

Taaltoetsen en schooltaalbeleid

Is meten echt beter weten?

Kris Van den Branden

In 2006 voerde het Centrum voor Taal en Onderwijs (K.U.Leuven) in opdracht van de Vlaamse minister van Onderwijs een studie uit naar de haalbaarheid en wenselijkheid van centrale taaltoetsen in Vlaanderen. Met centrale taaltoetsen bedoelen we toetsen die door een gespecialiseerde instantie worden ontwikkeld en genormeerd, en die breed beschikbaar zijn voor het ganse Vlaamse onderwijsveld. Eén van de vragen die de studie diende te beantwoorden was of centrale, door de overheid geïntroduceerde taaltoetsen een positieve invloed kunnen hebben op het taalbeleid dat door scholen wordt gevoerd. Het Centrum voor Taal en Onderwijs verzamelde rond deze vraag de nationale en internationale onderzoeksliteratuur. De belangrijkste resultaten van deze studie worden in twee *VONK*-artikelen samengevat. In dit eerste artikel gaan we in op het nut van centrale taaltoetsen voor het in kaart brengen en opvolgen van leerlingresultaten. In het tweede artikel (dat in het volgende *VONK*-nummer verschijnt) gaan we dieper in op het inzetten van taaltoetsen in functie van een structureel schooltaalbeleid¹.

INLEIDING

SCREENING VERSUS DIAGNOSE VAN LEERLINGPRESTATIES

Inleidend moeten we even stilstaan bij het verschil tussen screening en diagnose. Bij *screening* (of 'signalering') wil men een prestatie van een leerling in kaart brengen en vanuit een vergelijking met een criterium of met de prestaties van andere leerlingen of de onderlinge vergelijking van de leerling met zichzelf, een uitspraak doen over het kwaliteitsniveau van de geleverde prestatie. Dat kwaliteitsniveau kan men op diverse wijze uitdrukken, gebruikmakend

van schalen, percentielen, absolute cijfers, codes of zelfs verbale commentaar. De basisfunctie van screening is een uitspraak te doen over de mate waarin de prestatie als voldoende of onvoldoende kan worden beschouwd, waarbij deze twee termen uiteraard met verdere verfijningen genuanceerd kunnen worden. Men kan screenen op leerling-, groeps-, klas- of schoolniveau.

Diagnose gaat verder dan screening. Diagnose heeft tot doel om achter de prestatie te kijken. Men wil ontleden (a) welke processen aanleiding hebben gegeven tot de

De basisfunctie van screening is een uitspraak te doen over de mate waarin de prestatie als voldoende of onvoldoende kan worden beschouwd.

geleverde prestatie, en/of (b) welke aspecten van de prestatie wel succesvol waren, en welke andere aspecten niet. Bij diagnose heeft men de ambitie om een gedetailleerd zicht te krijgen op de wijze waarop de leerling de toetstaak heeft aangepakt en op

Diagnose heeft tot doel om achter de prestatie te kijken.

de subniveaus van taakuitvoering die hij wel en niet beheerst, om op die manier veel meer informatie te verkrijgen over de wijze waarop de doelcompetentie bij de leerling kan worden bevorderd.

Daarbij is het vaak van belang om het geheel van kennis, vaardigheden en attitudes die bij het leveren van de doelprestatie om de hoek komen kijken, te ontleden, en ze in de prestatie van de leerling na te gaan, teneinde een scherper zicht te krijgen op hoe de verdere ondersteuning van de leerling het best vorm kan krijgen. Hieruit volgt dat wie lessen wil trekken over het beter ondersteunen van leerlingen via screening slechts vrij algemene informatie verkrijgt (voornamelijk over *wie* voor diepgaandere ondersteuning in aanmerking komt), terwijl diagnose kan leiden tot gerichtere info over *welke vorm* de ondersteuning het best aanneemt en op welke vaardigheden, kennis en attitudes ze zich primair hoort te richten om enig leereffect tot stand te brengen.

De complexiteit en uitgebreidheid van de diagnose nemen toe naarmate de competentie die in kaart wordt gebracht, complexer wordt. Dit wordt treffend geïllustreerd met het verschil tussen technisch lezen en begrijpend lezen. Van *technisch lezen* (het vlot en automatisch decoderen van schriftbeelden) wordt algemeen aangenomen dat het een vaardigheid is die een overzichtelijk en beschrijfbaar geheel van deelprocessen en deelvaardigheden omvat. In de meest

gebruikte methodes voor technisch lezen worden deze afzonderlijke deelprocessen dan ook systematisch gedemonstreerd en uitgelegd aan de leerlingen en vervolgens ingeoefend, teneinde tot een verregaande automatisering van het decoderingsproces te komen. De relatieve eenvoud van deze 'mechanistische' vaardigheid leidt er ook toe dat het mogelijk is een leerlingvolgsysteem voor technisch lezen te ontwikkelen waarbij prestaties van leerlingen niet alleen worden gescreend, maar ook verregaand gediagnosticeerd, met het oog op het gericht ondersteunen op deelvaardigheden die te traag worden ontwikkeld.

In de in Vlaanderen meest gebruikte leerlingvolgsystemen voor technisch lezen gebeurt een eerste screening (signalering) van technisch lezen met zogenaamde 'minutentoetsen' die op regelmatige tijdstippen worden afgenomen, voornamelijk in het eerste tot en met het derde leerjaar. Bij deze screenings-toetsen laat men leerlingen tegen de tijd woordbeelden hardop verklanken. De tijdsfactor en correctheid geven een indicatie van de mate waarin de leerling vlot en automatisch (tijdsdimensie) en accuraat (correctheid) technisch leest. Leerlingen die vanuit deze eerste screening een prestatie leveren die onder het gewenste niveau ligt of die plots een bruske daling in normcategorie vertonen, worden aan een diagnose onderworpen (de zogenaamde 'analyse') die tot handelen moet kunnen leiden (vandaar 'handelingsgerichte diagnostiek'). Bij deze analyse moeten de leerlingen bijkomende technisch-leesprestaties leveren (bijvoorbeeld het hardop voorlezen van contrasterende paren) om te ontdekken welke aspecten van het decodeerproces zij al wel hebben verworven en welke niet, en welke factoren remmend optreden bij het vlot verwerven van de technisch-leesvaardigheid. Vervolgens kunnen dan aangepaste oefeningen aan de leerling worden aangeboden.

Daar waar het aantal mogelijke bepalende deelprocessen bij technisch lezen nog vrij overzichtelijk is, ligt dat bij *begrijpend lezen* helemaal anders. De cognitieve operaties, (voor)kennis en de basisattitudes die nodig zijn om een begrijpend-leestaak tot een goed einde te brengen, zijn bijzonder veelomvattend, zelfs in die mate zo dat tot op heden de taalwetenschap de complexiteit van begrijpend-leesprocessen nog niet volledig in kaart heeft weten te brengen. Begrijpend-leesprocessen vereisen een complexe interactie tussen tekst en lezer: problemen bij het opbouwen van leesbegrip kunnen op ontelbaar verschillende wijzen ontstaan. Bij diagnose van begrijpend-leesproblemen zal met dat uitgangspunt rekening moeten worden gehouden, wat de diagnose in vergelijking met technisch lezen, niet minder noodzakelijk, maar wel veel complexer maakt (cf. infra).

