaarlijks het ministerie. Om dit te kunnen doen treffen de RVC's met de Commissie Testaangelegenheden Nederland (COTAN) van het NIP (nog steeds volgens de Regeling) een voorziening "die op onafhankelijke basis zorg draagt voor een beoordeling van screeningsinstrumenten of toetsen en tests in termen van de functionele vereisten zoals hierboven gesteld." De functionele eisen waarover gesproken wordt in de 'Regeling' betreffen: de validiteitsaspecten, de betrouwbaarheid en de normering. Tevens kunnen de RVC's volgens de Regeling nog beoordelen of de instrumenten positief scoren op: " - praktische bruikbaarheid (...) - geschiktheid voor de doelgroep (...)".

Door RVC-VO Landelijk is de Certificeringscommissie (CC) in het leven geroepen als de voorziening waarvan sprake is in de geciteerde tekst. In de commissie hebben vertegenwoordiger(s) van de COTAN, de samenwerkingsverbanden VO en de RVC-VO Landelijk zitting en staat onder voorzitterschap van een bestuurslid van RVC-VO Landelijk.

De commissie brengt in de junivergadering van RVC-VO Landelijk een advies uit in de vorm van een Lijst van Instrumenten en een verantwoording van de gemaakte keuzes bij de samenstelling van de lijst en de inhoudelijke beoordeling van de daarop voorkomende instrumenten. Wanneer dit advies door de RVC's wordt overgenomen (dan wel geamendeerd overgenomen) wordt dit advies doorgestuurd naar het ministerie van OCW. Wanneer het ministerie aan dit advies zijn goedkeuring hecht, publiceert zij deze als OCW regeling (www.cfi.nl).

Om de opdrachtgever (RVC-VO Landelijk en dus indirect het Ministerie voor OCW) evenals de toekomstige gebruikers inzicht te geven in de beweegredenen van de commissie die tot de samenstelling van de huidige lijst hebben geleid, wordt - zoals vermeld - jaarlijks een Verantwoording opgesteld. Deze verantwoording zal bestaan uit de volgende onderdelen:

- a) waarom toegestane instrumenten voorzien zijn van bepaalde beoordelingscategorieën (par.2);
- b) de gebruikte gegevens op basis waarvan de commissie haar beslissingen heeft genomen (par. 3);
- c) de gehanteerde criteria vertaald naar beoordelingscategorieën (par. 4);
- d) kort verslag van het onderzoek: de enquête (2006) en onderzoek van de database (2007 en 2008) in resp. par. 5 en 6;
- e) verantwoording van die aspecten op de lijst die afwijken van vorige jaren (par.7).

2. Lijst van toegestane instrumenten voorzien van beoordelingscategorieën

Wat vooraf ging

In het verleden is door de commissie en de landelijke RVC's altijd een strenge interpretatie gekozen van de in de inleiding geciteerde tekst "die op onafhankelijke wijze zorg draagt voor een beoordeling van screeningsinstrumenten of tests". Men ging er bij haar werkzaamheden van uit dat er een beoordeling moest komen in termen van al dan niet toegestaan, eventueel aangevuld met een aantal adviezen. Er valt bij deze wetstekst evenwel net zo goed te denken aan de interpretatie, dat de beoordeling in een afraden dan wel aanraden (en allerlei gradaties daartussen) van het gebruik van een instrument, waarbij men de uiteindelijke keuze voor het gebruik laat aan de deskundige gebruiker, onverlet de eigen verantwoordelijkheid van de onderscheiden RVC om als daar redenen toe zijn een kritisch standpunt in te nemen. De commissie is vorig jaar in haar advies om een aantal redenen explicieter geworden in haar advies aan de vergadering van de RVC-VO Landelijk.

De eerste reden was, dat steeds nadrukkelijker wordt opgepikt, dat de zorg in het onderwijs - en met name ook in de leeftijdscategorie van 12 tot 18 - op een heel gecompliceerde wijze is geregeld. Dit geldt ook voor de regelingen van de voor de indicatiestelling te gebruiken diagnostische instrumenten. De commissie heeft twee jaar geleden dan ook gemeend een eerste stap naar meer eenheid te moeten zetten: voor de leerlinggebonden financiering wordt om allerlei redenen gewerkt met een lijst voorzien van beoordelingscategorieën. De commissie is hier in principe in meegegaan. Op dit punt is niet geprobeerd om tot een volledig identieke testbenadering te komen. In beide sectoren is er weliswaar een overlap van taakstellingen in de testpraktijk, er zijn echter ook grote verschillen. Over het algemeen wordt in de LGF-praktijk in een veel extremer bereik gemeten ten aanzien van IQ, leerachterstanden en sociaal emotionele ontwikkeling. Het gaat om een ernstiger stoornis dan wel leerbeperking. Door de extremiteit en de grote specificiteit van de te meten aspecten in het LGF-traject zal daar ook niet altijd van (reeds) door de COTAN positief beoordeelde toetsen en tests uitgegaan kunnen worden. Wij houden in afwijking van de LGF het positieve COTAN-oordeel als "conditio sine qua non" overeind. Aan de andere kant is de individuele afname-eis zoals die voor de LGF - praktijk geëist wordt, voor ons veelal niet noodzakelijk. Alleen in gevallen als waar de betrokken leerling qua zwaarte van de problematiek mogelijk ook voor een indicatie van het REC (met name III en IV) zou kunnen komen of waar er sprake van is of zou kunnen zijn, dat de gevonden meetresultaten niet betrouwbaar kunnen worden geacht in verband met het niet optimaal kunnen presteren op een bepaald type test vanwege een beperking (zoals dyslexie of Nederlands als tweede taal) zal men tot individuele toetsing over moeten gaan. Enerzijds is dit nodig om tot een juist en afgewogen oordeel te komen, anderzijds kan het niet zo zijn dat een leerling-dossier ontstaat waarin mogelijk onbetrouwbare of onjuiste gegevens zijn opgenomen.

Een tweede reden om tot een lijst met beoordelingscategorieën over te gaan, was gelegen in het feit, dat de beleidstrend is om veel meer aan het beleid van de scholen over te laten. (Zie ook de ontwikkelingen bij Passend Onderwijs) Een advieslijst past beter in deze trend.

Een derde niet onbelangrijke reden was, dat de besluitvorming binnen de CC in intensieve interactie met het onderwijsveld plaatsvindt: onze adviezen en beslissingen zijn voor een groot deel gebaseerd op beoordelingen door samenwerkingsverbanden, cq. daar werkende diagnostici en RVC's die ons gevraagd (via enquêtes en open interviews) dan wel ongevraagd worden aangedragen. Aan de andere kant merken we dat waar de commissie in het verleden in sommige gevallen (*zoals in het geval welke IQ-tests wel of niet te gebruiken in geval van ernstige taalproblematiek of leerachterstanden*) tot aan- dan wel afraden besloten had, de terzake verantwoordelijken in de scholen die adviezen inderdaad over het algemeen ook opvolgden. Zo worden bij leerlingen met leerachterstanden op het PrO-niveau in de onderzoekspraktijk die toetsen en tests die in de lijst afgeraden worden ook nauwelijks nog gebruikt. Dit blijkt uit de monitor en uit

door ons uitgevoerde enquêtes. Een absoluut uitsluiten van een test bleek om die reden veelal niet meer noodzakelijk.

Een vierde punt betrof de opdracht erop toe te zien dat de instrumenten positief scoren op praktische bruikbaarheid en geschiktheid voor de doelgroep. Het zal duidelijk zijn dat de uitvoering van deze opdracht voor het overgrote deel zal moeten plaats vinden vanuit de praktijk; en wel op basis van commentaar dat door de diagnostici en RVC's de CC - gevraagd dan wel ongevraagd - wordt gegeven. Het zal tevens duidelijk zijn dat er bij betrokkenen geen algehele consensus heerst ten aanzien van de bruikbaarheid en geschiktheid van de door de CC opgesomde instrumenten. Daarom zijn relatieve uitspraken over de tests in de lijst meer in overeenstemming met de werkelijkheid en daar moet een wetenschappelijke verantwoording tenslotte toch op gericht zijn.

Een vijfde punt betrof de duidelijkheid van de rol die de CC zou moeten spelen in het spanningsveld van goed- en afkeuring. Nederland heeft een commissie van gezag (de COTAN). De Commissie Testaangelegenheden Nederland (COTAN) van het NIP (Nederlands Instituut van Psychologen) heeft als opdracht 'het bevorderen van de kwaliteit van tests en testgebruik in Nederland'. Door de COTAN worden psychodiagnostische- en didactische instrumenten beoordeeld aan de hand van criteria als uitgangspunten bij de testconstructie, kwaliteit van de handleiding en testmateriaal, betrouwbaarheid en validiteit. De instrumenten worden op relevante criteria door de COTAN tenminste als 'voldoende' beoordeeld. De COTAN heeft geen wettelijk gezag en kan het gebruik van een test niet verbieden en zij wil dat ook niet. Toch zijn de gronden waarop een test onvoldoende beoordeeld wordt soms zo graverend (bijvoorbeeld onvoldoende ten aanzien van betrouwbaarheid, begripsvaliditeit criteriumvaliditeit en de normering), dat het vanzelfsprekend is, dat voor het gebruik van een test binnen een taakstelling, zoals die van de RVC's een positieve beoordeling van de COTAN een eerste vereiste is. Het is daarom niet zinvol dat dit deel van de werkzaamheden door de CC nog eens over zou worden gedaan. Wel kan de CC via de Monitor beschikken over gegevens uit de praktijk en zich zo een beeld vormen van hoe een - overigens positief beoordeelde - test op validiteit, normen en dergelijke in onze bijzondere praktijk uitpakt. Het kan blijken dat er enige wrijving bestaat tussen wetenschappelijke goedkeuring en werking in die praktijk. Dit heeft de commissie onder andere vastgesteld in haar rapport over klassikale IQ-tests. (*De Greef & Van der Geest, 2005*). Het blijkt dat leerachterstanden sterker met sommige tests samenhangen, dan met andere. Kan de CC alleen op grond van een dergelijke vaststelling die test uitsluiten? Beslissend voor het advies van de CC was echter destijds, dat door een hoog IQ enerzijds (> 90) en tegelijkertijd behoorlijke leerachterstanden anderzijds, men gedwongen werd SE-problematiek als verklaring voor die achterstanden aan te voeren wilde men voor indicatiestelling in aanmerking komen. Een problematiek die niet altijd speelde. Wanneer de CC dan constateert, dat een bepaalde IQ-test(s) bij deze doelgroep veel hogere scores oplevert, wordt zij gedwongen maatregelen te nemen: in dit geval de test met deze eigenschappen uit de lijst te verwijderen. De constructeurs hebben deze beslissing veelal aangegrepen om noodzakelijke verbeteringen in de normen aan te brengen.

Een zesde punt betreft de toetspraktijk, waarbij er veelvuldig en vaak terecht over geklaagd wordt dat kinderen veel te vaak met testonderzoek worden lastig gevallen. In dergelijke gevallen moet de school of dienstdoende diagnosticus zijn eigen verantwoordelijkheid nemen en de afweging maken of hij toch van de waarschijnlijk minder betrouwbare toetsgegevens gebruik zal maken bij de aanvraag of dat hij opnieuw maar dan met een adequater instrument zal gaan kijken waar de leerling staat. Stel dat er een testresultaat voorhanden is van een test die positief beoordeeld is door de COTAN, maar die volgens de onderzoeksresultaten niet (zo) geschikt is. Moet er dan opnieuw getoetst worden? Men zou op zijn minst kunnen zeggen: dat hangt ervan af. Als het toetsresultaat in lijn is met eerdere gegevens van de leerling, of wanneer de toetsresultaten in overeenstemming zijn met de overige aangeleverde gegevens (te denken valt aan LVS, eerder IQ-gegevens, psychiatrische rapporten, onderwijskundig rapport, psychologisch rapport, e.d.) dan lijkt het goed verdedigbaar om niet opnieuw te testen maar in de aanvraag te beargumenteren dat de uitslagen zoals die verzameld zijn

een goed beeld van de leerling geven. Maar wanneer de testuitslag cruciaal is voor een te nemen beslissing, (bijvoorbeeld de gegevens liggen in of rond het overgangsgebied tussen LWOO en Pro) dan zou een RVC mogen verwachten dat een adequater toetsgegeven wordt aangedragen.

In het krachtenveld van COTAN, gebruikers en monitor lijkt een adviserende, voorlichtende taakstelling meer op zijn plaats en net zo effectief. Tenslotte, een categorieënsysteem maakt de afwegingen van de commissie duidelijker en meer verifieerbaar.

Tevredenheid met betrekking tot het adviessysteem?

