

# Naar een leerlingvolgsysteem voor het basisonderwijs

P. GILLIJNS

*Cito, Arnhem*

L. VERHOEVEN

*Katholieke Universiteit Brabant, Tilburg*

## Samenvatting

*Bij het verschijnen van de ARBO-nota "Voorrang aan achterstand" (1988) doet de term 'Leerlingvolgsysteem' zijn intrede in het Nederlandse onderwijs. Hoewel het begrip slechts een korte geschiedenis kent, heeft het zich zeer snel verspreid. In hoog tempo zijn uiteenlopende leerlingvolgsystemen ontwikkeld en aan het basisonderwijs aangeboden. In dit artikel presenteren we een schets van een leerlingvolgsysteem dat momenteel vanuit het Cito ontwikkeld wordt. We gaan eerst kort in op de achtergronden, het doel en de functie van leerlingvolgsystemen. Daarna geven we een beknopte, algemene schets van het Cito-Leerlingvolgsysteem. Welke informatie zo'n leerlingvolgsysteem voor de onderwijspraktijk oplevert, wordt geïllustreerd aan de hand van de resultaten van een normeringsonderzoek in groep 3 met een schaal voor spelling. In een discussie worden ten slotte de mogelijkheden en beperkingen van leerlingvolgsystemen afgewogen.*

## 1 Achtergronden

Uit onderzoek van de laatste tien jaar, blijkt dat het met de praktijk van het evalueren in het Nederlandse onderwijs weinig rooskleurig gesteld is. Onderzoek van Janssens (1986) wijst uit dat leerkrachten in het basisonderwijs over het algemeen een zeer geringe deskundigheid op het gebied van evalueren hebben en dat er op twee derde deel van de onderzochte scholen geen beleid ten aanzien van evalueren bestaat. Verder wordt door Kremers (1984) en Alberts en Kuijpers (1986) de evaluatiepraktijk in het voortgezet onderwijs als 'traditioneel' gekwalificeerd. Schriftelijke overhoring ten behoeve

van resultaatbepaling zijn weliswaar gemeengoed, maar andere vormen van evalueren komen weinig voor. In Britse en Amerikaanse literatuur worden soortgelijke bevindingen gerapporteerd, al concentreert het onderzoek zich daar vooral op het gebruik van gestandaardiseerde toetsen (Stiggins, Conklin & Bridgeford, 1986). Het belang van een planmatige en valide evaluatie wordt door de onderwijskundige ontwikkelingen van de laatste jaren steeds groter. Differentiëren, individualiseren, het vergroten van de zorgbreedte, of het creëren van een ononderbroken leerproces, vereisen een goed 'management van leervorderingen' (Janssens & Moelands, 1989) en kunnen feitelijk pas plaats vinden, wanneer leerkrachten zich een juist beeld vormen van de mogelijkheden en behoeften van leerlingen, de leervorderingen op de voet volgen en het onderwijsaanbod daarop afstemmen. In meerdere onderzoeken (Kool, 1982; Van der Leij, Kool & Wielenga, 1984; Van der Wolf, 1986; Janssens, 1986) is aangetoond dat leerkrachten vaak een globaal en weinig adequaat beeld hebben van de leervorderingen van hun leerlingen en dientengevolge het leerproces niet optimaal af kunnen stemmen op de instructiebehoefte. Dergelijke discrepanties tussen beeldvorming en feitelijke leervorderingen kunnen bij overschatting, gemakkelijk leiden tot 'didactische verwaarlozing' (Aarnoutse & Mommers, 1987). Wanneer leervorderingen worden onderschat is het gevaar van een 'pygmalion-effect' (Jungbluth, 1985) niet denkbeeldig, zeker wanneer leerkrachten onderwijsbeslissingen mede baseren op persoonlijke opvattingen en achtergrondinformatie van leerlingen en het onderwijsaanbod bij voorbaat afstemmen op negatieve verwachtingspatronen (Mens & Van Calcar, 1981). In ieder geval laat onderzoek ook zien dat goede en systematische evaluatieprocedures en het hanteren van een deugdelijk, uniform registratiesysteem niet alleen een positieve invloed hebben op het inzicht in leervorderingen (Van der Leij e.a. 1984; Janssens, 1986), maar ook op de leervorderingen zelf (Kool, 1982, Lowyck, 1983, Brandsma & Knuver, 1990, Scheerens, 1990).

De relatie tussen het systematisch volgen en vastleggen van de vorderingen van leerlingen en de effectiviteit van het onderwijs, kreeg vooral veel aandacht door het rapport "What works, research about teaching and learning" (United States Department of Education, 1986). Daarin werd gerapporteerd over Amerikaans onderzoek naar de kenmerken van effectieve scholen. In Nederland vond het rapport veel weerklank. Appelfhof kwam in 1986 tot de aanbeveling, dat iedere basisschool een systeem ontwikkelt, "waarmee de schoolgang van leerlingen gevolgd kan worden, zodat eventuele problemen vroegtijdig worden gesignaleerd" (Appelfhof, 1986, p. 35). In dat kader introduceerde de Adviesraad voor het basisonderwijs, speciaal onderwijs en voortgezet speciaal onderwijs (ARBO) in 1988 de term 'leerlingvolgsysteem'. In het buitenland bestaan dan al soortgelijke systemen. In de Verenigde Staten, in het district Portland (Oregon) bijvoorbeeld, wordt in 1977 al een omvangrijk "Achievement Testing Program" voor lezen, rekenen en later ook taal geïntroduceerd (Hathaway, 1975, 1981). Ook in andere districten worden dergelijke systemen ontwikkeld (LeMahieu & Wallace, 1986; Olsen, Cox, Price, Strozeski & Vela, 1990; Willis, 1990). In Engeland wordt een 'monitoring system' voor lezen en taal toegepast (Appelfhof, 1987). In 1988, in de nota "Voorrang aan achterstand", introduceerde de ARBO het begrip leerlingvolgsysteem, dat aanvankelijk omschreven wordt als een systeem waarbij "op geregelde wijze en op betrouwbare wijze het vorderingenniveau van de leerlingen wordt bepaald, om op basis daarvan het onderwijsaanbod beter te richten en de pedagogische aanpak en de didactische werkvormen toe te spitsen op de individuele behoeften" (ARBO, 1988a, p. 33). Volgens een recente telling bestaan in Nederland al meer dan 50 grotere en kleinere leerlingvolgsystemen! Deze systemen verschillen onderling op wezenlijke punten in opzet, functie en gebruiksmogelijkheden (Gillijns, in voorbereiding). De meeste leerlingvolgsystemen maken gebruik van bestaande toetsen en sluiten qua opzet en werkwijze aan bij het 'groepsgewijs schoolonderzoek' (Van der Leij, 1985), zoals dat in het speciaal onderwijs bestaat. De leervorderingen worden in de meeste gevallen gevolgd door ruwe toetsscores te transformeren naar didactische leeftijds-

equivalenten (Moelands, Mommers & Oud, 1990). Dergelijke systemen kunnen relatief snel ontwikkeld worden. Ze sluiten nauw aan bij gangbare evaluatieprocedures en -midelen, hetgeen vanuit oogpunt van implementatie wenselijk is. Anderzijds bestendigen ze het feilen van de bestaande evaluatiepraktijk, zeker wanneer verouderde of ondeugdelijk genormeerde toetsen in het systeem worden opgenomen. Bovendien zijn transformaties naar didactische leeftijdsequivalenten allerm minst onomstreden (Oud & Mommers, 1988).