DIRECT VERSUS INDIRECT SCREENEN

Een tweede onderscheid dat we ter inleiding moeten maken is dat tussen direct en indirect screenen. Van *directe screening* van een vaardigheid spreekt men als de prestatie die iemand moet leveren een directe doorslag is van de eigenlijke vaardigheid die men in kaart wil brengen. Eenvoudig gezegd: om te meten of iemand een krantenartikel kan begrijpend lezen, laat men hem een krantenartikel lezen en daarbij een opdracht vervullen waarmee de kandidaat kan tonen dat hij het artikel heeft

begrepen. Bij *indirecte screening* laat men de kandidaat een andersoortige prestatie leveren, waarvan men veronderstelt dat ze een indicatie geeft van de mate waarin iemand een bepaalde vaardigheid beheerst. Zo kan men iemand een schriftelijke woordenschattoets laten oplossen om op basis daarvan een uitspraak te doen over de spreekvaardigheid van de kandidaat; in het geval van indirecte screening moet de toetsafnemer echter wel de link kunnen aantonen tussen de prestatie en de vaardigheid waarover men uitspraken wil doen.

ANALYSE VAN BESTAANDE TAALTOETSEN

In het kader van deze studie maakte het Centrum voor Taal en Onderwijs een inventaris van de bestaande centrale taaltoetsen in Vlaanderen. In deze inventaris werden gestandaardiseerde en genormeerde toetsen en toetsbatterijen opgenomen die in Vlaanderen beschikbaar zijn voor het hele onderwijsveld, en die betrekking hebben op de toetsing van taalkennis en taalvaardigheid. Elk van de geïnventariseerde toetsen werd beschreven op een handige fiche die allerhande basisinformatie bevat. Ter illustratie hebben we één fiche opgenomen (figuur 1); de volledige inventaris is te downloaden op de website van het Centrum voor Taal en Onderwijs (www.cteno.be).

FIGUUR 1: Voorbeeldfiche van een gestandaardiseerde, genormeerde taaltoets

TALK. Taalvaardigheidstoets Aanvang Laatste Kleuterklas	
AUTEUR	Werkgroep Taaltoetsen
INSTELLING	Steunpunt NT2
PUBLICATIEJAAR	1996
UPDATE VAN DE NORMEN	Geen
ONDERWIJSNIVEAU	Kleuteronderwijs
DOELGROEP	Derde kleuterklas
GETOETSTE PRESTATIE	Luistervaardigheid
BESCHRIJVING	Eenmalige afname van 30 korte items, begin schooljaar geënt op schooltaalonderzoek
AFNAMEMODALITEITEN	Leerkracht Klassikaal 50'
MAATSTAF	Norm (andere leerlingen)
DIAGNOSE/SCREENING (MEETPRETENTIE)	Screening op leerlingniveau Mesoniveau
VOLGSYSTEEM	Neen
VLAAMS/NEDERLANDS	Vlaams

TAALTOETSEN VOOR BASISONDERWIJS

Onze analyse van de bestaande toetsen en leerlingvolgsystemen in Nederland en Vlaanderen leerde dat heel wat van de op dit moment sterk verspreide instrumenten voor een indirecte meting van taalvaardigheid kiezen (vooral dan bij de jonge kleuters), of voor de directe meting van technische deelaspecten van taalvaardigheid (technisch lezen en spelling bij uitstek).

Op *kleuterniveau* is er een behoorlijke lacune voor wat de directe meting van functionele schooltaalvaardigheid betreft. Vooral de functionele spreekvaardigheid van de kleuters wordt slechts mondjesmaat in kaart gebracht. De bestaande kleutertoetsen focussen vooral op woordenschatkennis en metalinguïstische kennis (leesvoorwaarden). Voor het einde van het kleuteronderwijs (of begin eerste leerjaar) is er momenteel geen screeningsinstrument voorhanden dat op directe wijze de functionele taaldoelen van

de Vlaamse ontwikkelingsdoelen en Nederlandse tussendoelen in kaart brengt.

Op het niveau van het *lager onderwijs* vinden we indirecte toetsing bijvoorbeeld terug in een aantal schrijftoetsen waarbij de leerling wordt gevraagd om andermans schrijfproducten te beoordelen of te verbeteren, eerder dan zelf te moeten schrijven. Een ander markant voorbeeld is het gebruik dat sommige Vlaamse scholen maken van hun leerlingvolgsysteem voor technisch lezen en spelling. Sommige scholen die niet beschikken over directe toetsen voor functionele taalvaardigheid, hanteren hun technisch-leestoetsen als indicator voor de begrip-leesvaardigheid van hun kinderen. De relatie tussen technisch leesvaardigheid en begrip-lezen is echter zeer complex: het is niet zo dat iedere leerling die sterk technisch kan lezen zich ook ontpopt tot een goede begrip-lezer. Hetzelfde geldt voor de relatie tussen spelling(toetsen) en schrijfvaardigheid.

Voor het lager onderwijs zijn er nochtans wel genormeerde en gestandaardiseerde toetsen voor de directe meting van functionele taalvaardigheid, maar het aanbod is niet volledig dekkend. De Cito-toetsen zijn uit Nederland afkomstig, wat soms tot een verschil in taalgebruik en terminologie leidt (van sommige toetsen werd ondertussen wel een Vlaamse versie geproduceerd, maar ook hier ligt de focus toch op technisch lezen en spelling). Het Vlaamse VLOT-leerlingvolgsysteem differentieert in haar normcategorieën vooral goed naar de lagere scores (omdat het zich vooral op scholen uit het Onderwijsvoorangsbeleid (OVb) richtte) en minder sterk naar de hogere scores. De Vlaamse periodieke peilinginstrumenten (voor begrip-lezen, luisteren en schrijven) vertonen een hoge betrouwbaarheid en goede indicaties qua

inhoudsvaliditeit en link met de eindtermen, maar beperken zich tot het einde van het lager onderwijs, en zijn in de eerste plaats bedoeld om de overheid informatie te verschaffen op systeemniveau, eerder dan op individueel leerlingniveau of schoolniveau.

TAALTOETSEN VOOR SECUNDAIR ONDERWIJS

Op het niveau van het secundair onderwijs focussen de bestaande toetsen vooral op de eerste graad van het secundair onderwijs. Veel gebruikte toetsen zoals TAS, TIST en Diataal proberen op een vrij directe wijze de leesvaardigheid van leerlingen in kaart te brengen, en zijn er vooral op gericht om scholen toe te laten leerlingen met taalproblemen te detecteren. Sommige van de bovenstaande toetsen maken gebruik van edittoetsen, die erin bestaan dat leerlingen overtollige woorden uit een tekst schrappen; dit zijn dus indirecte toetsen van algemene schooltaalvaardigheid. Wat ook opvalt, is dat luister- en spreekvaardigheid veel minder in de toetsen aanwezig zijn dan de schriftelijke vaardigheden. Gestandaardiseerde, genormeerde taal(vaardigheids)-toetsen voor de tweede en derde graad secundair onderwijs zijn nagenoeg onbestaande. Op dit vlak bestaat dus nog een stevige lacune.