In de raadpleging van de samenwerkingsverbanden, cq. diagnostici en RVC's is in de laatste enquête (2006) ook gevraagd naar de mate van tevredenheid met het nieuwe adviessysteem met zijn categorieën van 0 tot IV. *De enquête was dat jaar nadrukkelijk gericht op de RVC's en hun kamers enerzijds en de diagnostici in de samenwerkingsverbanden anderzijds. Van de eerste groep kreeg de CC 18 en van de tweede groep 34 ingevulde enquêtes terug.* Dit mag wat de RVC's betreft een zeer geslaagde respons genoemd worden. Voor de samenwerkingsverbanden lijkt dat in principe iets minder goed. De respons is hier iets meer dan 40 percent. We hadden op de respons hier geen invloed, aangezien de enquête op hun verzoek via het Platform is verspreid. Overigens zij erop gewezen, dat reacties na een herhaalde oproep vaak van mindere kwaliteit zijn. We gaan er daarom van uit, dat we de gemotiveerde samenwerkingsverbanden op dit punt in de enquête hadden. De tevredenheid met het nieuwe systeem was groot. Van de RVC's waren de reacties voor 83% positief, lopend van tamelijk positief (23%) tot ronduit positief (60%). Bij de samenwerkingsverbanden waren de reactie nog positiever: 91% waarvan 10% tamelijk positief en 81% ronduit positief. In hun stemverklaring gaven de gematigd positieven aan, dat het aanvankelijk tot onzekerheid in het veld had geleid, of dat men eigenlijk vond dat er in een categorieënsysteem een groter onderscheid tussen geschikt voor LWOO dan wel Pro gemaakt zou moeten worden. Dit laatste is vorig jaar een punt van overleg in de CC geweest. Tot dan hadden we gemeend nog niet zo ver te moeten gaan. De redenen daarvoor waren:

- a) we gingen ten aanzien van de IQ-meting al zover om voor potentiële Pro-leerlingen het gebruik van individuele toetsen ten stelligste aan te raden;
- b) ten aanzien van de leerachterstanden is het hele scala van basisschooltoetsen voorhanden, waarbij we er toen van uitgingen dat vooral ook de CITO-normen van 2003 voor Pro-leerlingen een uitkomst zouden kunnen zijn. Al bestond ook toen al het advies de toetsversie niet meer dan 20 DLE af te laten wijken van het toetsresultaat. Tot 2004 was men bij ontstentenis van de Normtabel 2003 sowieso verplicht door te toetsen tot men het individuele niveau van de leerling bereikt had;
- c) daarenboven hoefde vanaf een zeer lage score (vanaf DLE 20) bij technisch lezen het begrijpend lezen niet getoetst te worden. Dit is vorig jaar beperkt tot alleen geldend voor Pro-leerlingen; wat de SE-problematiek betreft, die speelt natuurlijk ook wel bij potentiële Pro-leerlingen, maar deze problematiek is voor het verkrijgen van een Pro-indicatie geen beslissingscriterium. Dat neemt niet weg dat in het belang van het kind de SE-problematiek in een aantal gevallen wel bepaald en gekwantificeerd zal moeten worden, al was het maar om tot een goede diagnose en daarop aansluitend handelingsplan voor deze leerling te komen. In dergelijke gevallen zijn vragenlijsten die door de kinderen zelf ingevuld moeten worden voor dit type leerlingen niet zo geslaagd. Het mondeling afnemen van de lijsten is niet toegestaan in verband met de normering. Misschien dat daisyspelers, Cd-romspelers en dergelijke een uitkomst zijn om de afname te standaardiseren. Ook de software pakketten Kurzweil 3000 (versie 10) of Sprint (Plus Pro) mogen hier niet

onvermeld blijven. De normering zou dan op het gebruik van deze hulpmiddelen afgestemd moeten worden.

De juni vergadering 2006 van de landelijke vereniging RVC-VO heeft de Commissie suggesties voor hun werkzaamheden meegegeven die maken dat we vorig jaar, maar ook dit jaar uitvoerig op het punt LWO/PrO terug moeten komen. Door nadrukkelijker van de Commissie te vragen de gegevens van de databestanden bij de besluitvorming te betrekken, kreeg de CC gegevens in handen over de digitale dossiers die maakten dat we met opmerkelijke voorstellen in dezen moesten komen. Zo kan men in het databestand vaststellen dat het hierboven onder a. genoemde advies geen brede navolging heeft gevonden in de praktijk (meestal wordt een klassikale test gebruikt, ook daar waar de leerachterstanden van dien aard waren dat men zich echt moest afvragen of de leerling de IQ-items wel begreep; NT2-leerlingen idem). Dit jaar viel in het onderzoek van de databestanden vooral ook op, dat in het twijfelgebied van IQ 75 tot 80 er geen enkel verschil in het testgebruik bestaat. Elke IQ-test is hier proportioneel ingezet. Ook zie je in dat databestand, dat ingeval er zeer grote leerachterstanden op technisch en/of begrijpend lezen aan te wijzen waren in de IQ-range van 90 tot 120, dat er tóch bij deze leerlingen schriftelijke zelfbeoordelinglijsten op het gebied van de sociaal emotionele problematiek werden afgenomen. We komen op deze zaken uitgebreid terug.

3. De gehanteerde criteria en de gebruikte gegevens op basis waarvan de commissie haar beslissingen heeft genomen.

3.1 De COTAN-beoordeling

De regeling, zoals die in het Staatsblad 2003, nr. 262 (p. 40 en 41 is geformuleerd, voorziet in de vraag aan welke criteria screeningsinstrumenten of tests dienen te voldoen. De daar genoemde RVC-voorziening (= certificeringscommissie) draagt ten eerste zorg voor een beoordeling in termen van de functionele vereisten: validiteit, betrouwbaarheid en normering. Het zal uit het voorafgaande duidelijk zijn dat hier het oordeel van de COTAN wat ons betreft bepalend is. De facto betekent het dat, wanneer een test niet op bovenstaande punten positief is beoordeeld (dat wil zeggen minimaal van een voldoende als beoordeling is voorzien) een test niet in aanmerking komt om opgenomen te worden in de lijst. Een aantal nuanceringen is hier echter op zijn plaats. Ten eerste, een IQ-test waarbij de criteriumvaliditeit niet voldoende is valt per definitie af. Men mag er namelijk van uitgaan dat intelligentie een behoorlijk stabiele eigenschap betreft en dat het daaruit voortvloeiend IQ een vrij stabiel gegeven moet zijn, dat als zodanig over een voor onze doeleinden behoorlijke prognostische kracht moet beschikken: met een IQ tussen de 55 en zeg 75 zal een leerling zonder PrO voorziening het in het VO niet redden. Maar zie in de lijst de WISC III, daar heeft men nog geen onderzoek naar de criterium validiteit gedaan. We hebben in het verleden via de uitgever geprobeerd hier druk achter te zetten. Het ziet er echter niet naar uit dat we ooit een vaststelling van de criteriumvaliditeit mogen tegemoet zien. Hierop komen we in hoofdstuk 7 terug.

Ten tweede kan in het geval van de didactische toetsen niet of in elk geval veel minder gesproken worden over stabiele eigenschappen en dientengevolge kunnen aan meetresultaten op het gebied van leerprestaties geen al te grote voorspellende waarde gehecht worden. Het gaat veel meer om het vaststellen van een al dan niet succesvol afgelegde leerweg. Het beperkt zich tot diagnostisch instrument op enig relevant moment. En het stelt eerder het resultaat vast van een interactief proces tussen leerling en school, dan dat het voldoende in staat kan worden geacht te voorspellen hoe het met de leerling in de toekomst zal vergaan. Het gaat om een resultaat, waarop handelingsplannen gebaseerd kunnen worden om daardoor juist invloed uit te oefenen op de vastgestelde gegevens omtrent de prestaties van leerlingen. Dit zou een zinloze onderneming zijn als we van een al te grote stabiliteit van leerprestaties uit zouden

moeten gaan. In overeenstemming met uitspraken zoals door de COTAN gedaan zal de criteriumvaliditeit in geval van de leervorderingstoetsen niet als beslissend beoordelingsmoment worden meegenomen. In geval van diagnose en behandeling zijn betrouwbaarheid, normering en begripsvaliditeit dus meer aan de orde dan criteriumvaliditeit. We moeten ook vaststellen dat voor bijna geen instrument op het gebied van de leerprestaties onderzoek naar de criteriumvaliditeit is gedaan. Ten derde, ook in het geval van de sociaal emotionele instrumenten tilt de commissie niet al te zwaar aan de prognostische capaciteit van de test. Uitspraken op basis van dergelijke tests (faalangst, prestatiemotivatie e.d.) zeggen in eerste instantie meer iets over de momentane situatie en over de oorzaak tot dusver en dus ook minder over de stabiliteit en (on)veranderbaarheid ervan. Dat laatste zou ook vreemd zijn. Men kan toch niet aan de ene kant vaststellen, dat iets een stabiele eigenschap is en dan vervolgens allerlei pedagogische aanpakken (prijzen, positieve feedback, stimuleren, succeservaringen) en trainingen zoals sova-training en Faalangst Reductie Training uitvoeren om die stabiele eigenschappen te veranderen. Natuurlijk kunnen een aantal sociaal emotionele aspecten verbonden zijn met vastgestelde stoornissen of beperkingen die wel permanent van aard zijn. In alle gevallen gaat het er echter vooral om dat vastgesteld moet worden dat er iets aan de hand is in termen van faalangst, prestatiemotivatie en/of emotionele instabiliteit en dat aannemelijk moet worden gemaakt dat dit van negatieve invloed is op de leerlingprestaties.

De gedateerdheid van de normen.

Wanneer tests door de COTAN worden beoordeeld, dan gelden deze oordelen niet tot in lengte van jaren. Ook tests hebben een houdbaarheidsdatum. Voor intelligentietests is het Flynneffect berucht. Elke 5 jaar neemt de intelligentie van referentiegroepen een aantal punten toe. Hierdoor zijn IQ-tests zeer gevoelig voor verjaring. In de COTAN is voor groepen tests vastgesteld wanneer dergelijke tests niet meer als adequaat mogen worden beschouwd. Probleem hierbij is, dat de COTAN slechts om de zoveel jaar met een nieuw boekwerk komen en dat dus een test gedurende de tijd tussen de laatste publicatie en de geplande publicatie qua beoordeling zwevende is. Dit probleem wordt over het algemeen met aanvullingen opgelost, waarin ook - alleen op verzoek van de auteur - herbeoordelingen voorkomen.

De IQ-tests worden doorgaans na verloop van tijd voor het Flynneffect gecorrigeerd. Opgemerkt zij dat dit effect voor onze doelgroep (onder de IQ van 90) nauwelijks of veel minder speelt. Volgens de COTAN zijn normen na 15 jaar verouderd en na 20 jaar zijn de normen niet meer bruikbaar. Dit houdt in dat tests tot 15 jaar als goedgekeurd mogen gelden, tussen 15 en 20 jaar weliswaar als verouderd gekenschetst moeten worden maar alsnog als toegestaan mogen worden beschouwd. Tests ouder dan 20 jaar moeten worden gekwalificeerd als niet meer te voldoen volgens de COTAN normen. Deze beslissingen zullen in het categorieënsysteem hun weerslag moeten vinden. In een aantal gevallen kan dit tot verlegenheidsbeslissingen in de commissie leiden, omdat er anders mogelijk geen noodzakelijke tests meer voorhanden zouden zijn. Te denken valt hier aan de PMT-K.

3.2 Bevindingen uit de praktijk.

In de regeling wordt aan de RVC's de mogelijkheid geboden ook op het gebied van geschiktheid en bruikbaarheid een oordeel over de tests te vellen. Uiteindelijk gaat het hierbij om (inter)subjectieve gegevens. De diagnostici in het veld en de RVC-leden maken samen in de praktijk uit wat ze geschikt en bruikbaar vinden. Er zijn een drietal voor ons in principe beschikbare middelen om achter deze intersubjectieve beoordelingen van tests te komen.

Gegevens uit het landelijke databestand.

In 2005 is voor het eerst gebruik gemaakt van de gegevens van het landelijke databestand om na te gaan, of een test om wat voor reden dan ook wel aansluit bij een door de RVC's gewenste praktijk. In de verantwoording van de lijst en later nog eens uitgebreid in *De Greef en Van der Geest, 2005* is aangetoond dat de toenmalige NDT-versie essentieel afweek van de andere klassikale IQ-tests en de vastgestelde leerachterstanden. Zoals gesteld kwamen de scores op deze test veelvuldig boven de 90 uit en werden diagnostici (en in het verlengde daarvan de scholen) in verlegenheid gebracht, omdat men in dat geval sociaal-emotionele verklaringen moest zien te vinden voor het falen van de leerlingen met betrekking tot de schoolprestaties. Dit leidde nogal eens tot frustraties en verkrampde rapportages in het veld en uiteindelijk tot een groei van bezwaarschriften. De vraag of men hier zich nu moet beperken tot de mededeling dat een dergelijke test niet of minder geschikt en/of bruikbaar is voor ons soort werkzaamheden, of dat men op de stoel van de COTAN moet gaan zitten en alsnog de normen afkeuren is hier niet aan de orde. De gegevens bleken zo manifest en overtuigend, dat de correcties die de commissie aanbeval (a. alleen nog maar de test gebruiken in de periode van oktober tot december en b. voor het laatst op de lijst, in afwachting van nader onderzoek) hebben ervoor gezorgd, dat er een nieuwe versie van de NDT is met bijgestelde normen. Uit verder onderzoek in het databestand is gebleken dat de NDT hierdoor een voor onze doeleinden veel beter instrument is geworden dan als klassikaal instrument in de lagere regionen beter voldoet dan de andere klassikale IQ-tests. Twee jaar geleden meldden we al dat de Certificeringscommissie vanaf dat moment systematischer gebruik zou moeten maken t.b.v. haar analyses van de Landelijke database, met inmiddels alle aanmeldingsgegevens van 6 jaar. Deze suggestie werd door de Landelijke vergadering van de RVC-VO overgenomen. Zie verder ook paragraaf 6.

Spontane reacties uit het veld (diagnostici uit de samenwerkingsverbanden, RVC's, testmakers)

Al gedurende een aantal jaren, eigenlijk vanaf het begin, heeft men (diagnostici, samenwerkingsverbanden, leden van RVC's, schoolmanagement) de commissie -al dan niet via de helpdesk van de Landelijke vereniging- weten te vinden met op- en aanmerkingen. Genoemd mag worden het symptomatisch rapporteren van de te hoge moeilijkheidsgraad van het Drempelonderzoek met betrekking tot het begrijpend lezen. Uiteindelijk heeft dit tot een evaluatieonderzoek geleid van *Melis (2005)*, dat bepalend is geweest voor de beoordeling van deze test en tot de toezegging van de auteur de normering te herzien. Het zal duidelijk zijn, dat hoe waardevol dergelijke op- en aanmerkingen ook zijn, we niet goed kunnen inschatten, hoe systematisch ze meegenomen behoren te worden bij de uiteindelijke beoordeling van een test voor opname in de lijst. Ze zijn wat ons betreft enerzijds waardevol gebleken voor het richten van onze aandacht op bepaalde problemen in de lijst en ze zijn daarom veelvuldig gebruikt als richtinggevend voor onze naspeuringen. Meer dan eens zijn ze achteraf ook vaak een ondersteuning van onze op andere wijze reeds verkregen methodologisch meer onderbouwde oordelen.

Enquêtes en interviews

Het hoofddeel van onze argumentaties voor opname en beoordeling voor de lijst haalden we tot vorig jaar uit enquêtes en interviews.