## 2 Kenmerken van een leerlingvolgsysteem

### 2.1 Definiëring

Niet iedere willekeurige verzameling toetsen kan als een leerlingvolgsysteem worden gekwalificeerd. Een leerlingvolgsysteem wordt gekenmerkt door een aantal specifieke eigenschappen. In algemene zin is het te definiëren als een verzameling van meetinstrumenten die een lange-termijn-evaluatie (Janssens, 1985) op uiteenlopende leerinhouden mogelijk maken van zowel individuele als groepen leerlingen. De termijn lange-termijn-evaluatie verwijst naar een signalering van leervorderingen over langere perioden, dat wil zeggen hooguit enkele malen per jaar. In het advies "Regulier en speciaal onderwijs in samenhang" (1988b) heeft de ARBO het begrip leerlingvolgsysteem nader uitgewerkt en gedefinieerd als "een concreet hulpmiddel voor het signaleren van het achterblijven van de vorderingen ten opzichte van de door de school nagestreefde doelen en tussendoelen en voor het registreren van die vorderingen" (p. 64). Het feit dat een leerlingvolgsysteem gericht is op lange-termijn-evaluatie impliceert dat de aanwijzingen voor diagnostisering en remedieering beperkt zijn. Leerlingvolgsystemen bestrijken immers betrekkelijk grote leerstofgebieden waardoor slechts globale informatie omtrent eventuele hiaten in de beheersing bij leerlingen kan worden opgespoord. Dit neemt echter niet weg dat een leerlingvolgsysteem zodanig kan worden opgezet dat het redelijk zicht biedt op de sterke en zwakke kanten van de leerling. Dit kan bijvoorbeeld door aan een bepaald meetinstrument vervolgttoetsen te koppelen die belangrijke deelcomponenten van de getoetste vaardigheid meten. Zo lijkt het zinvol om aan een schaal voor technisch

lezen in de aanvangsfase taken voor grafeemkennis en auditieve synthese als vervolgoetsen toe te voegen. Het leren beheersen van de elementaire leeshandeling veronderstelt immers het vlot en correct koppelen van grafemen aan fonemen en het samenvoegen van fonemen tot woordpatronen (Doctor & Coltheart, 1980). In de handleiding bij een leerlingvolgsysteem kunnen op deze wijze procedures voor diagnostisering worden aange-reikt samen met aanwijzingen voor remediëring.

## 2.2 Meetprocedures

Een leerlingvolgsysteem vereist naast een groot aantal inhoudelijke en praktische eisen vooral ook psychometrische eisen (Gillijns, in voorbereiding). Het betrouwbaar volgen van individuele leervorderingen, de kern van een leerlingvolgsysteem, is geen sinecure. Het probleem van de typische onbetrouwbaarheid van individuele meetresultaten is lange tijd 'opgelost' door uit te gaan van statistische momentopnames van de relatieve positie van leerlingen binnen een groep. Transformaties naar didactische leeftijdsequivalenten is zo'n techniek. Een adequaat leerlingvolgsysteem moet echter voorzien in longitudinale vorde-ringenbepaling. Dit vereist een psychometri-sche schaling van toetsscores. Hiervoor zijn op dit moment twee concurrerende proce-dures gangbaar: LISKAL en itemrespons-

theorie. LISKAL is een computerprogramma waarin elementen van LISREL en het Kal-man-filter zijn geïntegreerd (Oud, Van Ber-cken & Essers, 1986). LISREL (Jöreskog & Sörbom, 1981) maakt het mogelijk relaties tussen variabelen in de tijd te onderzoeken. Het model voorziet in een meetcomponent en een structurele component. Aan de hand van de meetcomponent kan worden bepaald welke hypothetische constructen ten grondslag liggen aan een verzameling toetsscores die op een bepaald meetmoment zijn afgenomen. Met behulp van de structurele component kan ver-volgens worden nagegaan in welke mate de score op een latente variabele voorspeld kan worden vanuit dezelfde variabelen op een eer-der tijdstip gemeten of vanuit andere variabe-len (Oud, 1990). In het programma worden longitudinale factorscores geschat met behulp van het Kalman-filter. Dit is te beschouwen als een combinatie van predictieve en cross-

sectionele schatters. De optimaliteit van het filter maakt schattingen mogelijk die bijna altijd kleinere standaardfouten opleveren dan predictieve of cross-sectionele schattingen.

Ook met behulp van itemresponsmodellen al dan niet in combinatie met groei modellen is het mogelijk de groei van afzonderlijke leerlin-gen op adequate wijze te beschrijven. Deze modellen geven een meettheoretische model-specificatie van het toetsgedrag van leerlingen (Verhelst & Eggen, 1989). Binnen zo'n model geldt dat de kans op het oplossen van een opgave uitsluitend wordt bepaald door de vaardigheid van de leerling en de moeilijkheid van de opgave. Daarbij geldt de aanname dat aan het oplossen van de toetsopgaven slechts een persoonsvariabele ten grondslag ligt. De verzameling opgaven op een schaal is immers homogeen te noemen. Verder geldt als belang-rijke eigenschap dat opgavenparameters onaf-hankelijk van persoonsparameters zijn te schatten. Bij het schalen van toetsopgaven krijgen leerlingen in de loop van de tijd vaak opgavenverzamelingen met gedeeltelijke overlap (ankeropgaven) voorgelegd. Door middel van equivalering van scores kunnen zo opgavenparameters die op verschillende tijd-stippen zijn bepaald tot een gemeenschappe-lijke schaal worden herleid (Verstralen, 1984; Glas & Verhelst, 1984).

Een sterke kant van de LISKAL-benade-ring is dat de leervorderingen van leerlingen niet alleen kunnen worden gevolgd, maar dat het tevens mogelijk is een voorspelling te doen over de te verwachten ontwikkeling van de leervorderingen. Een minder sterke kant is dat het proces van verandering of groei niet expli-ciet in het gehanteerde model wordt gespecifi-ceerd (Voeten, 1990). Het model, dat in feite gebaseerd is op de klassieke testtheorie, heeft betrekking op individuele verschillen, maar laat de intra-individuele variabiliteit niet aan bod komen. "Men kan zich afvragen of het gerechtvaardigd is relaties, vastgesteld op inter-individueel niveau zonder meer door te trekken naar intra-individueel niveau" (Voe-ten, 1990, p. 127). Verder kunnen ook bij de inhoudsvaliditeit van de procedure vraagte-kens worden geplaatst, in zoverre dat met latente variabelen wordt gewerkt die niet in psychometrisch opzicht 'geschaald' zijn.

Schalen op basis van de eisen van de item-responstheorie komen aan deze bezwaren van de klassieke testtheorie tegemoet. Hiermee

kunnen de leervorderingen in absolute, maar ook in relatieve zin gevolgd worden. Het is verder mogelijk om met een beperkte deelverzameling van toetsopgaven uitspraken te doen omtrent de relatieve schaalpositie die een leerling inneemt. Door het hanteren van groeimodellen en filterprocedures (Kamphuis & Moelands, 1990) kan de betrouwbaarheid van vaardigheidsschattingen aanzienlijk worden verhoogd en is eveneens een predictie van de te verwachten vaardigheidsontwikkeling mogelijk. Het gebruik van een geautomatiseerd verwerkings- en registratieprogramma is daarbij overigens noodzakelijk. De itemresponsstheorie stelt echter aan ieder item een strenge homogeniteitseis. Wanneer dat zou leiden tot de onderverdeling in een relatief groot aantal domeinen, kan dat de praktische bruikbaarheid van het systeem schaden. In hoeverre het mogelijk is voor een beperkt aantal domeinen grote verzamelingen van items samen te stellen die aan de eisen voldoen is slechts proefondervindelijk vast te stellen.

### 3 *Het Leerlingvolgsysteem van het Cito*

#### 3.1 *Uitgangspunten*

Het leerlingvolgsysteem dat het Cito ontwikkelt, zal een serie schalen omvatten voor de *lange-termijn-evaluatie* en een systeem voor handmatige of geautomatiseerde registratie van leervorderingen. Het systeem is uitdrukkelijk bedoeld als een *hulpmiddel*, om op *onderdeelniveau* de vorderingen vast te stellen in de basisvaardigheden taal/lezen, rekenen, wereldoriëntatie en informatieverwerking (zie Tabel 1). Een à twee maal per jaar kunnen *globaal* de leervorderingen van leerlingen worden vastgesteld en kan de leerkracht bepalen wat de relatief sterke en zwakke kanten van een leerling zijn en of de leerstof daarop is afgestemd. Gemiddelde scores op groeps- of schoolniveau kunnen aanleiding zijn tot een herbezinning van de door de school gevolgde werkwijze en op die manier de aanzet vormen tot schoolzelfevaluatie (Van der Schoot, 1991), zeker wanneer na herhaalde afnames, toevallige fluctuaties zijn geëlimineerd en er een trend zichtbaar wordt.