WANNEER BEGINNEN MET SCREENEN?

Uit het bovenstaande overzicht blijkt dat de bestaande screeningsinstrumenten op zeer verschillende momenten in de onderwijsloopbaan van de leerlingen worden afgenomen. Veel toetsen hebben expliciet de bedoeling om een tijdige screening mogelijk te maken, namelijk op het moment dat er

nog tijd is voor diagnose en een langdurige leerlinggerichte ondersteuning. In dit verband tonen meta-analyses van schooleffectiviteitsonderzoek dat voor schoolsucces in het algemeen, en voor schoolse taalvaardigheid in het bijzonder, het opvolgen van de leerlingen (bijvoorbeeld via het hanteren van een leerlingvolgsysteem) een kenmerk is van krachtig onderwijs. Screening op leerlingniveau heeft vaak de functie van het alarmsignaal dat gaat knippen en dat schoolteams alert kan maken op het feit dat sommige leerlingen extra zorg behoeven.

Screening op leerlingniveau heeft vaak de functie van het alarmsignaal dat gaat knippen.

TOETSEN IN DE KLEUTERSCHOOL?

In veel landen (o.a. Nederland, Duitsland, Verenigde Staten) begint men zeer vroeg met de eerste screenings. Reeds op 4-jarige leeftijd (en soms zelfs vroeger) worden leerlingen aan taaltoetsen onderworpen met de bedoeling om vroegtijdig taalachterstanden op te sporen of om aan bepaalde leerlingen gerichte ondersteuning te kunnen geven. Zo kreeg ook in Vlaanderen een mogelijke beginmeting voor 4-jarigen aandacht in de pers naar aanleiding van de actuele praktijk in Noord-Rijnland-Westfalen (Duitsland). Na moeizame debatten keurde het parlement van die Duitse deelstaat een resolutie goed om bij alle 4-jarige kleuters vast te stellen in welke mate ze Duits spreken. De idee hierachter is dat vroegtijdige detectie van taalachterstand kan leiden tot betere ondersteuning. In Engeland, Schotland, Australië en Nieuw-Zeeland wordt momenteel van 4-jarige kleuters de PIPS-toets afgenomen. In Nederland is een vertaling van deze toets gemaakt (de OBIS-toets). Nog in Nederland bieden o.a. de Toets Tweetaligheid en de

TAK-onderbouw, beide van het Cito, de kans om op jonge leeftijd een indicatie te krijgen van de taalvaardigheid van jonge kleuters. In Vlaanderen is het Kleuterobservatie-instrument Taalvaardigheid (KOBI-TV), bedoeld voor afname in de tweede kleuterklas, sterk verspreid.

Al deze instrumenten roepen in de wetenschappelijke onderzoeksliteratuur de vraag op vanaf welke leeftijd kleuters überhaupt toetsbaar zijn. Uit een screening van bestaande kleutertoetsen voor de eerste kleuterklas in het doctoraatsonderzoek van Verhelst (2002) bleek dat kinderen voor de leeftijd van 4 jaar nauwelijks te toetsen zijn. Ook andere onderzoekers komen tot deze conclusie. Jonge kleuters kunnen zich nauwelijks langer dan 10 minuten concentreren en moeten al erg vertrouwd zijn met de testafnemer vooraleer ze aanspreekbaar zijn. Als het om anderstalige kleuters gaat, is er bovendien de vraag naar de begripelijkheid van de instructies. Bij kinderen die in het Nederlands worden getoetst terwijl ze die taal nog niet goed spreken, is niet duidelijk waardoor een fout antwoord veroorzaakt wordt: door het niet begrijpen van de vraag of door het niet weten van het goede antwoord. Volgens ontwikkelingspsychologen vertoont kleuterontwikkeling bovendien zulke grillige patronen en zoveel fluctuaties in prestaties (*"kleuters verschillen meer van zichzelf dan zij van anderen verschillen"*) dat toetsen van jonge kleuters een heel precaire aangelegenheid wordt. Op basis van haar studie van kindvolgsystemen, concludeerde Bijlsma (2004):

"Jonge kinderen presteren in een testsituatie niet altijd optimaal. Zij zijn niet altijd gemotiveerd om mee te doen, voelen zich niet altijd veilig bij onbekenden, weten niet altijd wat van hun in de testsituatie wordt verwacht. Dit alles kan de prestatie negatief

beïnvloeden. Het beeld van de ontwikkeling van het kind is het meest betrouwbaar vast te stellen in een voor kinderen veilige en vertrouwde situatie, waarin het kind actief en spontaan betrokken is bij het op het kind ingestelde activiteiten.” (Bijlsma, 2004, p. 25)

In opdracht van de Nederlandse onderwijsminister Netelenbos onderzocht ontwikkelingspsycholoog Kohnstamm (Commissie Indicatiestelling Onderwijsachterstanden, 1996) de zin van een centrale test voor 4-jarigen. Kohnstamm bestudeerde bestaande testen in Nederland, Engeland en Amerika, en kwam tot de conclusie dat ze zeer weinig voorspellende waarde hebben, in puur psychometrische zin wel ‘in zekere zin betrouwbaar’ kunnen zijn, maar dat de validiteit (en voorspellende kracht) van de testen veel problematischer is. Verhoeven & Vermeer (2005) en de Commissie Indicatiestelling Onderwijsachterstanden concludeerden dat de kleutertoetsen slechts indicaties van taalachterstand bevatten, die überhaupt niet waterdicht zijn, noch naar de beslissing over welke kleuters precies extra zorg behoeven, noch naar wat voor soort ondersteuning de kleuters dan precies nodig hebben. De Nederlandse staatssecretaris Netelenbos zag onder andere op basis van Kohnstamms advies van de toets af.

Een blik op de Cito-toetsen voor kleuters, en de bovenvermelde internationaal opgang makende PIPS-toets, toont aan dat met dit soort kritische overwegingen wel degelijk wordt rekening gehouden door de toetsontwikkelaars. Zo presenteren deze toetsen zich in de eerste plaats als (genormeerde) indicatoren van taalachterstand die moeten worden aangevuld met de observaties die de leerkracht tijdens de dagdagelijkse bezigheden in de kleuterklas vaststelt. Bovendien hoeden de handleidingen van deze tests zich om uitspraken te doen over

met welke specifieke didactische ondersteuning op lage prestaties moet worden gereageerd. PIPS vermijdt dat zelfs volledig, terwijl de Cito-handleidingen verwijzen naar algemene principes van een goede taalvaardigheidsaanpak (zoals veel taalaanbod, veel vertellen en voorlezen, aandacht voor woordenschat, etc.). Het meest voor de hand liggende gebruik van de toetsen is volgens de handleiding dat van een herhaalde screening. De gebruikers wordt aangeraden om bij de leerlingen de toets af te nemen, vervolgens een krachtige taalvaardigheidsaanpak te ontwikkelen en daarbij gedifferentieerde ondersteuning te voorzien aan de leerlingen die laag scoorden, en ten slotte (bijvoorbeeld een jaar na de eerste afname) de toets opnieuw af te nemen om te kijken of het kind in ontwikkeling is.