Sinds haar aantreden heeft de commissie met grote regelmaat ten behoeve van de opstelling van de lijst geënquêteerd in het veld om er achter te komen hoe bepaalde instrumenten ervaren worden in termen van geschiktheid en bruikbaarheid. Tot het veld werden in eerste instantie de RVC's gerekend. Zij moeten tenslotte bij de beoordeling van dossiers kunnen vertrouwen op, dan wel een goede inschatting kunnen maken van de kwaliteit van een instrument. Ze kunnen bovendien bogen op een grote ervaring met hoe een instrument in de praktijk uitpakt binnen het geheel aan gegevens dat de RVC's via dossiervorming ten dienste staat. Ze monitoren als het ware de toetspraktijk van de regio.

Daarnaast rekent de commissie natuurlijk ook de bevoegd deskundigen, die verantwoordelijk zijn voor testafname en –rapportage binnen het traject van dossiervorming en aanmelding bij de RVC, tot de experts op het voor ons relevante testgebied. Het zal duidelijk zijn dat deze deskundigen in de praktijk op een geheel andere wijze geconfronteerd worden met de (on)mogelijkheden en eventuele tekortkomingen van tests. Zij kunnen bijvoorbeeld aangeven in welke mate een instrument aansluit bij het beeld dat over de leerling bestaat en bij de leerling-prestaties en zijn functioneren zoals die blijken uit het onderwijskundig rapport. Als daar systematische afwijkingen plaatsvinden, moet dat consequenties hebben voor de geldigheid van het betreffende instrument. De afgelopen jaren heeft de commissie systematische gegevens verzameld binnen dit veld. In 2005 hebben we gebruik gemaakt van de kennis binnen de RVC's en van de bevoegd deskundigen uit het onderwijsveld voor zover het de IQ-tests betrof. We kunnen rustig stellen dat de respons uiterst representatief was. Alle RVC's hebben gereageerd en hetzelfde gold voor het overgrote deel van de samenwerkingsverbanden. Ook zijn er in dit geval interviews gehouden met ambtelijk secretarissen en testdeskundigen van RVC's, bevoegd deskundigen en met een enkele testconstructeur. De resultaten van deze werkzaamheden kunt u nalezen bij De Greef & Van der Geest (2005). In 2006 heeft de Certificeringscommissie de RVC's via de Landelijke Vereniging RVC-VO en de bevoegd deskundigen via het Landelijk Platform Samenwerkingsverbanden benaderd om achter de beoordeling van de geschiktheid en praktische bruikbaarheid te komen van de in de lijst opgenomen instrumenten. We hebben in de commissie de inschatting gemaakt dat er de afgelopen twee jaar niet veel substantieels gebeurd is in de testpraktijk met betrekking tot het gebruik van tests. Er zijn weliswaar een beperkt aantal tests toegevoegd aan de lijst, maar de periode is te kort geweest om hierover nu al gericht te kunnen vragen naar de ervaringen in de praktijk (bijvoorbeeld: de RVC's begonnen rond maart pas met de behandeling van de dossiers voor het komende jaar.) Uit de databestanden blijkt verder dat tests die tot dusver weinig gebruikt werden ook nu weer weinig gebruikt werden, zodat men niet mag aannemen dat er grote verschuivingen in de opinies van de respondenten zullen zijn opgetreden. Graag hadden we de mening van betrokkenen vernomen over de veranderingen en aanscherpingen in de lijst: de splitsing in LWOO en Pro in termen van de beoordelingscategorieën, het in bepaalde gevallen doortoetsen en het vermijden van schriftelijke zelfbeoordelingvragenlijsten in geval van ernstige leeszwakte. Het Platform van de Samenwerkingsverbanden heeft hier zojuist onderzoek over gedaan. De uitslagen zijn echter te laat om ze in de huidige lijst mee te kunnen nemen. Hoewel de vorig jaar voorgestelde veranderingen eigenlijk niet eens zoveel veranderingen inhielden, maar vaak eerder slechts een explicitering waren van adviezen en opmerkingen die veelal in noten bijgevoegd waren, kunnen met name de herinvoering van een zij het gematigder doortoetsen tot enig –hopelijk tijdelijk- ongenoegen hebben geleid. We hebben om boven genoemde redenen besloten om de gegevens van de enquête van twee jaar geleden weer als uitgangspunt te nemen voor de beoordeling van de instrumenten op dit punt. Voor de volledigheid herhalen we de rapportage hiervan in par. 5.

4. De gehanteerde criteria vertaald naar beoordelingscategorieën

In het bovenstaande hebben we de criteria opgesteld waarmee de tests worden bekeken. De volgende zaken staan hierbij centraal:

1. de COTAN beoordeling (uitgangspunten bij testconstructie, kwaliteit testmateriaal en kwaliteit van de handleiding, normen, betrouwbaarheid, begripsvaliditeit en criteriumvaliditeit; let wel op de nuancering die we aanbrachten in par. 3.2);
2. de ouderdom van de normering;
3. de beoordeling door het veld;
4. de onderzoeksgegevens uit het landelijke databestand.

Aangezien we ook dit jaar - in tegenstelling tot twee jaar geleden- via het landelijke databestand systematisch over zeer betrouwbare algemene gegevens over onze

doelgroep konden beschikken, zal het duidelijk zijn, dat we op grond van deze andersoortige extra gegevens onze categorisering hebben kunnen cq. moeten bijstellen.

Uitgangspunt voor opname in onze lijst is en blijft de positieve beoordeling op de relevante aspecten: betrouwbaarheid, normering, begripsvaliditeit en veelal de criteriumvaliditeit. Wanneer hier niet op alle punten een "voldoende of hoger" beoordeling gegeven is, wordt het instrument niet in de lijst opgenomen. (m.u.v. de didactische toetsen en de SE-vragenlijsten waar we aan de criteriumvaliditeit niet al te grote prioriteit toekennen).

Op basis van de ons ter beschikking staande gegevens en analyses is de commissie in het verleden gekomen tot de volgende categorisering.

Categorie 0: Er is nog onvoldoende ervaring met dit instrument in het veld en het komt nog te weinig voor in het landelijke databestand om een oordeel over de geschiktheid dan wel bruikbaarheid voor de doelgroep(en) LWOO- en/of PrO-leerlingen uit te kunnen spreken;

Categorie I: volgens de uitslag van de enquête is dit instrument geschikt dan wel bruikbaar voor de afname bij potentiële LWOO- en/of PrO-leerlingen; de analyse van de gegevens uit het landelijke databestand ondersteunt de enquête uitslag. De normering is niet ouder dan 12 jaar;

Categorie II: volgens de enquête uitslag is dit instrument geschikt dan wel bruikbaar voor afname bij potentiële LWOO- en/of PrO-leerlingen; ook uit het landelijke databestand blijken geen negatieve aspecten voor de beoordeling op dit punt; de normering is echter nogal verouderd (12 à 15 jaar; ingeval een test de enige is op zijn gebied zelfs ouder dan 15 jaar);

Categorie III: meer dan de helft van de respondenten vonden dit instrument niet zomaar geschikt of bruikbaar voor afname bij potentiële LWOO- dan wel PrO-leerlingen en/of de analyse van de gegevens uit het landelijke databestand geeft aan dat aan het gebruik van dit instrument voor de doelgroep(en) (potentiële LWOO- dan wel PrO-leerlingen) negatieve effecten kleven die van dien aard en omvang zijn **dat aangeraden wordt het gebruik van een ander instrument voor het te onderzoeken criterium in overweging te nemen;**

Categorie IV: 70 percent of meer van de respondenten vonden dit instrument niet geschikt of bruikbaar voor potentiële LWOO- dan wel PrO-leerlingen en/of de analyse van de gegevens uit het landelijke databestand geeft aan dat aan het gebruik van dit instrument bij de doelgroep(en) zeer negatieve effecten kleven die van dien aard en omvang zijn **dat het gebruik van dit instrument bij de onderhavige doelgroep ernstig ontraden moet worden.**

Het zij vermeld, dat categorie IV in het verleden betekend zou hebben dat we het instrument van de lijst hadden afgevoerd. Nu hebben we de term *ernstig ontraden* gehanteerd. Ook categorie III houdt een advies in om te zien naar een ander instrument. Hoe hiermee nu in de praktijk om te gaan? Meestal gaat het om intelligentie- en SE-tests die klassikaal en/of schriftelijk zijn afgenomen. Dit is dan in het jaarlijkse ritme van schoolonderzoek en testprocedures gebeurd. In een flink aantal gevallen acht de commissie (en wel op basis van allerlei onderzoek) klassikale/schriftelijke tests niet of niet voldoende geschikt voor delen van onze leerling-populatie. Dit hangt meestal samen met ernstige taal- en leesproblemen van de leerling, als gevolg van dyslexie, ernstige spraak- en taalstoornissen of Nederlands als tweede taal. Het zal duidelijk zijn dat het bevoegde gezag (en binnen de scholen de deskundigen) en de RVC-VO hier hun eigen verantwoordelijkheden hebben en moeten nemen en dat uiteindelijk in de interactie tussen beide instanties uitgevonden dient te worden hoe met deze problematiek in het algemeen en individuele gevallen om te gaan. In de commissie is hierover ook uitvoerig

gesproken. We kunnen ons voorstellen, dat men in bepaalde gevallen en omstandigheden de klassikaal en schriftelijk verzamelde gegevens in de beschikkingsprocedure toch accepteert. Ten eerste dient vermeld, dat voor sommige leerlingen een zoveelste toetsmiddag niet goed op te brengen valt. Ook is duidelijk dat in de korte spanne tijds die tussen de verzameling van de gegevens en de RVC beslissing ligt, onnodig werk voorkomen dient te worden. Daarnaast is met hertesten geld gemeoid. Dus als een en ander niet echt nodig is, dan moet dit in goed overleg voorkomen worden. Wanneer lijkt nu het hertesten onnodig? Wanneer de klassikaal en schriftelijk afgenomen tests in overeenstemming zijn met vroegere op een andere manier verzamelde gegevens (*bijvoorbeeld: een recente NIO met een vroegere WISC-R*), en/of wanneer gegevens harmonisch overeenkomen met andersoortige gegevens in de aanvraag (*bijvoorbeeld van een leerling is een NPV-J afgenomen hoewel hij voor lezen slechts een DLE van 25 heeft, maar de gegevens komen overeen met eerder jeugdpsychiatrisch onderzoek en/of met gegevens uit het onderwijskundig rapport*) kan ervan worden uitgegaan dat de aangeleverde gegevens kloppen en niet door een categorie I of II test vervangen hoeven te worden. In het andere geval (*bijvoorbeeld de NIO komt veel lager uit dan op grond van vorig intelligentie onderzoek of op grond van resultaten op de inzichtvakken mocht worden verwacht, of de leerling geeft in de NPV-j aan sociaal adequaat te handelen, waar de leerkracht in de AVL/DVL aangeeft, dat dat geenszins het geval is*) zal men toch tot andere evidentie (*bijvoorbeeld via hertesten met een ander instrument, of observatie van de psycholoog, of het invullen van vragenlijsten door leerkracht of ouders*) moeten beschikken: niet alleen zou in het andere geval de beslissingsprocedure op mogelijk onjuiste gegevens berusten, maar ook zouden foutieve persoonlijke gegevens over de leerling in omloop komen en daaraan wil toch niemand meewerken.

5. Kort verslag van de enquête: opzet en voornaamste uitkomsten

In de eerste paar maanden van 2006 heeft de commissie een enquête uit laten gaan naar de RVC's en naar de bevoegd deskundigen in het veld. In totaal kregen we 18 ingevulde vragenlijsten terug vanuit de RVC's en 34 vanuit de samenwerkingsverbanden. De respondenten waren goed over het land verspreid. De uitkomsten van deze enquête schragen ook nu nog het overgrote deel van onze beslissingen, zodat hier de belangrijkste uitkomsten vermeldenswaardig zijn in termen van verantwoording.

Inhoudelijk zag de vragenlijst er als volgt uit: van elke test/vragenlijst werd gevraagd of deze

- a) van zelden tot regelmatig gebruikt werd;
- b) (on)geschikt was voor LWOO;
- c) (on)geschikt voor Pro;
- d) of men al dan niet problemen had met dit instrument;
- e) Daarnaast was er ruimte om commentaar te leveren.

Het gaat hier niet aan, om uitgebreid de vele gegevens van onze enquëtering weer te geven. We zullen ons hier tot een aantal hoofduitkomsten beperken, die hun impact voor de uiteindelijke beoordeling in de lijst hebben. Is men geïnteresseerd in de gedetailleerde uitkomsten dan kan men zich tot de commissie wenden.

Het viel op, dat de beide partijen in de enquête veelal met elkaar overeenstemden (en dat er slechts een paar interessante verschillen in de antwoorden tussen de RVC-deskundigen (RVC) en de bevoegde deskundigen uit de samenwerkingsverbanden (SWV) te noteren vielen.) Zo geven bijvoorbeeld zowel RVC als SWV aan, dat (in deze volgorde) de NIO, WISC-R, NDT, GIVO gebruikt worden; vervolgens minder gebruikt SON-R, SON-R-verkort en IVO; tenslotte NSCCT en de Internet versie van de IVO bijna niet. Deze vastgestelde eenstemmigheid betekent, dat we eigenlijk meestal geen onderscheid hoeven te maken tussen beide groepen. En, dat de uitkomsten, via verschillende kanten tot ons gekomen vergelijkbaar en waarschijnlijk representatief zijn. Over het algemeen was SWV wel een tikkeltje milder in zijn oordeel over de geschiktheid, dan wel het

problematisch zijn van een instrument. Natuurlijk vermelden we hier de afwijkende standpunten wanneer die afwijkingen opvallend zijn,

Intelligentietests. Bij de intelligentietests viel alleen op dat de GIVO door beide groepen voor meer dan 50 percent als ongeschikt werd gekenmerkt in het geval van potentiële Pro-leerlingen. SON-R en WISC-R vielen op alle hierboven genoemde aspecten heel positief op. De SWV groep geeft – in tegenstelling tot RVC - tenslotte in meerderheid aan, dat de NDT problematisch is. Ook in 2006 werd in een behoorlijk aantal gevallen weer gemeld, dat de NDT te hoog meet in vergelijking met de leerachterstanden. Dit wordt ook door een aantal RVC's gemeld. Een zaak om in de gaten te houden. De onvrede haalt de kritische 50 percentscore echter niet. (het zij hier echter alvast vermeld, dat onderzoek in het databestand duidelijk maakt dat dit in tegenstelling tot ons onderzoek in 2005, nu niet meer het geval is. Een na-ijlingseffect?)