Een dergelijk leerlingvolgsysteem zal in eerste instantie een signaalfunctie hebben. Wanneer de ontwikkeling van leerlingen niet naar wens verloopt of zelfs stagneert is het

Tabel 1 *Overzicht van het toetsinstrumentarium in het Cito-Leerlingvolgsysteem*

Vaardigheid	Leerjaar
ordenen	1 - 2
algemene taalvaardigheid	4 - 8
woordenschat	3 - 8
luistervaardigheid	3 - 8
technisch lezen	3 - 8
begrijpend lezen	3 - 8
spellen	3 - 8
stellen	6 - 8
informatieverwerking	6 - 8
wereldoriëntatie	6 - 8
rekenen	3 - 8

verzamelen van aanvullende informatie gewenst, enerzijds om het signaal te verifiëren en anderzijds om specifieke problemen of hiaten te lokaliseren. Volgens Van Dongen (1984) moet daarbij zowel gekeken worden naar andere aspecten binnen en buiten het kind als naar onderliggende factoren. Het kijken naar andere aspecten binnen en buiten het kind, zal de leerkracht grotendeels door middel van gerichte observaties zelf moeten uitvoeren, al biedt het leerlingvolgsysteem de mogelijkheid om de vaardigheid op verwante taken te analyseren. Zo kan een score op begrijpend lezen bijvoorbeeld gerelateerd worden aan de score op technisch lezen. Bovendien kan de leerkracht de in het registratiesysteem gedocumenteerde leerroute of schoolloopbaan analyseren (Janssens, 1985), door scores op een en dezelfde schaal in de tijd te vergelijken. Om informatie over onderliggende factoren te kunnen verzamelen worden aan sommige schalen vervolgotoetsen toegevoegd. Dat zijn diagnostisch gerichte en individueel af te nemen criteriumtoetsen (zie ook 2.1). Nadat aanvullende informatie verzameld is, kan de leerkracht in principe geschikte leertaken selecteren. In het leerlingvolgsysteem van het Cito worden, daar waar mogelijk, didactische aanwijzingen en oefenvormen beschreven die leerkrachten kunnen gebruiken ten behoeve van remediering. In een aantal gevallen wordt tevens verwezen naar aanwijzingen en oefeningen in bestaande onderwijsmethoden.

#### 3.2 *Onderzoeksopzet*

Ten behoeve van de constructie van de schalen is onderzoek uitgevoerd bij een tweetal cohorten: een Rotterdams/Schiedams en een lande-

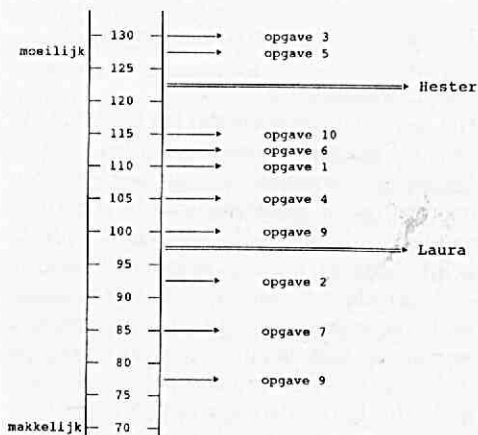


lijk cohort. Allereerst is in januari 1987 een cohort van circa 1800 leerlingen samengesteld uit scholen voor basis- en speciaal onderwijs in de gemeente Rotterdam en het onderwijsvoorzorgsgebied Schiedam. Bij deze leerlingen, die destijds in groep 3 zaten, zijn inmiddels gedurende een aantal jaren longitudinale data verzameld. Over het schooljaar verspreid, kregen zij voor een aantal leerstofgebieden opgaven voorgelegd. In de regel ging het om nieuwe opgaven, die vooraf waren uitgetoetst bij een landelijke steekproef van leerlingen. De data van dit cohort zijn gebruikt om schalen te construeren, van waaruit modules voor verschillende vaardigheidsniveaus zijn samengesteld. Dat resulteerde in enkele experimentele versies van schalen die ter beschikking zijn gesteld aan de Rotterdamse en Schiedamse basisscholen. Inmiddels zijn er ook enkele experimentele schalen van het leerlingvolgsysteem ingezet op basisscholen in de Westelijke Mijnstreek en op scholen die deelnemen aan het Experiment Gerichte Aanpak Amsterdamse basisscholen (EGAA-project) (Haanen, Roders, Lagendaal & Van der Wolf, 1990). Eind 1989 is een normeringsonderzoek bij een landelijk cohort van circa 3600 leerlingen gestart. Voor de samenstelling van dit cohort werd dezelfde procedure toegepast als bij het onderzoek in het kader van de Periodieke Peiling van het Onderwijsniveau (Wijnstra, 1988). De totale populatie van Nederlandse scholen werd, op basis van de schoolscore in het schooljaar 1989/1990, opgedeeld in drie strata van ongelijke omvang. In stratum 1 (schoolscores  $\leq 1.05$ ) zijn vooral scholen vertegenwoordigd met relatief weinig arbeiderskinderen en relatief veel kinderen uit gezinnen met een hogere sociaal-economische status. Scholen in stratum 3 (schoolscores  $\geq 1.15$ ) kenmerken zich door een relatief hoog aantal allochtone leerlingen. De scholen in stratum 2 (schoolscores 1.06 - 1.15) vormen een middenpositie tussen de beide andere strata. Per stratum werd een gelijk aantal scholen ( $N \approx 60$ ) getrokken. Het primaire doel van dit onderzoek is het verzamelen van landelijke referentiegegevens bij de verschillende schalen. Tevens wordt dit onderzoek gebruikt om reeds geconstrueerde schalen op hun geldigheid te onderzoeken, om aanvullende opgaven in te zetten, zodat bestaande schalen beter gevuld kunnen worden en om data te verzamelen voor de constructie van nieuwe schalen.

Om gefaseerd en op korte termijn meetinstrumenten uit te kunnen geven, worden in eerste instantie schalen voor groep 3 en 4 geconstrueerd. In een latere fase worden deze schalen zo mogelijk uitgebreid met opgaven voor groep 5 en 6 en voor groep 7 en 8, en worden additionele schalen geconstrueerd.

### 3.3 Meettechniek

In kort bestek lichten we de meettechniek toe die ten grondslag ligt aan het leerlingvolgsysteem. Voor nadere details verwijzen we naar Kamphuis en Moelands (1990). Door de toepassing van een itemresponsmodel, worden longitudinale data gecalibreerd, dat wil zeggen dat de moeilijkheidsgraad van de toetsopgaven en de vaardigheid van de leerlingen op een latente schaal worden afgebeeld (Zwarts, 1990). Die schaal is bij het leerlingvolgsysteem zo gekozen dat de gemiddelde moeilijkheidsgraad van de items gelijk is aan de waarde 100. In Figuur 1 staat een voorbeeld van een schaal met daarop afgebeeld 10 opgaven en de vaardigheidsschatting van twee fictieve leerlingen.