Deze inzet van de kleutertoets is grotendeels conform de filosofie van het Vlaamse Kleuterobservatie-instrument Taalvaardigheid (KOB-TV), dat bij 4-jarigen (tweede kleuterklas) wordt afgenomen. Ook met de afname van deze toets kan de leerkracht de scores van de leerling vergelijken met Vlaamse normen en wordt de school vervolgens aangeraden om een krachtige taalvaardigheidsaanpak te ontwikkelen in heterogene groepen (waarbij wordt vermeden dat de taalaanpak naar de zwakpresterende kleuters wordt gereduceerd tot het geïsoleerd aanleren van woordenschat en grammatica in aparte klasjes), en vervolgens op een later moment het instrument opnieuw af te nemen om de groei van taalvaardigheid systematisch op te volgen. Ook hier wordt duidelijk aangegeven dat een dergelijke kleutertoets, gezien de leeftijd van de kinderen, niet geschikt is voor een

***“Kleuters
verschillen meer
van zichzelf dan zij
van anderen
verschillen.”***

volledige diagnose van de taalkenmerken van het kind: daarvoor is gerichte en brede observatie van de kleuter doorheen de klasdag(en) nodig.

TOETSEN VANAF DE LAGERE SCHOOL

Op basis van een studie in 2002 komt de Nederlandse Onderwijsraad tot de conclusie dat pas vanaf de leeftijd van circa 6 jaar centrale toetsing van kleuters veel meer toepassingsmogelijkheden heeft:

"Praktisch biedt een meetmoment omstreeks het zesde jaar ook voordelen:

- toetsing vergt minder tijd;*
- toetsen zijn klassikaal afneembaar; en*
- de meting is betrouwbaarder dan op vierjarige leeftijd."* (Onderwijsraad, 2003, p. 35-36)

Vanaf de leeftijd van 5 à 6 jaar (in Vlaanderen derde kleuterklas – eerste leerjaar) is

Tragere groei kan een krachtig signaal vormen dat een nadere diagnose van de oorzaken ervan nodig is om gepaste ondersteuning te bieden.

op het vlak van toetsing dus veel meer mogelijk dan bij nog jongere kinderen, en kunnen leerlingen op het vlak van taalvaardigheid makkelijker met regelmatige tussenpozen doorheen het onderwijs gescreend worden. Dat heeft dan het voordeel dat leerlingen niet alleen systematisch kunnen worden vergeleken met andere leerlingen of criteria, maar ook met

zichzelf, dat met andere woorden hun groei in kaart gebracht kan worden.

Een aantal bestaande leerlingvolgsystemen (LVS) spelen sterk in op dit 'longitudinale' aspect van evaluatie. LVS-pakketten die met normcategorieën werken over leerjaren

heen, raden de leerkrachten in de handleiding aan om de beweging van een leerling over de normcategorieën scherp in de gaten te houden. Als een leerling met regelmatige tussenpozen wordt gescreend (bijvoorbeeld eenmaal per trimester, of eenmaal per schooljaar) en de leerling zich steeds in dezelfde normcategorie handhaaft, maakt hij een 'normale' of 'gemiddelde' groei door; als de leerling in normcategorie afdaalt, is zijn groei trager dan die van de normgroep, en als de leerling in normcategorie stijgt, is er sprake van leerwinst (sterkere groei dan de 'gemiddelde' leerling). Tragere groei kan een krachtig signaal vormen dat een nadere diagnose van de oorzaken ervan nodig is om gepaste ondersteuning te bieden.

Een nadeel van herhaalde screening is echter dat veel tijd naar toetsing gaat. Er moet dus worden gezocht naar toetsen die op een minimale tijdsspanne (ook voor de leerling), en met een relatief economische inzet van menskracht (schoolpersoneel) kunnen worden afgenomen, maar die desondanks maximale informatie opleveren. In dit opzicht zijn klassikaal af te nemen toetsen te verkiezen boven individueel af te nemen toetsen, als zij tenminste dezelfde kwaliteit aan informatie opleveren. Voor de meeste op de markt zijnde LVS-systemen geldt dat zij klassikaal, en binnen een relatief beperkte tijdsspanne zijn af te nemen; uitzonderingen vormen enerzijds de toetsen voor technisch lezen (die individueel worden afgenomen) en de toetsen voor spreekvaardigheid, die niet alleen afzonderlijk moeten worden afgenomen, maar bovendien relatief veel inzet van schoolpersoneel vragen omdat zij pas nadien kunnen worden gescoord op basis van uitgebreide scoringswijzers en audio-opnames.

HET BEOORDELEN VAN LEERLINGPRESTATIES

Prestaties die verwijzen naar een bepaald construct, moeten beoordeeld worden. In sommige van de bovenstaande gevallen is de beoordelingsmethode vrij eenduidig en waterdicht, zoals het tellen van het aantal woorden dat een kind hardop heeft voor-gelezen, of het tellen van het aantal juist aangekruiste vragen (met multiple choice in een luisteropdracht). In andere gevallen is de beoordeling complexer, en open voor meer interpretatie. Dat is het meest uitgesproken het geval voor spreek- en schrijftoetsen (vooral schrijftoetsen waarbij de leerling wordt gevraagd om een functionele schrijftaak uit te voeren). Hiervoor moeten door de ontwikkelaars uitgebreide beoordelingswijzers worden opgesteld, die vervolgens door getrainde beoordelaars worden toegepast. Uit wetenschappelijk onderzoek blijkt dat zonder uitgebreide beoordelingswijzers de scoring van spreek- en schrijftaken aan zeer veel willekeur en subjectiviteit onderhevig is.

HET BELANG VAN EEN MAATSTAF

De 'score' van de leerling moet vervolgens aan een maatstaf getoetst worden, om te bepalen of er van een probleem sprake is. In vele bovenvermelde toetsen wordt *normgerelateerd* gewerkt. Dat houdt in dat de prestatie van een individuele leerling wordt vergeleken met de scores van een uitgebreide steekproef, waarvan de scores in percentielen of normcategorieën zijn ondergebracht. Leerlingen worden hierbij dus met andere leerlingen vergeleken. Verschillende technieken worden gebruikt om de normcategorieën te definiëren: zo maakt de ene toets gebruik van het gemiddelde van de steekproefscores en de standaarddeviatie, terwijl andere toetsen gebruik maken van decielen.

In een aantal gevallen wordt de prestatie van de leerling (ook) vergeleken met een *criterium*. Bijvoorbeeld in het geval van de Vlaamse periodieke *peellingsinstrumenten* wordt een cesuur bepaald door een commissie van experts, die een uitspraak doet over het niveau dat leerlingen zouden moeten halen om de eindtermen voor de desbetreffende vaardigheid te bereiken. Ook uit predictief validiteitsonderzoek kan een criterium worden afgeleid: de analyses van zulk onderzoek tonen aan dat indien de prestaties van een leerling onder een bepaalde drempelscore vallen, de kans op 'problemen' aanzienlijk worden. Een criterium voor prestaties op de ene toets kan met andere woorden bepaald worden op basis van prestaties op een andere toets, die bijvoorbeeld later in de schoolloopbaan wordt afgenomen. Zo kunnen scores op een kleutertoets worden gelinkt via wetenschappelijk onderzoek (bij voorkeur longitudinaal onderzoek waarbij leerlingen gevolgd worden) met toetsen in het lager onderwijs die genormeerd zijn of van een cesuur zijn voorzien, en kan, mits de correlaties voldoende hoog zijn, vervolgens een criteriumscore voor de kleutertoets worden bepaald.