Sociaal-emotionele problematiek.

De screeningsinstrumenten/vragenlijsten die hier nadrukkelijk genoemd werden om hun veelvuldig gebruik, zijn de NPV-J, AVL, SVL en PMT-K. Een aantal instrumenten werd nauwelijks of niet gebruikt hoewel ze al sinds jaar en dag op de lijst stonden. Het gaat hier voornamelijk om de AUTI-R en de AVL-ADHD vragenlijst. Maar ook om SCHOBL-R, SSAT, VAK en VvGK.

Op de vraag of de instrumenten geschikt zijn, antwoordden SWV en RVC niet identiek. De RVC was duidelijk ontevreden over de AVL-ADHD en de AUTI-R (voor 50 percent van de antwoorden), terwijl SWV of de antwoorden niet invulde (resp. 23 en 29) of positief gestemd was (10 dan wel 5). De andere weinig gebruikte instrumenten kregen heel weinig beoordelingen.

Omdat de sociaal-emotionele problematiek voor de indicatie van Pro geen rol speelt, werd hier niet apart naar gevraagd.

Tot slot, over het algemeen werden er geen problemen gemeld bij het gebruik van de SEP-instrumenten.

Technisch lezen. SWV en RVC meldden, dat voor het technisch lezen eigenlijk alleen de Brus en DMT veelvuldig worden gebruikt. Ver daarna volgen Klepel en 'Leestechiek & Leestempo'. Alle toetsen werden geschikt bevonden voor zowel LWOO als Pro. Volgens RVC is de Klepel minder geschikt voor Pro-leerlingen. Dat is echter merkwaardig omdat deze toets juist voor orthopedagogische doeleinden is ontwikkeld en op de meest eigenlijke wijze het technisch lezen meet. In het verleden is meer dan eens vastgesteld, dat zwakke taalgebruikers, zoals NT2-leerlingen op de nonsens tests net zo goed scoren als de meer talige leerlingen. Verder correleren bij zwakke leerlingen de Brus, Avi en Klepel hoog met elkaar. Men heeft met geen van de besproken tests problemen.

Begrijpend lezen. Ook hier weer grote eenstemmigheid over het regelmatige gebruik: het Drempelonderzoek en de Toetsen Begrijpend Lezen van CITO en de CLIB-toetsen (inmiddels van de lijst verdwenen) werden vaak gebruikt. In mindere mate werden de Aarnoutse-toetsen gebruikt, m.n. de A 4/6.

Wat de geschiktheid betreft: hier liepen de meningen uiteen. De RVC vond in grote meerderheid het Drempelonderzoek hier ongeschikt met name voor de potentiële LWOO-leerling, terwijl SWV met deze toets beduidend minder moeite had. Wanneer we het onderzoek van Melis erbij nemen, dan kondigt zich hier misschien een belangentegenstelling aan. Begrijpend lezen in het Drempelonderzoek geeft namelijk lagere waarden in termen van DLE dan concurrerende toetsen. Aangezien het kwantitatieve onderzoek van Melis de oordelen van RVC ondersteunen, waren we in eerste instantie geneigd om voor de beoordeling de uitspraak van RVC te volgen. Echter, nu we sinds vorig jaar beschikken over alle gegevens in de database kunnen we nu voor grote aantallen nagaan in welke mate er een verschil valt aan te wijzen tussen Drempelonderzoek en bijvoorbeeld de CITO toetsen begrijpend lezen. (Zie hiervoor paragraaf 6)

Onder het hoofdje 'problemen met' viel op dat wanneer men de niet-invullers van deze vraag (plm. 40 percent) even buiten beschouwing laat, er nogal wat problemen bestaan met deze instrumenten. Het Drempelonderzoek wordt hier relatief veel genoemd; in mindere mate ook door SWV en wel (a) omdat dit onderzoek hier laag dan wel extreem hoog scoort of (b) (voor SWV ook belangrijk) nogal wat geld kost. Hier werd naast het drempelonderzoek ook de CITO-toetsen vaak genoemd. Ook de toonzetting is hier vaak niet mis te verstaan. Aarnoutse krijgt ook zijn deel: het dle-bereik is dermate klein dat het geen inzicht verschaft (zou verschaffen).

Op het Drempelonderzoek na (tussen de 60 en 70 % van de wel op deze vraag antwoordende RVCs als problematisch beschouwd) vallen de toetsen niet binnen de kritisch grenzen van een lage categorisering op de lijst. Vergeleken bij de andere onderdelen, is men ten aanzien van de tests voor begrijpend lezen het minst tevreden met de bestaande situatie.

Tenslotte, hoe begrijpelijk ook, een aantal zaken - vooral door SWV gemeld- zijn niet relevant. Het gaat dan voornamelijk om kritiek op testinstanties die aan elke testafname verdienen. (Drempel en CITO). Hier hebben we geen bevoegdheid. Alleen de kwaliteit van het geleverde werk staat bij de Certificeringcommissie ter discussie.

Spelling. Van de spellingsinstrumenten worden vrijwel uitsluitend en in deze volgorde de diverse SVS-versies (volgens het databestand plm. 50%) en het Drempelonderzoek (40%) en in heel bescheiden mate ook de Toelatingstoets voor Leerwegondersteunend en Praktijkonderwijs (3%) en de diverse toetsversies van Aarnoutse (2.7%) gebruikt. De TPVO en de internetversie daarvan kwamen toen eigenlijk niet of nauwelijks voor. De geschiktheid van de toets voor het LWOO staat ook in de hierboven weergegeven volgorde. Hetzelfde geldt voor de geschiktheid voor Pro-leerlingen en het niet problematisch zijn van het instrument. Uit het feit dat er nauwelijks aanvullende opmerkingen gemaakt werden, mag men afleiden, dat er nauwelijks problemen bestaan op dit gebied van toetsing.

Inzichtelijk rekenen.

Voor het inzichtelijke rekenen staan ons maar een klein aantal toetsen, te weten vijf, ter beschikking en daarvan worden er eigenlijk maar twee regelmatig gebruikt. Het gaat hierbij om Rekenen & Wiskunde en in iets mindere mate om het Drempelonderzoek. Slechts heel af en toe werd de Toelatingstoets voor Leerwegondersteunend en Praktijkonderwijs als regelmatig voorkomend genoemd. Ook hier bleken de beide TPVO versies nauwelijks of niet gebruikt te worden.

Alle toetsen werden geschikt voor LWOO-leerlingen genoemd. Het hoeft geen betoog dat deze vraag voor toetsen waarvan aangegeven was dat ze nauwelijks gebruikt worden, heel vaak niet was ingevuld.

Hetzelfde kan ook gezegd worden met betrekking tot de geschiktheid voor Pro-leerlingen. TPVO wordt hier even vaak geschikt als ongeschikt genoemd. We mogen daar echter geen al te grote betekenis aan hechten. Het gaat om slechts twee positieve en twee negatieve beoordelingen. Bovendien heeft de TPVO (in tegenstelling tot de internetversie) slechts een ondergrens van 31 DLE.

Slotopmerkingen over de enquête.

Het is gebleken, dat er in het veld redelijk grote overeenstemming bestaat over gebruiksvoorkeuren, geschiktheid en bruikbaarheid van de toegestane instrumenten. Ook blijkt meer nog dan in het verleden, dat de meningen ook over de jaren heen vrij consistent zijn. We hebben daarom in de commissie vastgesteld, dat een groots opgezette enquête zoals eerder verspreid en uitgevoerd, in kwantitatief opzicht niet meer die interessante resultaten opleverde, zoals we die in het begin kregen. Aan de andere kant troffen we wel veel stellingnames en uitwerkingen aan die voor ons interessant waren en waar we in de komende jaren iets meer mee zouden willen en kunnen doen. Zo neemt in de opmerkingen het aantal kritische geluiden toe van het gebruik van klassikale IQ-tests in geval van dyslexie en - voor ons van nog groter belang - in geval van

potentiële Pro-leerlingen, niet alleen van de kant van de RVC's maar ook van Inspectie en scholen. Naast de GIVO duikt hier de NIO nog wel eens exemplarisch op. In de afgelopen twee jaar hebben we geprobeerd met het inhoudelijk interessante materiaal van het databestand enige antwoorden op vragen uit het veld te vinden. Zie daarvoor paragraaf 6.

6. Analyse van het landelijke databestand

6.1. Inleiding

Door het raadplegen van het landelijke databestand was de commissie in staat belangrijke informatie te verzamelen over:

- a) de frequentie van het gebruik van de diverse ter beschikking staande instrumenten als een objectieve maatstaf om vast te stellen welk instrument bij voorkeur in de praktijk wordt gebruikt. Je zou kunnen stellen, dat door de scholen hiermee indirect aangegeven wordt dat ze de frequent gebruikte instrumenten ook de meest geschikte dan wel handigste instrumenten vinden, hetgeen weer van belang is voor de vaststelling van de categorieën 0 vs. I t/m IV en om eventueel te besluiten om een instrument van de lijst af te halen wanneer dat instrument jarenlang niet of nauwelijks wordt gebruikt, en
- b) om na bepaalde bewerkingen van de data over de kwaliteit van deze instrumenten uitspraken te kunnen doen ten aanzien van het gebruik bij onze doelgroepen in termen van bruikbaarheid en geschiktheid. We moeten in het geval van de kwaliteit van de diverse instrumenten namelijk goed in de gaten houden, dat onze populatie van leerlingen (de potentiële LWOO- dan wel Pro-leerlingen) niet zo maar een homogene groep genoemd mag worden. Aan de ene kant hebben we te maken met leerlingen zonder echte leerstoornissen of al te grote SE-problematiek, met alleen maar over de hele lijn wat minder cognitief vermogen en wat achterblijvende leerprestaties. Daarnaast hebben we ook te maken met leerlingen die op deze gebieden veel verder achterblijven, en met leerlingen met duidelijke leerstoornissen en/of psychische beperkingen. Het kan niet zo zijn dat alle door ons goedgekeurde instrumenten voor alle leerlingen uit onze doelgroep ook zomaar geschikt zijn. Het is van belang na te gaan in de databestanden hoe de diverse instrumenten anders presteren afhankelijk van de variatie in de doelgroep. Mochten er grote verschillen bestaan tussen de diverse instrumenten ten aanzien van deze variaties in de doelgroep, dan zal dat zijn consequenties moeten hebben voor de categorisering van die instrumenten.

6.2. De vraagstellingen

De visie zoals in paragraaf 6.1 verwoord, werd in juni 2006 door vergadering van RVC-VO Landelijk overgenomen. In het schooljaar 2006/2007 is er binnen RVC-VO Landelijk intensiever dan gebruikelijk over de kwaliteit van de meetprocedure en inzonderheid de kwaliteit van de instrumenten gediscussieerd. Zowel schriftelijk als ter vergadering. Globaal genomen heeft men zich achter het visiestuk van het dagelijks bestuur en de concretisering daarvan zoals verwoord in de Toelichting geplaatst, zoals die voor de vergadering van 14 februari 2007 is toegestuurd. In concreto betekende dit voor de Certificeringscommissie dat men voor haar onderzoek van de databestanden aan de slag kon gaan en wel toen met de volgende kwesties:

1. in welke mate is het (on)verantwoord zich op klassikale schriftelijke IQ-tests te beroepen, als de leerachterstanden aangeven, dat de leerlingen qua lezen en taal deze tests moeilijk of niet aan zullen kunnen, en dat men daarom moet vrezen, dat men in dat geval niet alleen het IQ meet, maar dat men tegelijkertijd ook het afwezig zijn van leerprestaties meet. We moeten met andere woorden vrezen voor contaminatie-effecten waarbij het IQ lager uit zal vallen, dan bij andere individuele of

non-verbale tests. We stellen hier niet ter discussie of de beoordeling van de COTAN al dan niet juist zou zijn, maar hooguit, of onze bijzondere doelgroep wel past binnen de normale afloop van overigens betrouwbare en valide instrumenten;

2. in welke mate is het verantwoord uit te gaan van de normtabellen, zoals aangegeven door het CITO. Wat gebeurt er als een leerling een veel te moeilijke of veel te makkelijke toetsversie te behappen krijgt. Men kan zich voorstellen, dat faalangst en frustratie aan de ene kant of aan de andere kant gebrek aan concentratie (omdat er van de opgave geen uitdaging uitgaat), maakt dat de uitkomsten niet de optimale prestatie weer zullen geven;
3. in welke mate is het (on)verantwoord dat men gebruik maakt van SE tests die een te groot beroep doen op het taal- en leesvermogen van onze doelgroep;
4. is het mogelijk op basis van de samenhang tussen IQ/SEP te komen tot aanbevelingen voor de instrumentenlijst.

De Commissie heeft deze vier vragen het vorig jaar tot uitgangspunt gekozen voor haar onderzoek. Hierbij moet aangetekend, dat we de eventuele samenhang tussen SEP en leerachterstanden voorlopig moeten hebben laten rusten, en wel in verband met de ons toegestane tijd. (Overigens, zou dan een en ander bijgesteld moeten worden aan de database, omdat tot op heden geen uitslagen op SE gebied worden vastgelegd, maar alleen welk instrument gebruikt is. In plaats daarvan zijn we vorig jaar wel een oude kwestie nagegaan, te weten:

5. de vraag of er grote verschillen aan te wijzen zijn tussen het begrijpend lezen gedeelte van het Drempelonderzoek enerzijds en de andere tests op dit gebied (i.c. de Cito-toets) anderzijds.

Dit jaar hebben we nog een aantal andere kwesties opgepakt. Gedeeltelijk vloeien deze kwesties voort uit de veranderingen in de lijst zoals die vorig jaar zijn ingevoerd. Voor een ander deel zijn het oude vragen (die overigens aan actualiteit niet ingeboet hebben) die nu via gericht onderzoek in de databestanden in principe te beantwoorden zijn. Het gaat hierbij om de volgende vragen.