Figuur 1 Een voorbeeld van een vorderingschaal

De schaalwaarde van een opgave geeft aan dat een leerling met een overeenkomstige vaardigheid een kans van 50% heeft om deze of even moeilijke opgaven juist te beantwoorden. Naarmate de vaardigheid van een leerling verder beneden de schaalwaarde van een opgave ligt, wordt de succeskans steeds kleiner. Naarmate de vaardigheid verder daarboven ligt, wordt de succeskans steeds groter. In feite worden op deze manier gecalibreerde opgavenbanken ontwikkeld, waaruit toetsmodulen van uiteenlopende moeilijkheids-

graad zijn samen te stellen. Bij de samenstelling van de modules spelen naast psychometrische overigens ook leerstof-inhoudelijke argumenten een rol. Om de signaalfunctie van het systeem gestalte te geven, is het gewenst van elke inhoudelijke leerstofcategorie een voldoende aantal opgaven op te nemen (zie ook 3.1 en 4.4). Wanneer een schaal voldoende gevuld is, kan in principe voor elk vaardigheidsniveau een optimale toetsmodule worden samengesteld. Door opgaven te selecteren, waarvan de gemiddelde moeilijkheidsgraad het vaardigheidsniveau van de leerling benadert, krijgt men een schatting met een minimale meetfoutvariantie en dus de grootst mogelijke testinformatie. Op deze wijze kan de toetsprocedure worden verfijnd, zonder de toets te verlengen. Om te bepalen of een leerling een makkelijker of moeilijker module moet maken kan bijvoorbeeld een korte voortoets worden afgenomen (Lord, 1980). Het is ook mogelijk om de score van een vorig meetmoment hiervoor als criterium te gebruiken. Met behulp van een te ontwikkelen computerprogramma en een gecalibreerde opgavenbank kan een adaptieve toetsingsprocedure gerealiseerd worden (Kingsbury, 1990). Hierbij selecteert het programma steeds opgaven van een moeilijkheidsgraad die de lopende vaardigheidsschatting benadert. Kamphuis en Moelands (1990) hebben aangetoond dat de meetnauwkeurigheid bij zowel modulair als adaptief toetsen aanzienlijk toeneemt. De mogelijkheid tot adaptief toetsen met behulp van de computer zal, onder andere om financiële redenen, binnen het leerlingvolgsysteem vooralsnog echter niet gerealiseerd worden, maar kan in de toekomst mogelijk een prominente rol spelen. Een procedure van modulair toetsen, of 'toetsen op maat', is wel in het leerlingvolgsysteem opgenomen, zij het onder de voorwaarde dat de organisatorische consequenties voor de leerkracht tot een minimum beperkt blijven. Daarbij speelt ook de face-validity (Drenth, 1975) van de samen te stellen modules een rol. Om een maximale testinformatie te bereiken moet men leerlingen modules voorleggen met een gemiddelde oplossingskans van 50%. Leerkrachten geven leerlingen in de regel echter geen toetsen van een dergelijke moeilijkheidsgraad en vinden dat zelfs ongewenst.

### 3.4 Registratie en interpretatie

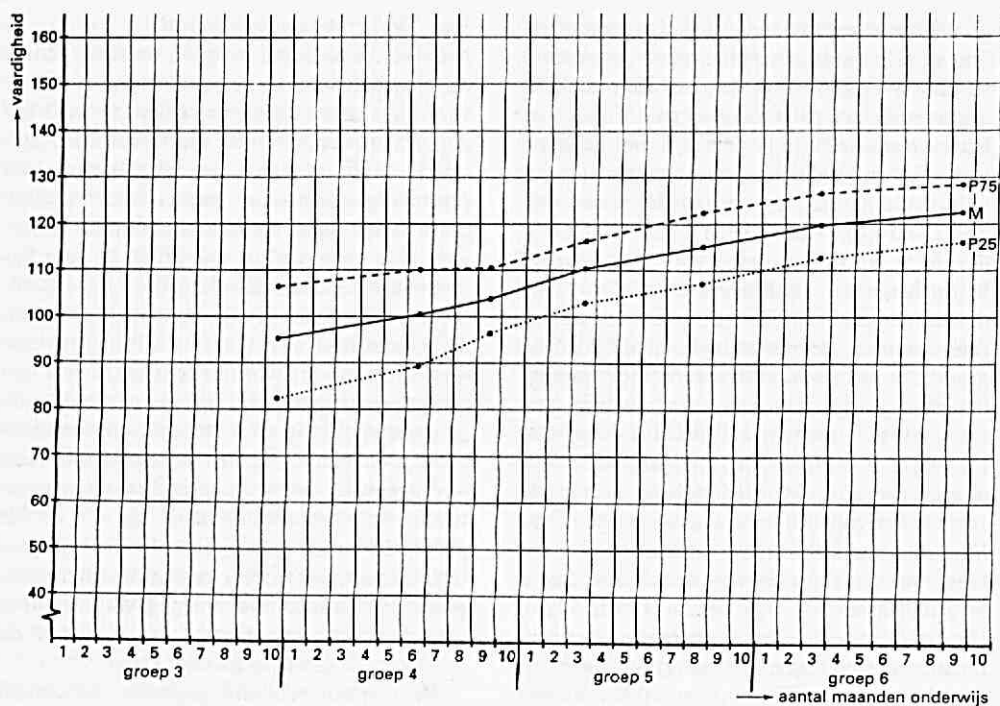
Na elke toetsafname worden de ruwe scores naar schaalscores getransformeerd. Deze vaardigheidsschattingen vormen de basis om de leervorderingen in de tijd te volgen. Registratie vindt plaats door schaalscores van opeenvolgende meetmomenten in een grafiek uit te zetten tegen het aantal maanden onderwijs. Hoe steiler de lijn, hoe groter de vaardigheidsontwikkeling. De data die bij verschillende deelpopulaties zijn verzameld, worden als referentiegegevens ter beschikking gesteld. Op transparanten worden de schaalscores van de referentiegroepen, op de verschillende meetmomenten in de vorm van curven afgebeeld. Per referentiegroep worden drie curven weergegeven, corresponderend met drie posities in de populatie, namelijk het 25<sup>e</sup> en 75<sup>e</sup> percentiel en het gemiddelde. De transparanten van een (of meer) relevante referentiegroep(en) kunnen over een individuele curve gelegd worden (zie Figuur 2), waardoor de interpretatie meer perspectief krijgt.

Er worden referentiegegevens verzameld bij de totale steekproef en bij de volgende relevante subpopulaties: autochtone en allochtone leerlingen en de drie, op basis van de schoolscore onderscheiden, strata van scholen. De wenselijkheid voor het geven van aparte normeringsgegevens voor allochtone leerlingen bij het leren lezen is eerder onderbouwd door Verhoeven (1987). Het gehanteerde itemresponsmodel maakt het bovendien tot op zekere hoogte mogelijk de positie van een leerling op de schaal te vertalen in al dan niet beheerste opgaven en dus in concrete onderwijsdoelstellingen. Daarbij moet overigens wel het kansmodel dat aan de schaal ten grondslag ligt betrokken worden.

## 4 Illustratie: het onderdeel spelling

Voor het aanvankelijk spellingonderwijs zijn drie experimentele meetinstrumenten geconstrueerd: de Schaal Vorderingen in Spellingvaardigheid-1 (in het vervolg de SVS-1) het Fonemendictee (in het vervolg FD) en de Toets voor Auditieve Analyse (in het vervolg TAA).

De SVS-1 is bedoeld voor het longitudinaal meten van de vaardigheid in het spellen van onveranderlijke één- en tweelettergrepige woorden bij leerlingen van groep 3 en 4. Het



Figuur 2 Een voorbeeld van de referentiegegevens die in het leerlingvolgsysteem gebruikt worden

P75 correspondeert met het 75<sup>e</sup> percentiel in de populatie

M correspondeert met het gemiddelde in de populatie

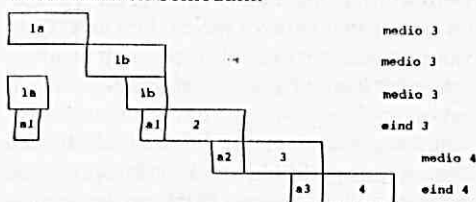
P25 correspondeert met het 25<sup>e</sup> percentiel in de populatie

FD en de TAA zijn bedoeld als vervolftoetsen, waarmee de beheersing van respectievelijk de klank-letter-koppeling en de vaardigheid in auditieve analyse kan worden vastgesteld. Deze beide vaardigheden vormen belangrijke deelcomponenten van spellingvaardigheid in de aanvangsfase van het onderwijs (Verhoeven, 1983). In het uiteindelijke leerlingvolgsysteem neemt de leerkracht deze toetsen individueel af bij leerlingen met een onvoldoende (vooruitgang in) spellingvaardigheid.