Het bovenstaande geeft aan dat er in essentie twee manieren bestaan om van een 'probleem' (in dit geval: 'taalachterstand') te spreken: *normgerelateerd* of *criteriumgerelateerd*. In het eerste geval vindt men een prestatie problematisch omdat ze te sterk achterop hinkt bij de gemiddelde prestatie van de leerlingengroep. Aan deze probleembepaling zal men

*Zonder uitgebreide
beoordelingswijzers
is de scoring van
spreek- en
schrijftaken aan
zeer veel willekeur
en subjectiviteit
onderhevig.*

de voorkeur geven indien men de verschillen tussen de prestaties van individuele leerlingen binnen bepaalde marges wil houden. Dit is niet de eerste bezorgdheid bij een criteriumgerelateerde bepaling van taalachterstand. In dat laatste geval gaat men na of leerlingen een minimumdrempel hebben behaald, en focust men minder op het feit dat een aantal leerlingen veel hogere beheersingsniveaus etaleren. In het Vlaamse onderwijslandschap wijst een benadering vanuit de eindtermen in de richting van een criteriumgerelateerde benadering. Het is dan ook niet toevallig dat bij de ontwikkeling van de periodieke peilingsinstrumenten voor begrijpend lezen, schrijven en luisteren telkens met een cesuurbepaling werd gewerkt, waarbij een commissie van experts een niveau aanbracht waarop gezegd kan worden dat de leerling de eindtermen haalt.

Deze cesuur kan dus geïnterpreteerd worden als de drempel waarover leerlingen moeten geraken qua beheersingsniveau.

In het Vlaamse onderwijs-landschap wijst een benadering vanuit de eindtermen in de richting van een criteriumgerelateerde benadering.

NORM OF DREMPEL?

Met betrekking tot de inzet van taaltoetsen om de kwaliteit van ondersteuning aan individuele leerlingen te verhogen, hebben zowel een normgerelateerde bepaling als

een drempelgerelateerde bepaling hun voor- en nadelen. Normgerelateerde probleembepaling vereist ten eerste dat met zorg wordt overdacht met welke andere leerlingen individuele leerlingen worden vergeleken. Indien de normgroep niet representatief is voor de populatie zou dit tot een scheeftrekking van de probleembepaling (of maatstaf) kunnen leiden. Ten tweede moet men in contexten waarin men pleit voor een

geïntegreerde en structurele aanpak van onderwijs- en taalachterstanden, er rekening mee houden dat zulk een aanpak niet alleen de zwakkere leerlingen ten goede komt, maar ook de betere leerlingen, met als gevolg dat hoeveel moeite men ook doet, en hoeveel winst men op het vlak van onderwijskwaliteit ook boekt, het verschil tussen taalzwakkere en taalsterkere leerlingen niet noodzakelijk kleiner wordt (en het 'probleem' dus niet opgelost geraakt).

Een criteriumgerelateerde bepaling is absoluut en dus vaster (over de tijd heen), maar kan het nadeel hebben dat leerlingen die net met de hakken over de sloot geraken, geen bijkomende ondersteuning krijgen. Drempels zijn ook niet absoluut. Dat heeft onder andere te maken met de relatie tussen de van de leerling gevraagde prestatie en het construct dat in kaart wordt gebracht. Met name voor zogenaamde directe toetsen stelt zich de vraag hoe 'direct' de meting dan wel is. Is bijvoorbeeld een drempeltaak waarbij de leerlingen wordt gevraagd om een uitnodiging te schrijven voor leeftijdsgenoten indicatief voor de globale schrijfvaardigheid van de leerling, of enkel voor het schrijven van uitnodigingen?

Wetenschappelijk onderzoek laat hier geen sluitende conclusies toe. Zo durven de correlaties tussen toetsen van dezelfde vaardigheid (gesitueerd binnen eenzelfde niveau) sterk van elkaar verschillen. Dit probleem is in de toetsliteratuur bekend als het probleem van *extrapolatie*. Dit probleem is nauw verwant met de mate van *eendimensionaliteit* van (taal)vaardigheden. Eendimensionaliteit leidt ertoe dat onder verschillende prestaties (bijvoorbeeld het uitvoeren van luistertaken van verschillende teksttypes) dezelfde eendimensionele vaardigheid schuilt. In toetsconstructie die vanuit een IRT-paradigma wordt opgezet, uit zich deze

eendimensionaliteit in het construeren van één schaal waarop de items van verschillende toetstaken samen kunnen worden geplaatst. Bij de ontwikkeling van de periodieke peilingsinstrumenten voor begrijpend lezen en luisteren in Vlaanderen bleken de items wel degelijk op één schaal te passen, terwijl dat voor functioneel schrijven niet het geval bleek te zijn.

"Dit onderzoek heeft aangetoond dat het niet mogelijk is om een enkele vaardigheidsschaal voor schrijven te ontwikkelen, maar wel een voor elk van de zes verschillende aspecten (van schrijfvaardigheid). De overheid moet zich realiseren dat over toekomstig peilingsonderzoek schrijven op een andere manier gerapporteerd zal dienen te worden dan over begrijpend lezen. In plaats van uitspraken te doen over de mate waarin leerlingen elk van de eindtermen schrijven halen, kan gedetailleerd aangegeven worden in welke mate de leerlingen elk van de deelvaardigheden voor schrijven realiseren, over alle eindtermen heen." (Colpin e.a., 2005)

Ook voor het Nederlandse PPOIN-instrument voor de periodieke peiling van schrijfvaardigheid bleek geen eengemaakte IRT-schaal uit de analyses te komen. Daar werden in het uiteindelijke instrument ook verschillende schalen onderscheiden, die verwijzen naar verschillende eisen die aan de uitvoering van een schrijfproduct worden gesteld, een appel doen op sterk van mekaar te onderscheiden mentale processen. Een gevolg van dit alles is dat men voor schrijfvaardigheid dus niet één drempel (cesuur) kon bepalen, maar in essentie verschillende cesuren moet onderscheiden. Dit schept uiteraard nieuwe problemen, zoals het bepalen van de relatie tussen, en

het relatieve gewicht van, de afzonderlijke cesuren. Dit alles sluit aan bij de aanwijzingen in de wetenschappelijke literatuur die stellen dat de productieve vaardigheden (spreken en schrijven) een complexer geheel van diverse mentale processen en deelvaardigheden vereisen dan receptieve vaardigheden (lezen en luisteren). Voor de productieve vaardigheden blijkt het dus minder eenvoudig om op basis van schrijf- of spreekprestaties van leerlingen uitspraken te doen over de schrijfvaardigheid en de spreekvaardigheid van leerlingen, eerder dan over bepaalde dimensies van hun spreek- en schrijfgedrag.