6. hoe ligt de frequentie van het gebruik van de diverse IQ-tests verdeeld over de functionele groepen die de Regeling onderscheidt: 55-75, 75-80, 80-90 en 90-120. Er zou een voorkeur voor individuele tests moeten bestaan bij een IQ tussen 75 en 80. Tevens is de beantwoording van deze vraag van belang voor vraag 7:
7. Om effecten te kunnen bestuderen van de veranderingen in de lijst voor de aanvragen die vanaf dit schooljaar plaatsvinden, is een beginmeting gedaan van met name de aanvragen die het afgelopen jaar tot 1 oktober 2007 zijn gedaan. We nemen aan dat dit jaar verschuivingen zullen plaatsvinden van bijvoorbeeld klassikale tests naar individuele tests in het bereik van IQ 55-80, dat er ingeval van leeszwakke meer gebruik zal worden gemaakt van beoordelingen door anderen en minder van zelfbeoordelingstests, etc.;
8. Er wordt een begin gemaakt met onderzoek naar effecten van toetsing door VO-scholen of toetsrapportage op basis van lvs gegevens zoals aangeleverd door de basisschool.

De waarde van de te verwachten bevindingen

Alvorens met de beantwoording van de vragen in de weer te gaan, een aantal opmerkingen vooraf.

Ten eerste, we hebben ons voor de vragen 1 t/m 5 beperkt tot de laatste vier jaargangen van het landelijke databestand. Gaat men verder terug dan krijgt men te veel gegevens over andere toetsversies en niet meer in gebruik zijnde instrumenten in handen, die de berekeningen nodeloos compliceren. Overigens levert deze beperking toch altijd nog zo'n 140.000 dossiers op waaruit we kunnen putten.

Ten tweede moeten we er altijd op verdacht zijn, dat we hier niet met een normale homogene populatie te maken hebben. Nemen we leerachterstanden als het centrale issue voor verwijzing naar LWOO of PrO waarop de leerlingen in eerste instantie door het bevoegd gezag voor een aanvraag zijn geselecteerd, dan zijn er heel verschillende oorzaken (intelligentie, huiselijke problemen, dyslexie, PDD-NOS, ADHD, andere SE

problematiek zoals faalangst, gepest worden, ga zo maar door) voor de leerachterstanden aan te geven die maken dat correlaties minder duidelijk kunnen zijn, verdelingen er anders uitzien e.d. Neem alleen al het geval, dat we alle leerlingen met een IQ tussen de 90 en 120 krijgen met leerachterstanden van meer dan anderhalf jaar. Leerlingen met normaliter bij dit IQ passende leerprestaties (namelijk gemiddelde leerprestaties) zitten niet in ons bestand. De bijzondere groep met wel grote leerachterstanden maar een gemiddeld IQ beslaan zo'n kleine 30 percent van alle leerlingen (plm. 40.000 leerlingen). Een dergelijk gegeven is van invloed op de hoogte van de correlatie tussen IQ en leerprestaties.

Ten derde hebben we geen zekerheid, dat de groepen die met elkaar voor een bepaalde variabele vergeleken worden op alle andere punten die ertoe doen vergelijkbaar zijn. Zo is het bekend dat het Drempelonderzoek meestal niet in de vertrouwde omgeving van de eigen klas plaatsvindt maar door de VO-school wordt afgenomen.

Wat hebben de twee laatste punten te betekenen voor de hoogte van de toetsuitslagen. In elk geval zal duidelijk moeten zijn, dat wanneer we conclusies willen trekken met betrekking tot consequenties voor de lijst, we dan eerst ook de mogelijkheid van allerlei triviale verklaringen de revue moeten laten passeren, voordat een meer prangende verklaring aan de orde kan zijn.

6.3. de klassikale vs. individuele IQ-tests

In het verleden heeft de commissie constant gewaarschuwd voorzichtig te zijn met het gebruik van schriftelijk af te nemen klassikale tests ingeval van grote leerachterstanden (met name op begrijpend lezen) of bij onvoldoende beheersing van het Nederlands. We zijn nagegaan hoe het gebruik van een specifieke IQ-test samenhangt met de leerprestaties van de leerlingen voor wie een beschikking is aangevraagd. We hebben hier specifiek gekeken naar IQ-scores rond de 60, 75 en 90 aangezien dit voor onze doeleinden opmerkelijke punten zijn op de IQ-schaal. We konden vaststellen dat rond IQ 60 er bij een schriftelijke test in termen van DLE een leerprestatie gehaald werd die veel hoger was en tot wel bijna 50 percent hoger kon oplopen op de diverse leerprestatietests, (bijvoorbeeld leerlingen die op de GIVO een score van 60 behaalden hadden gemiddeld een DLE van 29 op begrijpend lezen van de CITO, terwijl die leerlingen die op de SON-R of WISC-R een IQ van 60 behaalden, slechts gemiddeld een DLE van 20 scoorden op begrijpend lezen.) Bij een IQ van 75 bestaan dergelijke verschillen nog steeds maar zijn dan minder groot (zo'n 20 à 25 percent verschil in termen van DLE); en rond de 90 duiken de verschillen duidelijk naar rond de 5 à 10 percent. Alleen blijven de verschillen groot met de beide SON tests. Dit vindt mogelijk zijn verklaring in het feit dat men hier vaak te maken heeft met leerlingen die nog maar relatief kort in Nederland zijn en nog weinig onderwijs in Nederland hebben meegemaakt. Hoewel de leerlingen die met een individuele test getoetst zijn qua kenmerken zullen verschillen van die met de NIO e.d. zijn gemeten (bijv. lengte van zorggeschiedenis, verblijf op het speciaal onderwijs e.d.) kan dit niet zomaar als oorzaak voor een relatief hogere intelligentiescore worden geïnterpreteerd. Vooralnog lijkt dat de wijze van afname voor het gevonden verschil verantwoordelijk is.

We hebben dan ook gemeend deze bevindingen in de lijst tot uitdrukking te moeten brengen. Ten eerste zal dit mogelijk zijn door een onderscheid aan te brengen tussen een beoordeling per leerling-categorie, te weten LWOO dan wel Pro; en vervolgens om in geval van Pro aan te geven dat de beoordelingscategorie in dat geval IV is.

Enkele nuanceringsen zijn hier nog op zijn plaats. We zullen tot uitdrukking moeten brengen,

1. dat de GIVO het volgens de respondenten niet zo goed doet ook in geval van LWOO. Dit vinden we ook weerspiegeld in ons onderzoek met het databestand;
2. dat de SON-versies vooral geschikt zijn voor leerlingen die het Nederlands nog maar gedeeltelijk beheersen;
3. dat de NDT het beter doet dan de andere klassikale tests.

6.3.1. Onderhandelingsadviezen tussen scholen en RVC m.b.t. IQ-problematiek.

Er gaat heel wat kosten, tijd en moeite zitten in het verzamelen en bewerken van gegevens die nodig zijn voor een aanvraag bij de RVC-VO. Bovendien moet de hele procedure vaak voor een groot aantal leerlingen doorlopen worden in een korte spanne tijds. Als men dan als commissie een voorstel indient, dat op de genoemde grootheden een extra claim legt, moet men goed afwegen of de baten tegen de kosten opwegen. De onderhandelingsruimte zit wat ons betreft in het advieskarakter. Aan de ene kant kan een RVC hierdoor zijn kwaliteitseisen stellen en aan de andere kant kan het bevoegde gezag goede argumenten geven om in individuele gevallen toch van het advies in de lijst af te wijken. De commissie gaat er van uit dat de hoge kwaliteitseisen daar moeten gelden waar de keuze cruciaal kan zijn voor de verwijzing. Dit zou het geval kunnen zijn ingeval van beleid of strijdigheid van de gegevens en in het bijzonder daar waar de IQ gegevens niet aansluiten bij de rest van de gegevens over de leerling zoals blijkend uit andere tests, vroegere IQ testgegevens, onderwijskundig rapport e.d. Ook kan men als RVC niet meegaan met een via een categorie III of IV test verkregen IQ uitkomst dat onwaarschijnlijk geacht mag worden en de dossiervorming van de leerling negatief belast. Anderzijds is het niet zinvol een categorie I test te verlangen ingeval de aangeleverde gegevens harmoniëren met de andere aangereikte gegevens of onderbouwd kunnen worden met vroegere intelligentiegegevens over de leerling.

6.4. De CITO normtabel (2003)

In 2003 introduceerde het CITO een nieuwe normtabel voor de toetsen die de leervorderingen meten in het basisonderwijs. Op basis van vrij veel onderzoek en het toepassen van bijzondere statistische technieken was men in staat het voor het BaO gemakkelijker te maken door voor iedere leerling in de klas de leervorderingen met behulp van één testversie (voor welk leerjaar dan ook) vast te stellen, ondanks het feit dat de ene leerling op het niveau van groep 8 functioneert, terwijl de andere leerling nog druk doende is de leerstof van groep 4 onder de knie te krijgen. Hoewel deze hulp speciaal ontwikkeld is voor leerlingen die extra zorg nodig hebben (globaal dus het type leerling waarmee de RVC-VO op enig moment mee van doen heeft) kleven er toch belangrijke bezwaren aan de al te strikte doorvoering van deze normtabel.

Ten eerste zal duidelijk zijn dat men rekentechnisch wel kan vaststellen met een toets van midden groep 5 dat de leerling eigenlijk qua rekenen een niveau heeft van eind groep 6, maar men weet dan nog niet of het kind de typische leerstof van eind groep 5 en van groep 6 (bijvoorbeeld de staartdeling) beheerst.

Ten tweede – en nu hebben we het over onze doelgroep – moeten we erop bedacht zijn, dat leerlingen die op het niveau van groep 4 of begin groep 5 hun individuele leerlijn volgen met een toets van groep 7 en/of 8 niet alleen maar in verlegenheid worden gebracht omdat ze bijvoorbeeld de staartdeling nog niet in hun leerprogramma zijn tegengekomen, maar ook omdat ze er (voor de zoveelste keer?) mee worden geconfronteerd dat ze voor het onderwijs niet bepaald in de wieg gelegd zijn. Met andere woorden frustratie alom en waarschijnlijk met als gevolg dat de leerling gaat onderpresteren op de CITO-tests met normtabel 2003. Natuurlijk zouden we er eigenlijk van uit moeten kunnen gaan, dat het onderwijs de kloof tussen toetsversie en DLE wel niet te groot zal laten worden, zodat de sociaal emotionele schade nog mee zou kunnen vallen. U begrijpt het al: dit hebben we onderzocht. En dan word je niet bepaald gelukkig.

Samenvattend is de uitslag als volgt:

1. *begrijpend lezen*: van de 56.000 leerlingen krijgen 17.000 leerlingen een toetsversie die meer dan 2 jaar afwijkt van wat je op basis van hun vastgestelde DLE zou mogen verwachten; voor 22.000 leerlingen was er een verschil van een à twee jaar tussen toetsversie en vastgesteld DLE;

2. *spelling*: van de 64.000 leerlingen krijgen 7.000 leerlingen een toetsversie die meer dan twee jaar boven hun niveau ligt; voor 28.000 leerlingen was dit verschil 1 à 2 jaar;
3. *inzichtelijk rekenen*: 7.000 van 75.000 leerlingen krijgen een toetsversie die meer dan twee jaar boven hun niveau ligt; voor 28.000 leerlingen is dit een à twee jaar.

Met andere woorden, in 31.000 dossiers zijn leerlingen getoetst met toetsversies die zeer ver (meer dan 2 jaar) van deze leerlingen afstaan. In 78.000 dossiers was dit het geval voor toetsversies tussen de 1 en de 2 jaar. Wat is de verklaring voor dit niet optimale toetsen van leerlingen, die toch al uit het onderwijs dreigen te vallen? Met twee redenen moeten we ernstig rekening houden. Ten eerste zal het om het handelingsgemak van de leerkracht gaan. Nu er normtabellen zijn kan hij vertrouwend op het CITO toch de leervorderingen bepalen. Een tweede reden is, dat in diverse regio's de Inspectie van het onderwijs een toets op eindniveau schijnt te verlangen. Na veel discussie over hoe strikt we zouden moeten zijn en over hoe de restricties uit te drukken **is in de Commissie vastgesteld dat de toetsversie 10 DLE punten mag afwijken.**

6.4.1. (Onder)handelingsadviezen tussen aanvragers en RVC-VO met betrekking tot de Normtabel 2003

Vastgesteld moet worden, dat voor 2004 de Normtabel 2003 niet toegestaan was en dat men voor de tests waarvoor diverse toetsversies op basis van leerniveaus bestonden moest doortoetsen tot men binnen het DLE-bereik van de toetsversie was. Daarna gold het voorzichtige advies binnen de 20 DLE te blijven. Op grond van de vele afwijkingen die gevonden zijn, heeft de Certificeringcommissie en daarna RVC-VO Landelijk en het ministerie besloten om de grens bij 10 DLE afwijking te leggen. Ook hier komt het voor dat door de basisschool lvs-gegevens aangedragen worden die niet in overeenstemming zijn met de DLE-norm. Ook hier is onder bepaalde omstandigheden enige onderhandelingsruimte tussen RVC en de aanvragende school. De school is verantwoordelijk voor de aangedragen gegevens. Zij dient na te gaan of de gegevens in overeenstemming zijn met de in de lijst opgenomen voorwaarden, en zo niet in welke mate afgeweken wordt. Tevens dient nagegaan te worden in hoeverre op basis van bijvoorbeeld het OKR en de bijgeleverde lvs-gegevens de score op de niet toegestane toetsversie al dan niet noemenswaardig afwijkt van het totale beeld van de leerling. Uiteindelijk beslist ook hier de RVC-VO over de eventuele acceptatie van de onderhavige score. Technisch hoeft het geen probleem te zijn, omdat foutief verkregen scores op dit gebied niet door de data base geblokkeerd worden.

Er is bij de scholen en hun vertegenwoordigers enige beroering ontstaan over hoe de normtabel te hanteren en over hoe men door moet toetsen. De Certificeringcommissie heeft gemeend hiervoor door middel van het hier onder vermelde enige hulp te moeten bieden:

Toelichting bij het door- en terugtoetsen

Bij jaargroep gebonden toetsen zoals van het CITO-leerlingvolgsysteem of Aarnoutse dienen de RVC's uit te gaan van toetsen die aansluiten bij het werkelijke didactische niveau van de leerling. Dit betekent dat schoolverlaters op de basisschool soms toetsen moeten maken die in jaargroepen daarvoor gebruikt worden.