#### 4.1 Instrumentconstructie

Voor de constructie van de SVS-1 is gebruik gemaakt van opgaven uit de bestaande Cito-toetsen Woorddictee 1 (medio groep 3) en 2 (eind groep 3) (Verhoeven, 1980). Voor groep 4 zijn nieuwe opgaven geconstrueerd. Op basis van methoden-analyse zijn relevante spellingcategorieën onderscheiden en per categorie 10 opgaven geconstrueerd. Deze opgaven zijn uitgetoetst bij een landelijke steekproef van scholen. Op grond van psychometrische

analyse en leerkracht-oordelen zijn per categorie vier opgaven voor Woorddictee 3 (medio groep 4) en vier opgaven voor Woorddictee 4 (eind groep 4) geselecteerd. Deze vier woorddictees zijn volgens het afname-design in Figuur 3 voorgelegd aan de leerlingen van het eerder genoemde onderzoekscohort in Rotterdam en Schiedam.



Figuur 3 Afnamedesign woorddictee-toetsen 1a, 1b, 2, 3, 4. (a1 = ankeropgaven uit woorddictee 1, a2 = ankeropgaven uit woorddictee 2, etc.)

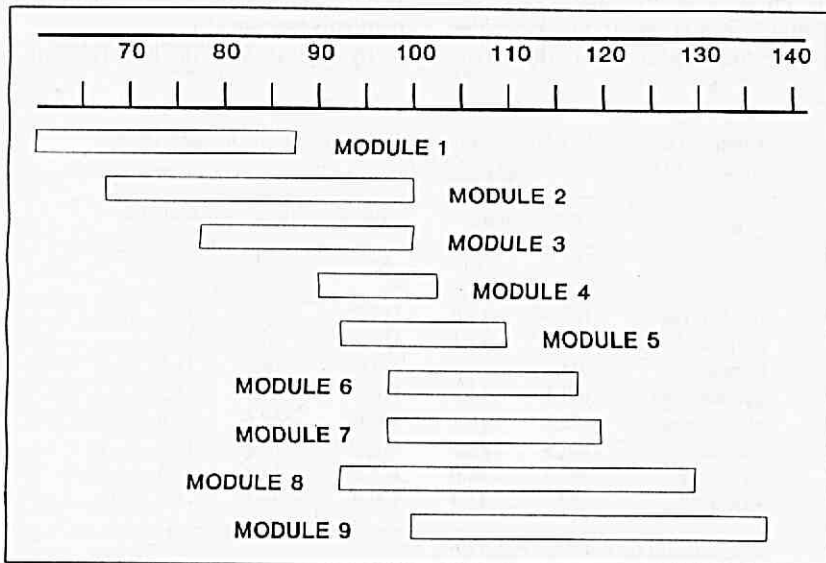
Bij de afname van Woorddictee 1 is dit cohort in drieën gedeeld. Eén groep maakte versie 1a, één groep maakte versie 1b en één groep maakte de helft van versie 1a en de helft van versie 1b. Woorddictee 2, 3 en 4 werd door

alle leerlingen van het cohort gemaakt. Bij deze afnames werden steeds 10 ankeropgaven, opgaven uit de voorgaande afname toegevoegd. Op deze wijze konden bij het calibratieonderzoek alle data met elkaar in verband worden gebracht. Daarbij werden alleen de gegevens betrokken van de leerlingen, die aan alle vier afnames hadden deelgenomen (n = 841). Met behulp van het computerprogramma RIDA (Glas, 1988) werden de itemparameters geschat en het gekozen Raschmodel getoetst. Met gebruikmaking van zogenaamde 'Warm-schatters' (Warm, 1989) werden de vaardigheden van de leerlingen uitgerekend. Op deze wijze ontstond een schaal van 176 spellingopgaven. Met het computerprogramma O.I.S. (1989) werden hieruit negen, in moeilijkheid oplopende modules samengesteld, waarmee min of meer 'op maat' kan worden getoetst. In Figuur 4 is te zien hoe deze modules een deel van de schaal bestrijken.

De modules met een even nummer fungeren op een vast meetmoment als een voortoets, op grond waarvan iedere leerling een eenvoudiger (één nummer lager) of moeilijker (één nummer hoger) module maakt. De ruwe score op de SVS-1 toets wordt bepaald door het aantal correct geschreven woorden. Met behulp van een tabel wordt deze ruwe score

getransformeerd in een schaa score.

Met het FD kan worden nagegaan of kinderen weten welke grafemen gerelateerd zijn aan de 34 fonemen van de Nederlandse taal. De toets wordt groepsgewijs afgenomen, waarbij de leerkracht telkens een foneem dicteert nadat dit eerst als eerste klank in een woordcontext is aangeboden. In het opgavenboekje staat bovendien per item een plaatje afgedrukt dat naar dat woord verwijst. Een voorbeeld: "Kijk naar het plaatje met het *mes*. Schrijf nu achter dat plaatje de *m* van *mes*". De score op de toets wordt bepaald door het aantal correct geschreven fonemen. Met behulp van een diagnoseschema kan bovendien per groep worden nagegaan welke fonemen vaak fout worden geschreven en wat de aard is van die fouten. Bij de TAA moeten leerlingen voorgezegde woorden in fonemen segmenteren. In totaal worden 20 woordpatronen aangeboden, oplopend in moeilijkheid qua orthografische structuur: 2 woorden van het type klinker-medeklinker (mk), 2 km-woorden, 2 mkm-woorden, 2 mmk-woorden, 2 kmm-woorden, 2 mmkm-woorden, 2 mkmm-woorden, 2 mmkmm-woorden, 2 mmmkm-woorden en 2 mkmmm-woorden. Onderzoek van Schreuder en Van Bon (1989) laat zien dat de complexiteit van het segmenteren van woorden in de aangegeven volgorde



Figuur 4 Overzicht van de modules die uit de spellingschaal zijn samengesteld



Tabel 2 *Verdeling van de scholen in de steekproef over de strata 1 (schoolscore  $\leq 1.05$ ), 2 (schoolscore 1.06 - 1.15), 3 (schoolscore  $\geq 1.15$ )*

	autochtoon	allochtoon	onbekend	totaal
stratum 1	185	5	26	216(36%)
stratum 2	194	3	0	197(32%)
stratum 3	87	91	16	194(32%)
	466(77%)	99(16%)	42(7%)	607

toeneemt. De toets kent een individuele afname. De score wordt bepaald door het aantal correct gesegmenteerde opgaven.

#### 4.2 Normering spellingschaal groep 3

In het kader van het landelijke, longitudinale onderzoek zijn in de maand januari en in de periode mei/juni van het schooljaar 1989/1990 enkele modulen uit de SVS-1, het FD én de TAA, afgenomen bij leerlingen uit groep 3 ( $n=607$ ). Deze groep leerlingen vormt een aselechte deelverzameling uit het landelijk cohort. Tabel 2 beschrijft de verdeling van de leerlingen over scholen.

In Tabel 3 zijn de resultaten van de afnames van de SVS-1 in groep 2 samengevat. Daarbij moet worden aangetekend dat volgens een bepaald afname-design de SVS-1 aan meerdere deelpopulaties is voorgelegd, dan aan die waarover wij hier rapporteren. Ten gevolge daarvan is bij deze deelpopulatie niet 'op maat' getoetst (zie 3.3), maar maakten alle leerlingen dezelfde modulen: bij de eerste

afname module 2 en 3, bij de tweede afname module 4 en 5. De meetnauwkeurigheid bleef daardoor lager dan dat bij een toetsing op maat het geval zou zijn.

Alle groepen van leerlingen boeken gedurende het leerjaar een significante vooruitgang in spellingvaardigheid. Tussen de verschillende deelpopulaties bestaan echter al bij de eerste afname aanzienlijke vaardigheidsverschillen; leerlingen op scholen in stratum 1 hebben een significant hogere spellingvaardigheid dan leerlingen op scholen in stratum 3. De mate waarin de deelpopulaties profiteren van het onderwijsaanbod verschilt aanzienlijk. De laatste kolom in Tabel 3 laat zien dat met name allochtone leerlingen en leerlingen op scholen in stratum 3 minder vooruitgaan in spellingvaardigheid dan autochtone leerlingen of leerlingen op scholen in stratum 1 of 2. De vaardigheidsverschillen tussen deze groepen bereiken bij de tweede afname het significantieniveau van 1%.