***Productieve
vaardigheden
vereisen een
complexer geheel
van diverse
mentale processen
en deelvaardig-
heden dan
receptieve
vaardigheden.***

HET PROBLEEM VAN DE PRODUCTIEVE VAARDIGHEDEN

Voor wie vanuit de eindtermen vertrekt, wordt het extrapolatieprobleem nog versterkt doordat een analyse van de eindtermen systematisch aantoonde dat voor sommige vaardigheden bepaalde eindtermen vrij fundamenteel van andere eindtermen verschillen, en dat bovendien sommige eindtermen moeilijk te operationaliseren zijn tot toetstaken voor gestandaardiseerde screening. Met name bij spreken en luisteren stelt zich het verschil tussen zogenaamd monologische (of eenrichtingsverkeer-) situaties en dialogische/polylogische situaties. Voor sommige eindtermen (zoals het luisteren naar een uiteenzetting of het luisteren naar een radiobericht) is de leerling enkel luisteraar; doch voor andere eindtermen (zoals het voeren van een telefoongesprek of het voeren van een discussie) is de leerling zowel luisteraar als spreker tegelijk.

De context van de taakuitvoering verschilt in beide gevallen fundamenteel, omdat in tweerichtingssituaties luisteraars een invloed op het gesprek kunnen hebben die in eenrichtingssituaties niet mogelijk is. Zo kan in echte dialogen de luisteraar om opheldering vragen indien hij iets niet begrijpt, de gang van het onderwerp sturen qua onderwerp, bepaalde hypothesen over de informatie die wordt overgebracht expliciet toetsen, etc. Toetsmethodologisch leveren tweerichtingssituaties veel problemen op. Al sluiten ze het sterkst aan bij natuurlijke, communicatieve situaties (en dus qua prestatie sterke, directe indicaties kunnen geven van de functionele communicatieve taalvaardigheid van de leerlingen), ze lenen zich vrij slecht tot situaties van gestandaardiseerde toetsing. De meest voorkomende problemen zijn:

- (1) de onderlinge afhankelijkheid van de prestaties van verschillende leerlingen;
- (2) de moeilijkheid om (zo gewenst) vaardigheden nog van elkaar te onderscheiden en dus apart te beoordelen;
- (3) de technologische noodzaak om opnames te maken van de gesprekken en deze pas nadien (aan de hand van een beoordelingsleutel) te kunnen scoren.

Het is dan ook geen toeval dat in de leerlingvolgsystemen die op de markt aanwezig zijn, zo weinig toetstaken voor spreek- en luistervaardigheid, zeker wat echte dialoogsituaties betreft, opgenomen zijn. Ook bij de ontwikkeling van het Vlaamse periodieke peilingsinstrument voor luistervaardigheid werden de echt dialogische situaties niet in de toetsontwikkeling opgenomen. Zelfs in leerlingvolgsystemen waar spreken en luisteren systematisch zijn uitgewerkt, wordt de voorkeur gegeven aan vrij kunstmatige situaties waarin de leer-

kracht de gesprekspartner vormt. Vanuit deze vaststelling wordt in de internationale toetsliteratuur dan ook sterk gepleit voor een evaluatie van spreek- en luistervaardigheid vanuit observatie in functionele klas- of buitenklassituaties, eerder dan vanuit strak gestandaardiseerde toetsing. Bij dit laatste dient de beoordelaar uiteraard ook gebruik te maken van gestandaardiseerde beoordelingswijzers of -schalen waarmee hij bepaalde aspecten van de spreek- en luisterprestatie kan afzetten tegen de prestatie van andere leerlingen of een bepaalde drempelnorm.

Dit alles maakt dat het toetsen van begrijpend lezen in grootschalige gestandaardiseerde screenings (zoals PIRLS en PISA) veel meer voor de hand ligt dan de toetsing van spreken, luisteren en schrijven. Bij begrijpend lezen kan met relatief waterdichte beoordelingswijzers worden gewerkt (meer dan bij schrijven en spreken), begrijpend-leestoetsen kunnen qua afname makkelijker uniform worden georganiseerd dan spreken en luisteren (zo kan bijvoorbeeld de invloed van de toetsafnemer sterker worden uitgeschakeld), begrijpend-leestoetsen vereisen geen inzet van audio-apparatuur (wat bij luisteren en spreken vaak het geval is), en ze hebben bij individuele afnames een hogere 'face validiteit' in termen van functionaliteit en levensechtheid dan spreek-, luister- of schrijftoetsen.

DIAGNOSE OP LEERLINGNIVEAU

Zoals eerder gesteld, bestaat er een fundamenteel verschil tussen screening en diagnose. De beide termen zijn echter geen complete tegengestelden, maar vormen eerder de uiteinden van een lang continuüm. Aan het ene einde van het

continuüm ligt een vorm van screening die een signaal geeft over de relatieve taalvaardigheid (of groei van taalvaardigheid) van de leerling, maar verder zeer weinig inzicht biedt in waar nu precies het probleem van de leerling inzake taalvaardigheid ligt. Eén van de meest prototypische voorbeelden van een dergelijke signaleringstoets is de edittoets waarbij leerlingen overtollige woorden uit een tekst moeten schrappen. Los van alle voordelen die de edittoets biedt (inzake hoge betrouwbaarheid, klassikale afname, empirisch-gevalideerde link met schoolsucces) biedt een edittoets bijzonder weinig houvast over welke (deel)aspecten van taalvaardigheid de leerling wel/niet heeft verworven, welke processen de leerling doorloopt bij het uitvoeren van taaltaken, en welke vorm van ondersteuning de leerling kan helpen om tot een versnelde groei van taalvaardigheid te komen. Dat heeft enerzijds te maken met het feit dat vooralsnog niet duidelijk is welke deelprocessen of taalvaardigheden de edittoets precies aanspreekt, en het zelfs onduidelijk is of het oplossen van een edittoets niet veeleer algemeen-cognitieve ('schoolse') denk- of toetsoplossingsvaardigheden en -strategieën vereist, eerder dan puur talige kennis, vaardigheden en attitudes.

KENMERKEN VAN KWALITEITSVOLLE DIAGNOSE

Aan het andere uiterste van het continuüm ligt een volwaardige diagnose, die toelaat een zicht te krijgen op zowel het proces en product van taakuitvoering door de leerling, waarbij de leerkracht inzicht verwerft in de (deel)processen, strategieën, (deel)vaardigheden, kennis en attitudes die nadere ondersteuning behoeven in zulke mate dat gerichte ondersteuning kan worden geboden. Zulk een volwaardige diagnose heeft als voornaamste kenmerken dat ze:

- (1) *dynamisch* is – verbonden aan processen van taakuitvoering – en product en proces voortdurend aan mekaar linkt (hoe komt het dat de leerling bij het uitvoeren van deze taak dit gedrag vertoont?);
- (2) *interactief* is, in de zin dat vooral via de interactie tussen leerling en degene die de diagnose stelt, kan worden scherpgesteld waar zich het probleem stelt (en ook wat de leerling al goed doet/kan/weet), en dat via de interactie interne processen naar boven worden gebracht (geëxpliciteerd);
- (3) *zowel taakspecifiek als taakoverschrijdend* wordt uitgevoerd, dit met het doel te bepalen of het probleem dat zich stelt, enkel optreedt bij het uitvoeren van specifieke taken die aan bepaalde kenmerken voldoen (bijvoorbeeld spreken in grotere groepen over een abstract onderwerp) of systematisch optreden over taaktypes (en eventuele vaardigheden) heen.