Adaptief toetsen

Kandidaten voor LWOO en Pro hebben leerachterstanden. De basisschool wordt aangeraden die leerlingen adaptief te toetsen. Neem dan toetsen af die de leerling qua niveau redelijk zou moeten kunnen maken. Een leerling met het niveau van groep 6 laat men niet ploeteren met de veel te moeilijke toets van groep 8. Dit is kindvriendelijker en de score is betrouwbaarder.

Door- of terugtoetsen

Als men bij gebruik van de toetsen van groep 8 een score bereikt die 10 of meer DLE punten lager ligt moet men terugtoetsen. Als men bij adaptief toetsen verkeerd geschat heeft en de uitslag 10 of meer DLE punten hoger of lager uitvalt moet met doortoetsen of terugtoetsen.

Voorbeeld: Een leerling behaalt op de M5 toets een score van dle 36. U verwachtte op grond van uw M5 keuze een score tussen dle 21 en dle 30 nu is het dle 36 geworden. (geen PrO score maar één die past bij LWOO). U legt ter nadere verifiëring van dit niveau nu de M6 of E6 toets voor om te bezien waar het niveau nu op uit komt. U voegt beide toetsscores toe aan het onderwijskundig rapport.

Tegengesteld werkt het niet anders, een M5 score en nu met een tegenvallende score van dle 15 vraagt van u om terug te toetsen of het echt groep 4 niveau is in dit voorbeeld graag met een E4 toets. Aan het eind van deze notitie is een tabel ter advisering hoe door en terugtoetsen opgenomen.

Bij deze werkwijze hoeft u niet drie keer door te toetsen. U kiest eerst op grond van uw verwachting. De eventueel tweede toets die u kiest sluit aan op het behaalde niveau. Die twee gegevens zullen bijna altijd voor elke RVC voldoende zijn. Wel is het van groot belang dat u precies de toetsinstructies van de betreffende toetshandleiding volgt.

Bovenstaande werkwijze voor toets keuze en het eventuele moeten doortoetsen is aan de orde bij de volgende toetsen:

- ✓ Cito leestechiek en leestempo
- ✓ Cito begrijpend lezen A en B en begrijpend lezen toets 1 t/m 4.
- ✓ Cito schaal vorderingen in spellingvaardigheid 1, 2 en 3 alle dictees van E3A/B t/m M8 A/B
- ✓ Cito rekenen algemeen 1996, Rekenen-Wiskunde CITO, 2002
- ✓ Spelling en begrijpend lezen toetsen van Aarnoutse

Advies tabel door- terug toetsen E toetsen CITO-LVS en toetsen Aarnoutse

| Niveau | Behaald dle | advies |
|----------|-------------|---|
| Groep E4 | < 10 | Kies een Toets van groep 3 |
| Groep E4 | > 30 | Kies een Toets van groep 5 |
| Groep E5 | < 20 | Kies een Toets van groep 4 ¹ |
| Groep E5 | > 40 | Kies een Toets van groep 6 |
| Groep E6 | < 30 | Kies een Toets van groep 5 ¹ |
| Groep E6 | > 50 | Kies een Toets van groep 7 |
| Groep E7 | < 40 | Kies een Toets van groep 6 ¹ |
| Groep E7 | > 60 | Doortoetsen niet nodig score is buiten LWOO criteria. |

Advies tabel door- terug toetsen M-toetsen CITO-LVS .

| Niveau | Behaald dle | advies |
|----------|-------------|---|
| Groep M4 | < 5 | Kies een Toets van groep 3 |
| Groep M4 | > 25 | Kies een Toets van groep 5 |
| Groep M5 | < 15 | Kies een Toets van groep 4 ¹ |
| Groep M5 | > 35 | Kies een Toets van groep 6 |
| Groep M6 | < 25 | Kies een Toets van groep 5 ¹ |
| Groep M6 | > 45 | Kies een Toets van groep 7 |
| Groep M7 | < 35 | Kies een Toets van groep 6 ¹ |
| Groep M7 | > 55 | Kies een Toets van groep 8 |
| Groep M8 | < 45 | Kies een Toets van groep 7 ¹ |

¹ Is de score lager dan het voorafgaande leerjaar dat u adaptief had gekozen, dan is het advies: neem de toets van het leerjaar waar de behaalde dle score naar verwijst. Dit kan twee of meer leerjaren lager worden als uw start niveau veel te hoog is geweest.

6.5. De schriftelijke zelfbeoordelvragenlijsten voor de vaststelling van de SE-problematiek

In het Visiestuk, gepresenteerd in de Landelijke Vergadering van de RVC-VO's van 14 februari 2007 is melding gemaakt van een veel gehoorde klacht onder RVC-leden en diagnostici dat de schriftelijke zelfbeoordelvragenlijsten veelal veel te moeilijk zijn voor leerlingen met grote leerachterstanden. Aangezien we volgens de Regeling alleen te maken hebben met leerlingen met een IQ van tussen de 90 en 120 die dergelijke vragenlijsten worden geacht in te vullen, moeten we deze klacht vooral zien in het licht van de leerachterstanden met betrekking tot begrijpend lezen. We zijn dan ook in het databestand nagegaan om hoeveel leerlingen het hier ging in de afgelopen vier jaren en hoe het met hun leesvaardigheid stond. Alvorens de resultaten van het onderzoek te geven, eerst nog dit. Er zijn vragenlijsten die aangeven dat leesniveau midden groep 7 nodig is om de lijst naar behoren in te kunnen vullen. In het primacohort onderzoek (in het kader van het onderwijsachterstandenbeleid) is ook welbevinden op school meegenomen. Ook hier bleek dat men vóór groep 7 sociaal emotionele problematiek niet goed met leerling-vragenlijsten aan kon pakken. Het nog tamelijk recent verschenen CITO instrument (VISEON) heeft ook op basis van onderzoek vastgesteld dat men pas in de bovenbouw –en dan alleen bij leerlingen die daar aan toe zijn- zelfbeoordelvragenlijsten kan afnemen. DLE 45 lijkt dus niet zo gek om als richtpunt aan te houden. In het bestand staat 50.000 keer het gebruik van een SE vragenlijst genoemd voor leerlingen met een IQ tussen de 90 en 120. 38.000 leerlingen hebben een zelfbeoordelvragenlijst ingevuld.

1. Zo'n 1.100 leerlingen hadden een DLE voor begrijpend lezen van 15 of minder (niveau groep 3 tot midden 4);
2. Een kleine 9.000 leerlingen hadden een niveau van 15 tot 30 (midden groep 4 t/m groep 5);
3. Een kleine 22.000 leerlingen hadden een niveau van 30 tot 45 (groep 6 tot midden groep 7);
4. Zo'n 6.000 leerlingen lazen op niveau hoger dan DLE 45.

Het DLE-gemiddelde van de leerlingen met een IQ tussen de 90 en 120 die zelf een vragenlijst hebben ingevuld was 36.6. Wanneer we dit laatste gegeven vergelijken met het gemiddelde DLE van de leerlingen voor wie de leerkracht of de ouders een vragenlijst hebben ingevuld, dan stellen we enigszins verbaasd vast, dat dit gegeven merkwaardig identiek is: 36.7. Blijkbaar is in de praktijk de leesvaardigheid geen beslissend criterium bij de selectie van de geschikte vragenlijst voor een individuele leerling. Volgens onze berekeningen moeten we voor 32.000 van de 50.000 SE-gegevens vraagtekens zetten bij de betrouwbaarheid van de uitkomsten.

Om aan deze ongewenste situatie te ontkomen stelt de commissie voor beperkingen in te stellen ten aanzien van het gebruik zelfbeoordelvragenlijsten door leerlingen. Het voorstel is: *Wanneer een leerling een DLE voor begrijpend lezen van 40 of meer heeft, dan kunnen alle zelfbeoordelvragenlijsten voor SEP worden gebruikt. Wanneer deze DLE tussen de 30 en de 40 ligt moet de onderzoeker nagaan of een zelfbeoordelvragenlijst wel een juiste keuze is. Onder DLE 30 wordt zeer sterk afgeraden een zelfbeoordelvragenlijst te gebruiken voor de vaststelling van SE problematiek.*

6.5.1 Handelingsadviezen voor de aanvragende school en de beoordelende RVC

Het is niet denkbeeldig, dat er al SE-vragenlijsten zijn ingevuld door de leerling bij andere gelegenheden en voordat met de leesvaardigheid rekening gehouden is. Stel dat de leesvaardigheid niet toereikend is om zomaar de scores te vertrouwen. In dat geval is het veel moeilijker om vast te stellen of de zo verkregen gegevens wel in lijn zijn met de overige gegevens in het dossier. Stel de leerling vindt zichzelf wel degelijk (sociaal) adequaat terwijl elders de leerkracht het tegenovergestelde beweert. Mag men nu zonder

meer stellen, dat de leerling de vragen niet begrepen heeft, of moeten we er rekening mee houden dat de leerling zichzelf wel degelijk adequaat vindt, wat op zich weer inadequaat kan zijn. We weten het niet. In dit geval is het beter dat de onderzoeker voor het psychologische rapport gebruik maakt van gegevens uit het onderwijskundig rapport, van beoordelingslijsten die door ouders en/of leerkracht zijn ingevuld en van gegevens op basis van eigen observatie en zich niet te wagen aan een interpretatie van de gegevens van de zelfbeoordeling van de leerling.

6.6 Drempelonderzoek vs. Cito-toets Begrijpend lezen.

Vorig jaar is gebruik gemaakt van het onderzoek van Melis² naar de verschillen tussen de uitkomsten op de Cito-toets Begrijpend lezen en het Drempelonderzoek op dit punt. Het ging hierbij om kinderen die twee keer getest zijn, zodat een goede vergelijking mogelijk was. Op grond van de uitkomsten van dit onderzoek werd aan het Drempelonderzoek Begrijpend lezen een III in de categorie indeling toegekend. De data waren echter van slechts relatief weinig kinderen (plm. 450 leerlingen) afkomstig. Ook kon men vragen stellen bij de representativiteit van de leerlingen. Alle leerlingen waren afkomstig uit één regio; mogelijk zijn de leerlingen van wie twee testgegevens worden verzameld geen doorsnee leerlingen van onze doelgroep, mogelijk zitten er bijzondere redenen achter het feit dat ze hertest zijn, en wel buiten de vertrouwde omgeving van de basisschool en hebben ze op een vrije middag een hele batterij tests moeten doen etc.

Nu we gebruik kunnen maken van het hele databestand kunnen de bovengenoemde bezwaren gedeeltelijk worden weggenomen: de leerlingen zijn uit heel het land afkomstig, we hebben hier de beschikking over grote aantallen en we kunnen de IQ scores variëren om achter de representativiteit van de gegevens in dit opzicht te komen. Ook weten we dat het drempel onderzoek niet alleen door de VO-scholen in een gecombineerde toetsessie wordt afgenomen, maar ook dat het onderzoek op de basisschool wordt gebruikt. Nadelen zijn, dat we geen twee meetgegevens van dezelfde kinderen hebben, en dat we ook niet zomaar weten in welke mate de twee groepen leerlingen op de overige kenmerken vergelijkbaar zijn.

De leerlingen uit dit onderzoekje zijn plm. 60.000 leerlingen die een Cito Begrijpend Leestoets gedaan hebben en 49.500 leerlingen die het Drempelonderzoek hebben afgelegd. Samen meer dan 80 percent van alle onderzochte leerlingen. Het merkwaardige feit doet zich voor, dat beide toetsen qua DLE boven het gemiddelde van alle leerlingen over alle toetsen liggen (CITO 32.8, Drempel 32.1 en het gemiddelde is 31.9). Dit ligt vooral aan andere CITO toetsen (samen zo'n 6.000 leerlingen) die beduidend onder het DLE gemiddelde scoren (Clib 30.8 en Cito Toelatingsonderzoek 30.6). De Clib is uit de lijst verdwenen. Voor het CITO Toelatingsonderzoek zal nagegaan worden of dit consequenties heeft voor de categorisering in de lijst. Het Drempelonderzoek bevindt zich dus in goed gezelschap Cito BL. Ze scoren hoger en de twee andere lager.

Beide groepen zijn opgedeeld in sprongen van vijf punten IQ (55-60, 60-65, etc.). In dat geval zien we dat Cito BL t/m IQ 95 een ietsjes hoger DLE (maximaal twee DLE punten) heeft dan het Drempelonderzoek. De gemiddelde DLE's per sprong van vijf IQ punten leidt nooit tot andere beoordelingen in termen van wel/niet LWOO en tussen LWOO en PrO. Op deze gronden is het niet opportuun om in het categorieënsysteem anders te behandelen dan de Cito-toetsen. Er wordt wel aangevoerd, dat de Drempelgegevens vooral van het Voortgezet onderwijs komen, dit in tegenstelling tot de Cito-toetsen die in het primair onderwijs zijn afgenomen waardoor de verschillen zouden kunnen worden verklaard. Het Drempelonderzoek is minder vertrouwd, in een andere omgeving afgenomen en dergelijke. Door deze minder vertrouwde context zouden de resultaten lager uitkomen. Dit is echter waarschijnlijk geen goede verklaring. Want daarmee kan niet verklaard worden dat Drempelonderzoek rekenen hogere DLE scores heeft dan de Cito toetsen.

² Exploratief onderzoek Begrijpend lezen Drempelonderzoek en NDT; Gerard N. Melis 2004. Interne notitie RVC Noord-Brabant

Al met al is er te weinig grond om het Drempelonderzoek Begrijpend lezen de categorie III te geven. Er zijn dan ook geen handelingsadviezen nodig omdat alle beperkingen voor het Drempelonderzoek begrijpend lezen vervallen.