In Figuur 5 is de ontwikkeling van de

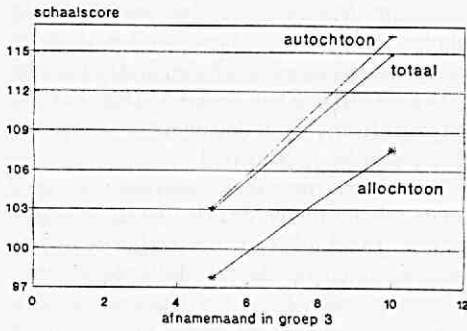
Tabel 3 *Gemiddelde schaalscores, standaarddeviaties en toename in de schaalscore bij de afnames van de SVS-1*

	eerste afname		tweede afname		toename
	X	sd	X	sd	
totale groep	102,9 <sup>b</sup>	15,0	114,9 <sup>a</sup>	13,8	12,0
jongens	101,5	15,1	114,2 <sup>a</sup>	14,3	12,7
meisjes	104,1	14,7	115,6 <sup>a</sup>	13,3	11,5
autochtonen	103,1	14,7	116,3 <sup>a,c</sup>	14,3	13,2
allochtonen	97,8	16,1	107,6 <sup>a</sup>	12,6	9,8
stratum 1	106,3	11,8	118,8 <sup>a,b</sup>	10,7	12,5
stratum 2	102,3	16,0	115,8 <sup>a,b</sup>	14,8	12,5
stratum 3	99,6	16,4	109,0 <sup>a</sup>	13,9	9,7

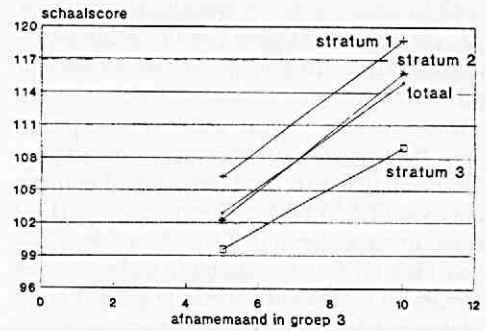
a significant ( $\alpha = 1\%$ ) verschil t.o.v. eerste afname

b significant ( $\alpha = 1\%$ ) verschil t.o.v. stratum 3

c significant ( $\alpha = 1\%$ ) verschil t.o.v. allochtone leerlingen



Figuur 5 De ontwikkeling van de spellingvaardigheid bij respectievelijk de totale populatie, de autochtone en de allochtone leerlingen in groep 3



Figuur 6 De ontwikkeling van de spellingvaardigheid bij de totale populatie en bij leerlingen op scholen in respectievelijk stratum 1, 2 en 3 (groep 3)

Tabel 4 Gemiddelde scores, standaarddeviaties en toename van de scores bij de afnames van het Fonemendictiee

	eerste afname		tweede afname		toename
	X	sd	X	sd	
totale groep	29,2	4,2	31,7 <sup>a</sup>	2,5	2,5
jongens	29,1	4,3	31,6 <sup>a</sup>	2,8	2,5
meisjes	29,4	4,1	31,9 <sup>a</sup>	2,1	2,5
autochtonen	29,4	4,0	31,9 <sup>a</sup>	2,4	2,5
allochtonen	27,3	5,0	30,7 <sup>a</sup>	3,0	3,4
stratum 1	30,5 <sup>b</sup>	2,5	32,3 <sup>a,b</sup>	1,3	1,8
stratum 2	29,4	4,3	31,6 <sup>a</sup>	3,1	2,2
stratum 3	27,6	4,9	31,2 <sup>a</sup>	2,7	3,6

a significant ( $\alpha = 1\%$ ) verschil t.o.v. eerste afname

b significant ( $\alpha = 1\%$ ) verschil t.o.v. stratum 1 en 2

Tabel 5 Gemiddelde scores, standaarddeviaties en toename van de scores bij de afnames van de Toets voor Auditieve analyse

	eerste afname		tweede afname		toename
	X	sd	X	sd	
totale groep	14,9	5,2	18,1 <sup>a</sup>	3,2	3,2
jongens	14,5	5,1	17,9 <sup>a</sup>	3,2	3,4
meisjes	15,2	5,1	18,3 <sup>a</sup>	3,1	3,1
autochtonen	15,0	5,1	18,4 <sup>a</sup>	3,1	3,4
allochtonen	13,6	5,5	16,6 <sup>a</sup>	3,6	3,0
stratum 1	16,1	4,3	18,1 <sup>a</sup>	3,5	2,0
stratum 2	14,8	5,7	19,1 <sup>a,b</sup>	2,4	4,3
stratum 3	13,6	5,1	17,1 <sup>a</sup>	3,1	3,4

a significant ( $\alpha = 1\%$ ) verschil t.o.v. eerste afname

b significant ( $\alpha = 1\%$ ) verschil t.o.v. stratum 3

spellingvaardigheid bij autochtone en allochtone leerlingen grafisch weergegeven. Figuur 6 toont de ontwikkeling van de spellingvaardigheid bij leerlingen op de drie strata van scholen.

### 4.3 Deelcomponenten spellingvaardigheid

De resultaten van de Toets voor Auditieve Analyse (TAA) en het Fonemendictiee (FD) staan in respectievelijk Tabel 4 en 5 beschreven. Ook bij deze toetsen blijken alle groepen leerlingen significant vooruit te gaan. Verder blijken ook hier verschillen in prestaties tussen de deelpopulaties te bestaan. In de laatste kolom van zowel Tabel 4 als 5 is te zien dat deze verschillen in de loop van het leerjaar niet groter worden. Leerlingen van scholen in stratum 3 blijken bijvoorbeeld op beide toetsen meer vooruitgegaan te zijn dan leerlingen van scholen in stratum 1. Desalniettemin halen leerlingen op scholen in stratum 1 bij zowel de

eerste als de tweede afname van het FD een significant hogere score dan leerlingen op scholen in stratum 3. Opvallend is de hoge score van leerlingen op scholen in stratum 2 bij de tweede afname van de TAA. Deze score is zelfs significant hoger dan de score van leerlingen op scholen in stratum 1.

Vervolgens zijn enkele analyses uitgevoerd om de relaties tussen de drie toetsen te onderzoeken. Tabel 6 laat de correlaties tussen de toetsresultaten op de verschillende afname-momenten zien. Over het algemeen zijn de correlaties redelijk tot hoog. De correlaties tussen de drie scores bij de eerste afname van de toetsen zijn hoger dan die tussen de scores bij de tweede afname van de toetsen. Dit is waarschijnlijk het gevolg van een plafondefect op de onderhavige toetsen.

We zijn ook nagegaan of de veronderstelde relatie van de SVS-1 met respectievelijk het FD en de TAA door de resultaten bevestigd

Tabel 6 *Correlaties tussen de verschillende toetsresultaten*

toetsen	2	3	4	5	6
1. TAA (eerste afname)	0,47	0,47	0,34	0,59	0,47
2. TAA (tweede afname)		0,47	0,39	0,45	0,36
3. FD (eerste afname)			0,67	0,74	0,68
4. FD (tweede afname)				0,60	0,55
5. SVS-1 (eerste afname)					0,70
6. SVS-1 (tweede afname)					

Tabel 7 *Stapsgewijze regressie-analyse met de score op SVS-1 (1<sup>e</sup> afname) als afhankelijke variabele en de scores op het FD en de TAA (1<sup>e</sup> afname) als onafhankelijke variabelen*

variabele	b	t	p	multiple R	0,79
FD (1 <sup>e</sup> afname)	2,14	20,1	0,0000	R <sup>2</sup>	0,63
TAA (1 <sup>e</sup> afname)	0,91	10,5	0,0000	standard error	9,17
(constant)		9,8	0,0000	F	465,79
				p	0,00

Tabel 8 *Stapsgewijze regressie-analyse met de score op SVS-1 (tweede afname) als afhankelijke variabele en de scores op de SVS-1 (eerste afname), FD en TAA (eerste en tweede afname) als onafhankelijke variabelen*

variabele	b	t	p	multiple R	0,76
SVS-1 (1 <sup>e</sup> afname)	0,33	6,6	0,0000	R <sup>2</sup>	0,57
FD (1 <sup>e</sup> afname)	1,12	6,6	0,0000	standard error	9,18
FD (2 <sup>e</sup> afname)	0,71	2,6	0,0095	F	148,60
TAA (1 <sup>e</sup> afname)	0,25	2,4	0,0174	p	0,00
(constant)		3,4	0,0006		

wordt. Het ligt in de lijn der verwachting dat leerlingen met een lage spellingvaardigheid de klank-letter-koppeling en/of auditieve analysevaardigheden minder goed zullen beheersen. Om dat te onderzoeken is met een stapsgewijze regressie-analyse het verband geschat tussen de spellingvaardigheid en de overige, gemeten variabelen. In Tabel 7 en 8 staan de resultaten.