Voor dit soort van diagnose wordt in de pedagogische literatuur en taalwetenschappelijke literatuur vooral gewezen in de richting van probleemanalyserende en leerbevorderende interacties tussen leerkracht en leerling, bijvoorbeeld aan de hand van mediërende ondersteuning die de leerkracht tijdens leertaken aan leerlingen aanbiedt, of aan de hand van hardopdenkprotocollen waarbij leerlingen na de taakuitvoering expliciteren hoe ze de taak hebben aangepakt, tegen welke problemen ze zijn aangestoten, hoe ze die problemen hebben proberen aan te pakken, etc. In de huidige literatuur rond

In de huidige literatuur rond alternatieve en brede evaluatie wordt sterk gehamerd op het belang van het integreren van leren en evalueren.

Van de voorhanden zijnde toetsen en leerlingvolgsystemen ambiëren slechts een paar instrumenten om verder te gaan dan de eerste stap van de screening.

alternatieve en brede evaluatie wordt sterk gehamerd op het belang van het integreren van leren en evalueren, het bezorgen van feedback op toetsprestaties van de leerling, het uitgebreid en analyserend nabespreken van taakuitvoeringen, en dergelijke. Ook dit past in het interactief en dynamisch diagnosticeren van taalproblemen.

ZIJN TAALTOETSEN DIAGNOSTISCH?

De vraag die zich hier stelt, is waar taaltoetsen zich bevinden op dit continuüm. In principe kunnen zij zich, al naargelang de aard van de toets, op verschillende punten op het continuüm bevinden, al kan worden vastgesteld dat de huidige bestaande taaltoetsen zich veeleer aan de kant van de pure screening bevinden dan aan de kant van de volwaardige diagnose. Enkel in de leerlingvolgsystemen voor technisch lezen en spellen wordt een stevige aanzet tot handelingsgerichte diagnose na signalering geboden. Het nadeel van deze instrumenten is evenwel dat zij zich richten op deelaspecten van de functionele vaardigheden waar het in de eindtermen om draait, namelijk technisch lezen als deelvaardigheid van begrijpend lezen, en spelling als deelvaardigheid van schrijven. De toetsen en leerlingvolgsystemen die ernaar streven om functionele taalvaardigheid in kaart te brengen, geven expliciet aan dat hun eerste functie die van screening (signalering) is, en dat zij slechts karige info bieden op het vlak van diagnostiek. Zo vermeldt bijvoorbeeld de handleiding van de Cito-toetsen begrijpend lezen expliciet dat deze toetsen bij het diagnosticeren van problemen een rol kun-

nen spelen, maar dat ze op zich genomen niet diagnostisch zijn.

Van de voorhanden zijnde toetsen en leerlingvolgsystemen ambiëren slechts een paar instrumenten om verder te gaan dan de eerste stap van de screening. Zo vermelden TIST (een Vlaamse toets voor taalvaardigheid aan het begin van het eerste jaar secundair onderwijs), Diataal/Diatekst (een Nederlandse toets voor hetzelfde moment) en TASAN (een Vlaamse toets voor taalvaardigheid aan het einde van de onthaalklas secundair onderwijs) dat ze kansen tot diagnostiek bieden. De manier waarop dit bewerkstelligd wordt, is in deze gevallen dezelfde: door het stellen of analyseren van verschillende types van vragen of opdrachten bij de taakuitvoering, wordt verwacht dat een scherper inzicht wordt verkregen in welke subvaardigheden, kennis en attitudes de leerling wel/niet heeft verworven.

VOORBEELDEN VAN DIAGNOSTISCHE TOETSEN

Bij Diatekst bijvoorbeeld wordt dit verkregen door vragen te stellen op drie niveaus: het 'microniveau', het 'mesoniveau' en het 'macroniveau'. Het *microniveau* heeft betrekking op het woordniveau en de inhoudelijke details van een tekst. Aan de hand van dit niveau kan volgens de makers gekeken worden in hoeverre leerlingen woorden herkennen en begrijpen en of zij in staat zijn om tekst te begrijpen wanneer daar complexe grammaticale structuren in gebruikt worden. Het *mesoniveau* heeft te maken met de samenhang tussen de verschillende delen van een tekst. Hier gaat het om de herkenning van functiewoorden en structuuraanduiders. Er kan bijvoorbeeld gevraagd worden naar het verband tussen twee zinnen. Het *macroniveau* ten slotte heeft betrekking op de tekst als geheel. Hier

wordt bijvoorbeeld gevraagd naar de hoofdgedachte van de tekst of de leerling moet een andere titel voor de tekst kiezen. Ook de functie van de tekst kan bij een vraag op macroniveau aan de orde komen.

Diatekst vervult dus met andere woorden zijn diagnostische functie door leerlingen vragen van verschillende aard te stellen, en te analyseren hoe de leerlingen daarop scoren. Dat komt sterk overeen met wat de bovenvermelde toetsen begrijpend lezen van het Cito ook doen, waarbij de laatste organisatie toegeeft dat dit niet met echte diagnostiek gelijkgesteld mag worden. Dat laatste heeft vooral te maken met het feit dat het stellen van verschillende soorten vragen hoegenaamd geen garantie op inzicht biedt in het denkproces dat de leerling heeft doorgemaakt om voor een bepaald antwoord te kiezen en de strategieën die hij al dan niet heeft toegepast. Bovendien mag het fout beantwoorden van bijvoorbeeld een woordenschatvraag op microniveau niet blindelings worden gelijkgesteld met een woordenschatprobleem. Zo wijst de auteur van Diatekst (Hacquebord, 1989; 1996) er zelf op dat de verschillende begripsniveaus van deze test niet als afzonderlijke deelvaardigheden moeten worden gezien, maar als interagerende aspecten van het leesproces. Begrip van microniveau kan immers tot stand komen op basis van macrobegrip en andersom.

TIST en TASAN, twee Vlaamse toetsen voor het secundair onderwijs, bieden de gebruiker naast een totaalscore op de toetstaken ook subscores op verschillende categorieën, die samenhangen met verschillende aspecten van de toetstaken, namelijk: 'vaardigheid' (welke vaardigheid wordt aangesproken, bijvoorbeeld receptief of productief taalgebruik?), 'talige ondersteuning' (biedt de tekst veel redundantie?), 'verwerkingsniveau' (welk

verwerkingsniveau (beschrijvend, structurend, etc.) wordt aangesproken?), en 'perspectiefneming' (abstracte info versus concrete info?). Wat voor Diatekst geldt, geldt ook hier. De subscores kunnen zeker een eerste aanzet tot diagnostische informatie bieden, maar dit zal om de hierboven genoemde redenen via meer procesgerichte dialoog en via observatie van de leerling tijdens taakuitvoeringen aangevuld moeten worden om het probleem van de leerling echt scherp te krijgen. Samenvattend betwijfelt Alderson (2000) dan ook sterk of men via toetsing lees- en andere taalproblemen kan diagnosticeren: zelfs als er aparte deelvaardigheden zouden meespelen in leesprestaties, dan is volgens hem quasi onmogelijk om die te onderscheiden via het gebruik van toetsen.