6.7 Het gebruik van de diverse IQ-tests in de relevante segmenten tussen IQ 55 en 120

In het verleden hebben we (zie De Greef en Van der Geest, 2005) onderzoek gedaan naar het gebruik en de effecten van diverse IQ-tests. Aanleiding was dat de voorlaatste versie van de NDT veel hogere uitslagen aan de bovenkant van de IQ-metingen vertoonden en wel zodanig dat veel meer leerlingen boven de 90 scoorden waardoor SE onderzoek noodzakelijk werd. Vastgesteld werd dat genoemde toets inderdaad een significante neiging vertoonde om hogere scores op te leveren: ondanks mindere leerprestaties waren de IQ-scores hoger en dat vooral in de hogere IQ prestaties van de doelgroep. Vorig jaar hebben we op een meer directe manier in de database onderzocht hoe de relatie tussen gebruikte IQ-instrumenten en leerprestaties ligt (zie paragraaf 6.3). We hebben daar vastgesteld dat men voor de klassikale tests hogere leerprestaties moet halen om een met de individuele tests vergelijkbare IQ score te verkrijgen. Dit betekent dat er van een zeker contaminatie-effect sprake is bij klassikale tests met de onderwijsprestatie: in die zin dat negatieve leerprestaties in dat geval de IQ-score negatief beïnvloeden. De nieuwste versie van de NDT is in dit geval een test die zich bijna als een individuele test gedraagt. Dit vastgesteld hebbende, kunnen we ons afvragen in welke mate welke tests in de voor ons relevante segmenten van de IQ scores worden ingezet.

In het segment van 55 tot 75 worden de klassikale tests in meer dan 80% van de gevallen ingezet. Hoofdverantwoordelijke is hier de NIO met 60% en daarna de NDT met 20% (Dit laatste is minder verontrustend, zie de opmerkingen over de NDT hierboven). In het voor onze beslissingen cruciale gebied tussen 75 en 80 waar we beslissingen moeten nemen over LWOO dan wel PrO zou men verwachten van de kant van de diagnosticus, dat hij het zekere voor het onzekere neemt door in dergelijke gevallen regelmatig individueel te toetsen. Dit blijkt echter niet het geval: slechts zo'n 7% wordt met een individuele test onderzocht. In dit segment zijn alle klassikale tests oververtegenwoordigd met uitzondering van de GIVO. Ook hier speelt de NIO en in mindere mate de NDT de boventoon.

De cijfers voor het jaar dat nu aan de gang is hebben we niet meegenomen. We mogen hopen dat de situatie zich ten positieve zal ontwikkelen.

6.8 De mate van gebruik van de diverse instrumenten

Sinds jaar en dag geven diagnostici en RVC's van een aantal instrumenten aan, dat ze weinig of niet gebruikt worden. De commissie is onzeker over hoe met dergelijke instrumenten om te gaan. Enerzijds zou men kunnen stellen, dat ze niemand in de weg zitten; anderzijds kan men ook stellen, dat de lijst op die manier een schijnrijkdom vertoont: door het nauwelijks gebruikt worden van de test bestaan er ook geen ervaringen in het veld met deze instrumenten waardoor we ook niet in staat zijn een geschiktheids- en bruikbaarheidsoordeel af te geven. De commissie is nagegaan in welke mate de diverse instrumenten in de praktijk gebruikt worden. **Met de bevindingen heeft ze nog niets gedaan met betrekking tot de lijst. Misschien dat hierover discussie in de landelijke vergadering kan plaatsvinden.**

De bevindingen kunnen ook meegenomen worden in ons onderzoek naar de effecten van de vorig jaar doorgevoerde veranderingen in de lijst.

6.9 Onderzoek naar dubbele toetsing op het gebied van de leerachterstanden

Voor de aanvraag naar LWOO of PrO geldt in principe dat de meest recente toets- en toetsuitslag op het gebied van de leerachterstanden opgegeven moet worden. Het komt echter meer dan eens voor, dat aanvragers zelf al melden dat bepaalde toetsen en toetsuitslagen niet betrouwbaar zijn blijkend uit de omstandigheden waaronder de toets is afgenomen. RVC's stellen nogal eens vast (o.a. blijkend uit een recente discussie in de landelijke vergadering) dat de door de aanvragende school aangeleverde scores niet altijd in overeenstemming zijn met gegevens uit het onderwijskundig rapport waaronder de lvs-gegevens. In de vergaderingen van de Certificeringcommissie is deze problematiek ook sinds jaar en dag aan de orde. Geopperd werd daar meerdere malen dat in het onderwijskundig rapport de lvs-gegevens verplicht worden meegeleverd. Zover zijn we nu nog niet. Er is ondertussen wel een kleine enquête gehouden onder de verschillende RVC's en hun kamers over de daar heersende praktijk. Het gaat dan meer om vermoedens en niet zo zeer om harde bewijzen. In deze enquête (24 responsen) kwam naar voren dat 23 van de 24 respondenten met het fenomeen vertrouwd waren (de 24^e respondent had deze vraag niet ingevuld). Van de respondenten geven er 8 aan dat dit minder dan 10% voorkomt. De rest geeft aan dat dit meer voorkomt. Vijf respondenten zeggen zelfs: in meer dan 50% van de gevallen.

De tweede vraag was interessanter. Als u deze scores van dezelfde leerling vergelijkt, welke scores zijn dan hoger, die van het lvs of die van het VO? De antwoorden zijn mogelijk nog interessanter: 18 respondenten geven aan dat de scores van het VO meestal lager zijn dan die van het lvs; 4 zeggen dat ze meestal overeenkomen en 2 hebben deze vraag niet ingevuld. We zullen echter over objectieve kwantificeerbare uitkomsten moeten beschikken, willen we met een voorstel aan de vergadering komen. We hebben vastgesteld dat 18 respondenten bereid zijn enig meerwerk te verrichten door zowel de gegevens van de centrale toetsdag van het VO en de toetsuitslagen van het Bao in te vullen.

7. Verantwoording van die aspecten op de lijst die afwijken van vorige jaren

Overal in de lijst waar men met de criteria zoals genoemd en uitgewerkt in paragraaf 4 kan volstaan, zal in deze verantwoording aan het instrument voorbijgegaan worden. Telkens wanneer er wat nieuws en of bijzonders aan de hand is (zoals een bijzondere opmerking die nog niet in de vorige lijst stond, een nieuwe toets op de lijst, of een afwijking ten opzichte van het vorig jaar) zal daarvan verantwoordend gewag worden gemaakt. Omdat we het vorig jaar in de organisatie en het explicieter maken van de lijst en hoe de instrumenten te gebruiken nogal ingegrepen hebben, worden ook die wijzigingen meegenomen.

7.1. De intelligentietests

Vorig jaar was een aantal grote wijzigingen te melden, die cruciaal waren en dit jaar mogelijk tot behoorlijke veranderingen in de toetspraktijk geleid hebben. Ten eerste hebben we een fundamentele scheiding aangelegd tussen LWOO en PrO wat de beoordeling betreft. Dit is in zoverre een verandering, dat we ook voorheen al ten stelligste aangeraden hebben om klassikale tests voor potentiële PrO-leerlingen het liefst te vermijden. Vorig jaar zijn we daarin verder gegaan door de bruikbaarheid van deze tests per doelgroep van een categoriebeoordeling te voorzien die in geval van PrO een III of een IV betekende. De redenen hiervoor hebben we eigenlijk al in de hoofdstukken 5 en 6 gegeven en met name in paragraaf 6.3: als de leerprestaties ver achterblijven, zoals bij potentiële PrO-leerlingen het geval is, dan blijken IQ prestaties op schriftelijke/klassikale tests veel lager uit te komen dan in mondelinge/individuele tests. Dit kan alleen verklaard worden uit het feit dat in het eerste geval niet alleen cognitieve capaciteiten, maar ook (gebrek aan) schoolsucces gemeten wordt. Hierdoor worden deze leerlingen benadeeld in de dossiervorming en de school en de RVC krijgt mogelijk een

fout beeld van de leerling. Hoe met deze III en IV in de toetspraktijk en aanvraagprocedure om te gaan is ook in het voorgaande, m.n. in par. 6.3.1 aangegeven.

Ten aanzien van de WISC- III ^{NL}

Na een turbulente introductiegeschiedenis is de WISC- III ^{NL} dan eindelijk twee jaar geleden door de COTAN van een gunstig oordeel voorzien. Op de door de COTAN gehanteerde criteria scoorde de test toen overal voldoende of hoger met helaas met één uitzondering: de criteriumvaliditeit was onvoldoende en wel omdat dit onderzoek nog niet was uitgevoerd. Op zich was dit voor ons een groot probleem, omdat we juist van de IQ-tests eisen (zie o.a. par. 3.1) dat het criterium validiteit dan wel prognostische capaciteit van een IQ-test in orde is. Er waren echter een paar verzachtende omstandigheden. Ten eerste was dit criterium ook voor zijn voorganger de WISC-R als onvoldoende aangemerkt en wel om de zelfde reden. Ten tweede, we hadden destijds voor het handhaven van de WISC-R gekozen, omdat het a. om de enige test ging die individueel afgenomen kan worden en daarnaast ook nog expliciet aan de verbale factor aandacht besteedt, b. deze test een schat aan diagnostische gegevens op kan leveren, wat speciaal van belang kan zijn voor die kinderen die bijzondere zorg meer dan nodig hebben en c. in onderzoek aannemelijk is gemaakt dat de lagere IQ's nauwelijks van het Flynn-effect profiteren. Ten derde is door makers en uitgever toegezegd, dat men onderzoek zal uitvoeren naar de criteriumvaliditeit van de test, zodat we over enige tijd over deze gegevens kunnen beschikken en nieuwe afwegingen kunnen maken. We hebben er toen dan ook voor gekozen de WISC-R na jarenlang trouwe dienst op te geven en dat we ondanks het blijvende probleem van de criteriumvaliditeit voor de WISC- III ^{NL} moesten gaan. Het ging om overigens vergelijkbare tests, waarbij de uitvoering van de nieuwe test moderner oogt en de normen van recente datum zijn. Toch hebben we toen gemeend, slechts een tijdelijke goedkeuring van twee jaar af te moeten geven. Van het moderne IQ-test onderzoek mogen we toch eisen dat controleerbare uitspraken over de criteriumvaliditeit worden gedaan. Die 2 jaar zijn inmiddels voorbij. Vandaar dat de Commissie zich zorgen maakt over hoe het verder moet. Het gaat namelijk niet aan om een toets op niet eigenlijke gronden te handhaven. Tot op heden hebben we niet de indruk gekregen dat ook maar een begin is gemaakt met het bepalen van de criteriumvaliditeit. Het gaat niet aan ons bij deze situatie neer te leggen. De commissie beveelt dan ook aan initiatieven te ontplooiën die aan deze onbevredigende situatie een einde maken.

SON-R en SON-R Verkort

Uit het onderzoek van het databestand kon men afleiden, dat de SON tests vooral gebruikt worden in geval de leerling het Nederlands nog niet voldoende beheerst. De correlatie met leerprestaties is niet bijzonder hoog, de spreiding daarentegen wel en de schoolprestaties blijven soms achter. De verklaring hiervoor is dat het hier gaat om leerlingen die nog niet zoveel onderwijs in Nederland of in het Nederlands gevolgd hebben of dit onderwijs niet optimaal konden volgen ten gevolge van nog op gang zijnde tweede taalverwerving. Bovendien wordt het verbale vermogen alleen maar indirect vastgesteld. Om deze aspecten tot uitdrukking te brengen is aan de SON in beide gevallen een dubbele categorisering gegeven. Ingeval van achterstanden in de taalverwerving hebben we de SON-R een II gegeven (de test doet het goed volgens de enquête en het databestand, maar hij is verouderd) en de SON-R Verkort een I. In het geval dat er van genoemde achterstanden geen sprake is, maar wel van onderwijsprestaties op PrO niveau gaan we voor beide instrumenten uit van een III.

De klassikale tests

De klassikale tests zijn vorig jaar allemaal voor de potentiële LWOO-leerlingen op de categorie I gezet behalve de GIVO vanwege het niet helemaal voldoen in de doelgroep (categorie III) en de NSCCT. Deze laatste blijft voor LWOO als PrO op de categorie 0 staan, aangezien hij zo weinig gebruikt wordt, dat we ons geen oordeel over de bruikbaarheid kunnen vormen. Voor PrO gaan we niet zo ver, aangezien het ook hier om een klassikale test gaat. Zoals gezegd, om de hierboven besproken redenen krijgen de

klassikale tests voor het PrO gedeelte allemaal een beoordeling IV. Hier geldt één uitzondering, namelijk de NDT die het in dit segment beter doet en daarom de beoordeling III meekrijgt.

Alhoewel de IVO bezig is om de normen voor het hele schooljaar geschikt te maken, zijn hierover bij ons nog geen beoordelingsgegevens van de COTAN bekend. Voor IVO en de internetversie daarvan (te weten de TPVO-IVO) blijft daarom tot nader order de beperking bestaan, dat hij alleen van januari tot en met mei afgenomen mag worden. Daarnaast moet ook voor de TPVO-IVO voor de PrO leerlingen een IV aangehouden worden.

Drempeltest, 2007

Er is een nieuw IQ-instrument te melden, die aan alle criteria van de COTAN heeft voldaan, te weten de Drempeltest uit 2007. Omdat er nog geen gebruiksgegevens bekend zijn zouden we voor zowel LWOO als voor PrO een categorietoekenning van 0 moeten geven. Voor LWOO doen we dat ook. Voor PrO lijkt dat echter niet gewenst aangezien het om een klassikale test gaat.

7.2. Toegestane instrumenten Criterium SE.

De schriftelijke zelfbeoordelvragenlijsten

De opmerkingen die hieronder volgen hebben betrekking op het restrictieve gebruik van YSR, CBSK, NPV-J 2005, PMT-K 1983, SVL, SSAT, VAK en tenslotte de leerlingenvragenlijst van de VISEON.