Van de variantie in de spellingvaardigheid bij de eerste meting in groep 3 wordt ruim 63% verklaard door de scores die op dat moment op het FD en de TAA zijn gehaald. Van de beide predictorvariabelen levert de score op het FD in relatief opzicht de grootste bijdrage. Ruim 57% van de variantie in de spellingvaardigheid bij de meting aan het einde van leerjaar 3 wordt verklaard door de score bij de eerste en tweede afname van het FD, de score bij de eerste afname van de SVS-1 en de score bij de eerste afname van de TAA. Van de predictorvariabelen leveren de scores bij de eerste afname van de SVS-1 en het FD de relatief grootste bijdrage.

#### 4.4 Diagnostisering en remediëring

De SVS-1 stelt de leerkracht in staat om de vorderingen in spellingvaardigheid zowel in persoonlijke (in vergelijking met zichzelf), sociale (in vergelijking met andere leerlingen) als in didactische zin (in vergelijking met de nagestreefde doelstellingen) te beoordelen (Oud & Mommers, 1988). Doordat de samenstelling van de SVS-modulen mede gebaseerd is op de verschillende categorieën die in de spellingleerstof voor groep 3 en 4 aan de orde komen, kunnen er van de scores signalen uitgaan. In iedere module komen enkele spellingcategorieën aan bod. Per categorie zijn minimaal vier woorden in een module opgenomen. Wanneer een leerling bijvoorbeeld drie van de vier woorden die betrekking hebben op het onderscheid ei-ij fout heeft geschreven, is dat als een signaal te beschouwen. Door die leerling een zogenaamd controledictiee voor te leggen, bestaande uit tien woorden van de signaal-categorie, kan met grotere zekerheid worden vastgesteld of hier inderdaad sprake is van een spellingprobleem. Bij de SVS-1 is voor elk van de 21 spellingcategorieën een controledictiee geconstrueerd. Deze zijn opgenomen in het Hulpboek (Geurts, Van den Bosch & Krom, 1990). Het Hulpboek biedt de leraar verder aanknopingspunten voor de afstem-

ming van het onderwijs op de toetsresultaten. Per spellingcategorie zijn didactische aanwijzingen en oefeningen opgenomen die gebruikt kunnen worden ten behoeve van remediëring. Tevens wordt verwezen naar aanwijzingen en oefeningen in methoden of leergangen die specifiek bedoeld zijn voor leerlingen met spellingproblemen.

## 5 Conclusies en discussie

Het leerlingvolgsysteem dat het Cito ontwikkelt biedt goede mogelijkheden om op basis van schalen die voldoen aan de eisen van de itemresponsstheorie de ontwikkeling van leervorderingen betrouwbaar te volgen. De illustratie van het onderdeel spellen laat zien dat de referentiegegevens en met name de uitsplitsing daarvan naar verschillende subpopulaties, de leerkracht de mogelijkheid biedt om de ontwikkeling van iedere leerling afzonderlijk en die van de totale groep, af te zetten tegen die van een of meerdere relevante referentiegroepen. Met de SVS-1 kan men tot op zekere hoogte ook nagaan of een lage spellingvaardigheid dan wel een stagnerende ontwikkeling daarin, te wijten is aan een onvoldoende beheersing van één of meer specifieke spellingcategorieën. Bij leerlingen die hardnekkige spellingproblemen vertonen, die niet met extra aandacht of oefening verholpen kunnen worden, kan met het FD en de TAA bovendien de beheersing van onderliggende deelvvaardigheden worden vastgesteld. De gehanteerde meetprocedures en technieken maken het mogelijk om in de toekomst meer geavanceerde evaluatietechnieken te ontwikkelen. Zowel wetenschappelijke ontwikkelingen binnen de psychometrie en de cognitieve psychologie als technologische ontwikkelingen op het gebied van personal computers en andere moderne media en hulpmiddelen bieden wat dat betreft mogelijkheden. Bunderson, Inoye en Olsen (1988) schetsen mogelijke ontwikkelingen in computerondersteunde evaluatie, in een raamwerk van vier generaties. Zij voorzien dat uiteindelijk vorderingschalen een centrale plaats in het curriculum gaan innemen en de basis zullen vormen voor een systeem van continue, door de computer gestuurde en beheerde evaluatie, met behulp waarvan de ontwikkeling van leerlingen kan worden gevolgd en het onderwijsleerproces



kan worden vormgegeven. Op een wat minder lange termijn kunnen, met behulp van een te ontwikkelen computerprogramma en schalen, adaptieve toetsingprocedures worden gerealiseerd. Op dit moment wordt bij het Cito-Leerlingvolgsysteem een computerprogramma ontwikkeld, waarmee de toetsresultaten kunnen worden verwerkt en gerapporteerd op de Comenius-computers die de basisscholen tussen 1991 en 1994 zullen ontvangen (Van Bergen, Van Os, Slijkhuis & Zwart, 1990; Moelands, Van den Bosch & Gillijns, 1991). Dit programma zal tevens de mogelijkheid bieden om toetsen interactief af te nemen. Het leerlingvolgsysteem heeft met andere woorden voldoende potentie om de bestaande evaluatiepraktijk in het onderwijs te verbeteren. Toch zijn er ook aanwijzingen (Stiggens e.a., 1986; Janssens, 1986) dat al te veel optimisme daarover misplaatst is. Zo stelde Janssens bijvoorbeeld vast dat slechts 10% van de leerkrachten die een serie leerdoelgerichte rekentoetsen hadden aangeschaft, het materiaal gebruikten overeenkomstig de bedoelingen van de ontwikkelaars. Deze geringe implementatie wijdt hij aan de concurrentie van methode-gebonden toetsen voor korte-termijn-evaluatie, welke mogelijk beter aansluiten bij de onderwijsinhouden en de mogelijkheden en behoeften van leerkrachten (Janssens, 1986, blz. 148). Lange-termijn-evaluatie-instrumenten leveren globale informatie op, welke zich niet ogenblikkelijk en zonder meer laten vertalen in concreet, didactisch handelen. Voor het volgen van de vaardigheidsontwikkeling is een grofmazige screening van relevante leerstof(deel)domeinen voldoende. Voor het aanpassen van het onderwijsleerproces van leerlingen die zich niet voorspoedig ontwikkelen is het onvoldoende. Door aanvullende gegevens te verzamelen, of aanvullende analyses uit te voeren, moet dan de koppeling gemaakt worden tussen de globale informatiestroom van het leerlingvolgsysteem en de meer concrete, of methodegebonden informatiestroom, op grond waarvan de leerkracht zijn dagelijkse didactische handelen vorm geeft. Tot op zekere hoogte wordt in het Cito-Leerlingvolgsysteem gepoogd om de leerkracht hulpmiddelen te geven om die koppeling te leggen. Het SVS-Hulpboek is daar een voorbeeld van. Een gerichte (na- en bij-)scholing en een goede begeleiding zijn in dit verband ook van belang. Een hechtere

integratie van evaluatie, instructie en curriculum tot een omvattend onderwijsleerpakket is een meer adequate oplossing voor dit probleem (Haertel & Calfee, 1983; Baker & Herman, 1983; LeMahieu & Wallace, 1986; Nitko, 1989). Dergelijke ontwikkelingen komen ook in het Nederlandse basisonderwijs op gang. Zo worden in de nieuwste versie het onderwijsleerpakket Veilig Leren Lezen, de materialen van het Cito-leerlingvolgsysteem die betrekking hebben op het aanvankelijk leesproces integraal opgenomen.