De bestaande taaltoetsen besteden bovendien bijzonder weinig aandacht aan de attitudes die in de taakuitvoering van taal taken van bijzonder groot belang kunnen zijn. Variabelen zoals motivatie, welbevinden, betrokkenheid, zelfsturing, en andere socio-emotionele factoren die creatief probleemoplossend gedrag kunnen beïnvloeden (zoals impulsiviteit of planmatigheid) worden niet in kaart gebracht. Het feit dat de hierboven vermelde diagnostische toetsen slechts focussen op een klein segment van de mogelijke probleemfactoren, maakt dat een rechtstreekse link met onderwijs-ondersteuning, puur op basis van wat de toets aan (diagnostische) informatie oplevert, een beweging op zeer glad ijs dreigt te worden.

Het stellen van verschillende soorten vragen biedt hoegenaamd geen garantie op inzicht in het denkproces dat de leerling heeft doorgemaakt om voor een bepaald antwoord te kiezen.

BESLUIT

Dit alles maakt dat taaltoetsen slechts een eerste stap kunnen zijn in de richting van gerichte, gedifferentieerde ondersteuning van leerlingen van wie via screening is vastgesteld dat de ontwikkeling traag of problematisch verloopt, of dat een bepaald niveau van presteren op een bepaald moment niet is gehaald. Screening en eerste diagnose via taaltoetsing kan het voordeel bieden dat leerkrachten niet alleen het probleem zien, maar ook zelfs een zeker beeld krijgen van de grootte van het probleem (bijvoorbeeld de afstand tussen de gemiddelde prestatie en de prestatie van bepaalde individuele leerlingen), maar om het probleem op een efficiënte en effectieve manier aan te pakken, moet de leerkracht alle kwalitatieve informatie die hij reeds heeft over het talig functioneren van de leerling inzetten, of bijkomende procesmatige informatie verzamelen. De evaluatie moet na het moment van de screening dus sterk 'verbreed' worden met andere brede evaluatietechnieken zoals observatie, interactieve mediatie, portfolio, zelfevaluatie door de leerling en dergelijke. De rol van sterk gestandaardiseerde toetsen is op dat moment grotendeels uitgespeeld; de rol van gepersonali-seerde, handelingsgerichte diagnose wordt dan des te belangrijker.

Voor de screening en diagnose op leerling-niveau kunnen we dan ook besluiten dat:

- (1) screening van taalvaardigheid ertoe kan leiden dat leerkrachten de vinger aan

de pols houden van de groei van hun leerlingen, en tijdig alarmsignalen kunnen verkrijgen over leerlingen bij wie de ontwikkeling te traag verloopt;

- (2) die screening van taalvaardigheid met taaltoetsen kan worden uitgevoerd, bij voorkeur met toetsen die een hoge mix van betrouwbaarheid, validiteit, efficiëntie en een kleine kans op negatieve backwash-effecten vertonen. Op dit vlak scoren doorgaans begrijpend-lees-toetsen relatief beter dan spreek-, luister-, schrijf-, deelvaardigheids- en edittoetsen;
- (3) taaltoetsing die met relatieve tussenpozen wordt uitgevoerd, het voordeel biedt dat de leerling in zijn groei gevolgd kan worden, maar dat gestandaardiseerde screening via taaltoetsing voor de leeftijd van 5-6 jaar niet altijd leidt tot echt betrouwbare en valide informatie, en niet tot gedetailleerde informatie in verband met diagnostisering en ondersteuning;
- (4) taaltoetsen algemeen op het vlak van diagnose van problemen en dus gerichte aanduidingen voor daadwerkelijke gedifferentieerde ondersteuning van individuele leerlingen slechts een beperkte waarde hebben; hiervoor zal nodig zijn dat de info uit taaltoetsen sterk aangevuld wordt met kwalitatieve, procesmatige informatie uit observaties van een breed scala van taak- en deeltaakuitvoeringen en interactieve procesanalyses.

*Kris Van den Branden
Centrum voor Taal en Onderwijs
Blijde-Inkomststraat 7
3000 Leuven
kris.vandenbranden@arts.kuleuven.be*

Noot

- 1 Dit artikel is een samenvatting van het eerste deel van hoofdstuk 2 van Colpin e.a. (2006). De volledige tekst van dit hoofdstuk (inclusief alle bibliografische referenties) kan gratis gedownload worden op de website van het Centrum voor Taal en Onderwijs (www.cteno.be).

Referenties

Om de leesbaarheid van dit artikel te verhogen, werd ervoor geopteerd om de vele bibliografische referenties die in de oorspronkelijke tekst van de haalbaarheidsstudie zaten, niet op te nemen. U kunt de integrale tekst met referenties downloaden op [<www.cteno.be>](http://www.cteno.be). Hieronder verwijzen we enkel naar de werken die in dit artikel geciteerd werden.

Alderson, J. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Bijslma, B., (2004). *Het kiezen van een kindvolgsysteem*. Leeuwarden: Partoer.

Colpin, M., Gysen, S., Jaspaert, K., Heymans, R., Van den Branden, K., Verhelst, M. (2006). *Studie naar de wenselijkheid en haalbaarheid van de invoering van centrale taaltoetsen in Vlaanderen in functie van gelijke onderwijskansen*. Leuven: Centrum voor Taal en Onderwijs.

Colpin, M., Heymans, R., Rymenans, R., Daems, Fr. & Van den Branden, K. (2005). *Ontwikkeling van een instrument voor periodiek peilingsonderzoek schrijven (PPON-S). Eindrapport*. Leuven/Antwerpen: UA en K.U.Leuven.

Commissie Indicatiestelling Onderwijsachterstanden (1996). *Zo onvoorspelbaar als het leven zelf. Advies van de Commissie Indicatiestelling Onderwijsachterstanden uitgebracht aan de Staatssecretaris van Onderwijs, Cultuur en Wetenschappen, mevr. T. Netelenbos*. Den Haag: OC&W/Sdu.

Hacquebord, H. (1989). Tekstbegrip van Turkse en Nederlandse leerlingen in het voortgezet onderwijs. *Studies over Taalgebruik, 6*, Dordrecht: Foris.

Hacquebord, H., (1996). *Tekstbegriptoets voor de Brugklas. Instrument voor signaleren en typeren van leesproblemen. Handleiding*. 's Hertogenbosch: KPC.

Onderwijsraad (2002). *Wat scholen vermogen*. Den Haag: Onderwijsraad.

Verhelst, M. (2002). *De relatie tussen mondeling taalaanbod en woordenschatverwerving van het Nederlands als tweede taal door 2,5-jarige allochtone kleuters in Brussel*. Doctoraatsproefschrift. Leuven: K.U.Leuven.

Verhoeven, L. & Vermeer, A. (2005). Het ongelijk van Netelenbos. Toetsing van kleuters en hun prestaties op de Cito Eindtoets Basisonderwijs. *Toegepaste Taalwetenschap in Artikelen*, 74, p. 123-134.