In de lijst voor de toegestane middelen voor de Sociaal Emotionele problematiek zelf is niet zo heel veel veranderd. De grote verandering zit in het bruikbaarheidsoordeel van de Commissie ten aanzien van de schriftelijke zelfbeoordelvragenlijsten, zoals de NPV-J, SVL en dergelijke. Zoals we in paragraaf 6.5 hebben uiteengezet, zijn dergelijke lijsten normaal toegesneden op de leesvaardigheid van midden groep 7, of heeft men in onderzoek vastgesteld dat voor groep 6 of lager (zie ook VISEON) men geen betrouwbare gegevens met dergelijke gegevens in handen krijgt. Op basis van onderzoek hebben we vastgesteld, dat men in de praktijk de leesvaardigheid van de leerling niet meeneemt als selectiecriterium bij de bepaling welk instrument men voor een leerling zal gaan inzetten. Omdat we in het verleden er geen oog voor konden hebben hoe het met de leesvaardigheid van leerlingen is gesteld die voor de dossiervorming en aanvraag een vragenlijst moeten invullen, hebben we tot dusverre geen duidelijk afgebakende criteria opgesteld. Nu blijkt dat de betreffende leerlingen ondanks een IQ van boven de 90 in 65 procent van de gevallen niet over een voldoende vaardigheid met betrekking tot begrijpend lezen beschikken zijn we genooddaakt grenzen aan de bruikbaarheid van zelfbeoordelvragenlijsten te stellen. We hebben de leesvaardigheid daarom als criterium in de lijst opgenomen. Wanneer een leerling een DLE voor Begrijpend Lezen (BL) heeft van 40 of meer, gaan we er zonder meer van uit dat de leerling in principe elke zelfbeoordelvragenlijst kan begrijpen en adequaat kan invullen. Wanneer de DLE voor BL tussen de 30 en 40 ligt moet de gedragswetenschapper eerst nagaan of een dergelijke vragenlijst een juiste keuze is en moet er tevens een ander type vragenlijst worden ingevuld (vragenlijst voor ouders en/of leerkracht) en beoordeeld. Onder een DLE van 30 voor BL wordt ten stelligste afgeraden een schriftelijke zelfbeoordelvragenlijst in te laten vullen en dus mee te leveren met de aanvraag.

AVL- 2002/DVL- 2006

De AVL-2002 hadden in het verleden een III. Hoewel er enige kritiek op deze lijst is en wel met name op het relatief kleine aantal items, en als screeningsinstrument dus eigenlijk alleen naast andere instrumenten zou mogen worden gebruikt, was de frequentie van kritiek niet van dien aard dat dit een III rechtvaardigde. De AVL-2002 komt hierdoor op een I. Daarmee kunnen we hier volstaan. Het lijkt vanzelfsprekend dat dit ook voor de DVL (de t.o.v. de AVL iets gewijzigde internetversie) zou moeten gelden. De DVL blijft voorsnog op een 0 staan omdat dit instrument nog nauwelijks gebruikt is en daarmee de gebruikerservaring ontbreekt op het moment van het vaststellen van deze lijst.

CBCL/TRF/YSR

Ook bij deze vragenlijsten geldt, dat ze als screeningslijsten alleen als onderdeel van een breder persoonlijkheidsonderzoek mogen worden ingezet. Men dient daarnaast nog een ander instrument uit de lijst voor SE-problematiek te gebruiken.

CBSK en CBS voor Adolescenten.

We hebben deze vragenlijst toegevoegd omdat de CBSK slechts gebruikt mag worden tot 12 jaar.

SVL, PMT-K

Deze tests zijn verouderd (resp. van 1990 en 1983). Het gaat hier echter om twee heel belangrijke en veel gebruikte tests voor de indicatiestelling, waarvoor bovendien geen goede alternatieven voorhanden zijn. Met name geldt dit voor de PMT-K. Naast het feit dat met dit instrument uitspraken gedaan kunnen worden over direct relevante sociaal emotionele aspecten die voor schoolsucces indicatief zijn en die in de Regeling expliciet worden genoemd, wordt ook de sociale wenselijkheid van de beantwoording door de leerling in kaart gebracht. Hierdoor zijn we bijvoorbeeld beter in staat vast te stellen in welke mate andere vragenlijsten voor de leerling betrouwbare gegevens zullen bevatten. Voor de PMT-K geldt bovendien dat dit instrument (samen met de NPV-j) het meest gebruikt wordt (30%). Voor de SVL geldt dit laatste minder (7%). Tenslotte: de mening van de ondervraagden over beide instrumenten is zeer positief: 80 % vindt de tests geschikt voor LWOO-leerlingen. Dit bij elkaar genomen, betekent dit dat we deze twee tests alsnog moeten handhaven in afwachting van een nieuwe normering, dan wel een beter alternatief. Het blijft **categorie II. Er moeten door belanghebbenden initiatieven ondernomen worden deze tests te actualiseren of van deugdelijke vervangers te voorzien.**

SAQI 2007

Voor de SVL is al een begin gemaakt: de Saqi is als de internetversie van de SVL geïntroduceerd. R zijn nieuwe normen, maar opzet en inhoud zijn als die van de SVL. We gaan er dan ook van uit dat deze vragenlijst zijn weg naar de gebruikers wel zal vinden. We stellen daarom de categorie I voor.

SEV

Er is een nieuwe vragenlijst met een zeer groot leeftijdsbereik ontwikkeld, met normen van 2005: de Sociaal Emotionele Vragenlijst (SEV). In dit instrument zijn die vragen meegenomen die voor ons werk essentieel zijn. De vragenlijst is door de COTAN op alle relevante punten positief beoordeeld. De test wordt ondertussen veelvuldig gebruikt. Dit betekent dat we het instrument met de categorie I in de lijst kunnen opnemen.

7.3. De leerachterstanden.

In hoofdstuk 6, en met name in de paragrafen 6.4 over de normtabel en 6.6 over het Drempelonderzoek Begrijpend Lezen, hebben we al een aantal veranderingen in de lijst met betrekking tot de leerachterstanden besproken. Het gaat in die paragrafen over de grootste veranderingen in dit deel van de lijst. We zullen hieronder per subcategorie nagaan welke veranderingen er ten opzichte van de vorige lijst zijn doorgevoerd. Alleen hier stellen we voor alle gevallen waar bij de test in de Opmerkingen "Normtabel 2003" staat dat daar overal voor geldt dat er geen groter verschil tussen het genormeerde afnamemoment van de toetsversie en vastgesteld DLE mag zijn dan 10 DLE punten. D.w.z. dat er een range is van 20 DLE punten waartussen de toetsversies zich mogen bewegen. Hoe ten aanzien hiervan te handelen, staat vermeld in paragraaf 6.4.1.

1. Technisch lezen

➤ Niveautoets & NiveautoetsPlus, CITO, 2007

De grootste verandering is wel de invoering van de Niveautoets & NiveautoetsPlus (CITO, 2007). Deze toets is speciaal ontwikkeld voor leerlingen met achterblijvende

schoolprestaties van minimaal anderhalf jaar, voor wie de CITO Eindtoets te moeilijk zou zijn. Het gaat hierbij dus precies om onze doelgroep. De Cotan heeft positief geoordeeld over deze test op alle voor ons relevante punten. De test kan dus opgenomen worden met categorie aanduiding 0. De commissie heeft maar één bedenking en die betreft de afnamedatum. De afname zal op de datum van de CITO Eindtoets vallen. Dat is vrij laat om op tijd (d.w.z. voor de zomervakantie) de aanvraagprocedure en indicatiestelling rond te hebben. Nu al zien we in vele dossiers in het onderwijskundig rapport veelvuldig de mededeling opduiken: 'gegevens Cito eindtoets worden nagezonden'. Vaak zal pas na de gegevensverzameling van de leerachterstanden duidelijk worden of en zo ja welke aanvullende gegevens verzameld dienen te worden. We zullen de ervaringen volgend jaar moeten afwachten. Op grond van de ervaringen en het gebruik van de eindtoetsen mogen we van categorie I uitgaan

➤ *Van beoordeling 0 naar I*

Het Drempelonderzoek en Technisch Lezen 345678, 90 A/B waren vorig jaar van een beoordeling 0 voorzien. Op grond van het gebruik kunnen we nu aan deze instrumenten een beoordeling I geven.

Drempelonderzoek.

Voor het Drempelonderzoek geldt wel dat de normering voor de 1^e klas VO onvoldoende bevonden is. Vandaar dat het bereik tot de basisschool beperkt blijft.

2. Ten aanzien van begrijpend lezen:

➤ *De begrijpend leestoetsen van Aarnoutse van groep 4 t/m 8*

Vorig jaar had dit toetswerk al veel kritiek te duchten. Dit had onder andere te maken met het feit dat men vanwege het kleine DLE bereik veel moest doortoetsen. In het onderzoek in de databestanden hebben we bovendien vastgesteld, dat deze toetsen vooral in het geval van PrO aanvragen veel grotere leerachterstanden opleverden dan de concurrerende toetsen op dit gebied. Om die reden is de categoriebeoordeling voor PrO verlaagd naar een IV. **Bovendien is de test dit jaar voor het laatst toegestaan vanwege het feit dat de normen verouderd zijn, de toets zo veel kritiek te duchten kreeg en (waarschijnlijk daardoor) ook nog maar weinig gebruikt werd.**

➤ *Drempelonderzoek en Begrijpend lezen 345678*

Bij het drempelonderzoek moeten we even terugverwijzen naar hoofdstuk 6. (par. 6.6.). Daar stelden we vast dat de beoordeling III op basis van het totale databestand niet terecht was. Vandaar de vorig jaar nieuwe beoordeling I. Begrijpend Lezen 345678 heeft de beoordeling I voor LWOO gekregen. Deze I kan niet voor de toets Begrijpend Lezen, toetsversie 678 ten aanzien van het PrO gelden, aangezien de ondergrens groep 6 heeft en dus correspondeert met een DLE van 30. Ook hier is om dezelfde reden als voor technisch lezen, het bereik beperkt tot uitsluitend de basisschool.

➤ *TPVO 2003 en TPVO Internetversie 2004*

Twee jaar geleden hadden we voor beide tests een 0 als beoordeling. Ondertussen kwamen de tests voldoende in het bestand voor om een oordeel uit te kunnen spreken. Beide toetsen leveren gemiddeld als resultaat een veel hogere DLE op voor potentiële LWOO leerlingen dan de andere toetsen (rond de 40 tegenover 32 voor de andere toetsen, vastgesteld voor plm. 1500 leerlingen). Vandaar dat we overgegaan zijn naar de beoordeling III voor LWOO leerlingen. De 2003 versie is niet geschikt voor potentiële PrO leerlingen (de toets gaat maar tot DLE 31). De internetversie heeft deze beperking niet, maar daar hebben we te weinig gevallen (24, maar ook met een relatief hoog DLE gemiddelde) om een oordeel uit te spreken.

Niveautoets & NiveautoetsPlus, CITO 2007

Hiervoor geldt hetzelfde als bij Technisch Lezen is gemeld, Cotan goedgekeurd, beoordeling I. Mogelijk vertragingproblemen voor de tijdsplanning in de aanvraagprocedure.

3. Ten aanzien van spelling:

➤ *Niveautoets & NiveautoetsPlus, 2007, Cito 2007*

Hier geldt hetzelfde als in het voorafgaande voor toetsen voor Technisch en Begrijpend lezen ten aanzien van deze toets is gesteld. Beoordeling 1.

➤ *Aarnoutse Spellingtest voor de basisschool*

Vorig jaar al had de spellingstoets van Aarnoutse een III als beoordeling. Dit op basis van de kritische geluiden uit het veld. Deze draaiden voor een groot deel om het zeer beperkte DLE-bereik. Dit jaar konden we ook nog vaststellen, dat de scores qua resultaat ver onder het gemiddelde van de overige tests uitkwamen. Dit gold met name ook voor de PrO groep. **Ook hier geldt dat de toets dit jaar voor het laatst op de lijst staat.**

➤ *TPVO 2003 en TPVO Internetversie 2004*

Hetzelfde als gesteld voor de toetsversies van begrijpend lezen van de TPVO (Internetversie), geldt ook voor de spellingstests. Zo mogelijk zijn hier de verschillen nog groter met de overige tests (10 of meer DLE punten hoger). Ook hier is de TPVO 2003 niet geschikt voor potentiële PrO-leerlingen (gaat tot DLE 31).

➤ *De toelatingstoets Cito 2001*

De toelatingstoets van het Cito is in de beoordeling het vorig jaar al ten aanzien van potentiële LWOO leerlingen teruggedaan van I naar III. Dit ligt aan het feit dat deze toets qua DLE wel een half jaar lager uitkomt dan de concurrerende tests. Bovendien wordt de toets heel weinig gebruikt.

Drempelonderzoek

Ook hier is het toetsbereik beperkt tot de groepen 5 t/m 8 van de basisschool.

4. Ten aanzien van Inzichtelijk Rekenen:

➤ *Drempelonderzoek.*

Bij het Drempelonderzoek stonden vorig jaar twee versies vermeld: de derde herijkte versie van 2001 en de vierde versie van 2006. We hebben vastgesteld dat de versie van 2001 niet meer gebruikt wordt. Dit zal aan de centrale regie van de uitgever liggen. We nemen daarom alleen de recentste versie op en wel vanwege het veelvuldig gebruik met de beoordeling I. In tegenstelling tot vorig jaar is het gebruik voor het eerste leerjaar van het VO vervallen. De normen waren hiervoor door de COTAN met een "onvoldoende" gekenmerkt.

➤ *TPVO 2003 en TPVO Internetversie 2004*

In tegenstelling tot de TPVO tests voor Begrijpend Lezen en Spelling bewegen de scores zich hier wel rond het gemiddelde, zij het dat de 2003-versie nog altijd aan de hoge kant zit. Ook hier geldt dat deze versie niet geschikt is voor potentiële PrO leerlingen vanwege de beperking tot DLE 31.

➤ *Niveautoets & NiveautoetsPlus, 2007*

Hier geldt weer wat ook in de andere subparagrafen gesteld is voor deze toets. Cotan goedkeuring, beoordeling I en we verwachten problemen met betrekking tot de tijdsplanning van scholen voor het doen van een aanvraag. Het zal moeilijk zijn om tot een indicatiebeslissing voor de zomervakantie te komen.

Juni 2008

Certificeringscommissie RVC-VO

Gebruikte literatuur

Greef, E. de, en T. van der Geest (2005), *Onderzoek naar het gebruik van klassikale intelligentietests, inzonderheid de NDT*, www.rvc-vo.nl

Melis, G. (2005), *Exploratief onderzoek begrijpend lezen en NDT*, www.rvc-vo.nl