### Literatuur

- Aarnoutse, C. A. J. & M. J. C. Mommers, Een vragenlijst als hulpmiddel bij een leerlingvolgsysteem. *Pedagogisch Tijdschrift*, 1987, 12, 330-340.
- Alberts, R. V. J. & H. L. M. Kuijpers, De beoordeling van leerlingresultaten door docenten in het AVO/VWO. In: W. J. Van der Linden (red.), *Moderne methoden van toetsconstructie en -gebruik*. Lisse: 1986.
- Appelhof, P. N., Diagnostiek vanuit innovatieoptiek In: J. J. F. Schroots (red.), *Diagnostiek in het onderwijs*. Lisse: 1986
- Appelhof, P. N., *Preventie functioneel analfabetisme in het Engelse onderwijs. Verslag van een EEG-studiebezoek aan de "Royal County of Berkshire", Engeland, februari 1987*, Utrecht: Schooladviescentrum, 1987.
- Adviesraad voor het Basisonderwijs, Speciaal onderwijs en Voortgezet Speciaal Onderwijs. *Voorrang aan achterstand: advies over een integraal beleid ter voorkoming en bestrijding van onderwijsachterstanden*. Zeist: 1988a
- Adviesraad voor het Basisonderwijs, Speciaal onderwijs en Voortgezet Speciaal Onderwijs. *Regulier en speciaal onderwijs in samenhang. Advies over de ontwikkelingsperspectieven en de groeibeperking van het speciaal en voortgezet speciaal onderwijs*. Zeist: 1988b.
- Baker, E. L. & J. L. Herman, Task structure design: beyond linkage. *Journal of Educational Measurement*, 1983, 20, 149-164.
- Bergen, J. van, E. van Os, H. Slijkhuis & R. Zwart, *Verslag haalbaarheidsonderzoek, projectplan en onderwijskundig ontwerp voor het project Automatisering Rapportage Cito-Leerlingvolgsysteem*. Arnhem: Cito, 1990.
- Brandsma, H. & A. Knuver, Effectieve basisscholen moeten het van sfeer en toetsen hebben. *Didaktief*, 1990, 20 (april), 22-26.
- Bunderson, C. V., D. K. Inoye en J. B. Olsen, *The four generations of computerized educational measurement*. Princeton: Educational Testing Service, 1988.

- Doctor, E. & M. Coltheart, Children's use of phonological encoding when reading for meaning. *Memory and Cognition*, 1980, 8, 195-209.
- Dongen, D. van, *Concrete aanwijzingen voor het invullen van het schoolwerkplangedeelte dat betrekking heeft op "lezen"*. Den Bosch: K.P.C., 1984.
- Drenth, P. J. D., *Inleiding in de testtheorie*, Deventer: 1975.
- Geurts, J. M., L. van den Bosch & R. S. H. Krom, *Hulpboek bij de Schaal Vorderingen in Spellingvaardigheid I*. Arnhem: Cito, 1990.
- Gillijns, P. *Leerlingvolgsystemen*. Onderwijskundige Brochuren Reeks, Tilburg: in voorbereiding.
- Glas, C. A. W., *Handleiding RIDA*. Arnhem: Cito, 1988.
- Glas, C. A. & N. D. Verhelst, *Testing the Rasch-model with incomplete data*, PSM-Report, PR-84-1-1, Utrecht: Rijks Universiteit Utrecht, 1984.
- Haanen, M., R. Roders, P. Lagendaal & K. van der Wolf, *Kleurrijk en effectief; eindrapportage van de eerste fase van het Amsterdamse EGAA-project*. Delft: 1990
- Haertel, E. & R. Calfee, School Achievement: thinking about what to test. *Journal of Educational Measurement*, 1983, 20, 119-132.
- Hathaway, W. E., *Practical advances in goal-based evaluation of school-learning*. Paper presented at the 1975 Annual Meeting of the A.E.R.A., 1975.
- Hathaway, W. E., A school-district-developed, Raschbased approach to minimum competency achievement testing. In: R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing*. McCutchan Publishing, 1981.
- Janssens, F. J. G., Een voorstel tot een evaluatiesysteem voor diagnostiserend onderwijzen. *Nederlands Tijdschrift van Onderzoek van het onderwijs*, 1985, 1, 150-159.
- Janssens, F. J. G., *De evaluatiepraktijken van leerkrachten. Een beschrijvend onderzoek naar het evalueren tijdens het rekenen*. Arnhem: Cito, 1986.
- Janssens, F. J. G. & A. H. J. Moelands, Management van leervorderingen. In: C. Dietvorst, P. G. A. J. Lem, J. Lowyck & S. A. M. Veenman (red.), *Management in de klas*. Culemborg: 1989.
- Jöreskog, K. G. & D. Sörbom, *LISREL V: Analysis of linear structural relationships by maximum likelihood and least square methods*. Chicago: 1981.
- Jungbluth, P., *Verborgene differentiatie. Leerlingenbeeld en onderwijsaanbod op de basisschool*. Nijmegen: I.T.S. 1985.
- Kamphuis, F. H. & A. H. J. Moelands, Longitudinaal meten van individuele leervorderingen met het Cito-Leerlingvolgsysteem. In: P. R. J. Simons & J. G. L. C. Lodewijks (red.), *Onderwijs Research Dagen 1990: Technologie/Methodologie*, Nijmegen: I.T.S., 1990.
- Kingsbury, G. G., Adapting adaptive testing with the microCAT testing system. *Educational Measurement: issues and practice*, 1990, 9, (no 2), 3-6.
- Kool, E., *Handelingsplannen in het buitengewoon onderwijs*. Paper Onderwijsresearch Dagen, 1982.
- Kremers, E. J. J., Evalueren van leerresultaten door leerkrachten in de onderbouw van het voortgezet onderwijs. *Pedagogische Studiën*, 1984, 61, 496-508.
- LeMahieu, P. G. & R. C. Wallace jr., Up against the wall: psychometrics meets praxis. *Educational Measurement: issues and practice*, 1986, 5, (no 1), 12-16.
- Leij, A. van der (red.), *Zorgverbreding: bijdragen uit speciaal onderwijs aan basisonderwijs*. Nijkerk: 1985.
- Leij, A. van der, E. Kool & A. Wielenga, *Vernieuwing van speciaal onderwijs. Een voorbeeld van LOM/MLK-integratie in theorie en praktijk*. Orthovisies 20, Groningen: 1984.
- Lord, F., *Applications of item response theory to practical testing problems*. Hillsdale NJ: 1980.
- Lowyck, J., *Effectiviteit van het onderwijzen*. Losbladig Onderwijskundig Lexicon DHI 2500, 1983.
- Mens, A. J. & C. van Calcar, *Onderwijsstimulering op weg*. Nijmegen: ITS / Amsterdam: SCO, 1981.
- Moelands, A. H. J., L. van den Bosch & P. Gillijns, Systematisch evalueren: Het Cito-Leerlingvolgsysteem. In: J. F. M. Claessen & J. J. van Kuyk (red.), *De grensverleggende basisschool*. Groningen: 1991.
- Moelands, A. H. J., M. J. J. Mommers & J. H. L. Oud, Leerlingvolgsystemen verklaard en vergeleken. *School & Begeleiding*, 1990, 7 (26) 19-28.
- Nitko, A. J., Designing tests that are integrated with instruction. In: R. Linn, (Ed.), *Educational measurement (third edition)*, New York 1989.
- O.I.S. (Optimale Item Selectie), *User's manual*. Arnhem: Cito, 1989.
- Olsen, J. B., A. Cox, C. Price, M. Strozeski & I. Vela, Development of a computerized test for statewide assessment. *Educational Measurement: issues and practice*, 1990, 9 (no 2), 7-10.
- Oud, J. H. L., Leerlingen volgen met LISKAL. In: C. Aarnoutse & M. Voeten (red.), *Gaat en onderwijst; Liber amicorum voor dr. M.J.C. Mommers bij gelegenheid van zijn afscheid van de Katholieke Universiteit Nijmegen*. Tilburg: 1990.
- Oud, J. H. L., J. H. L. van Bercken & J. H. Essers, Longitudinal factor score estimation using the Kalman filter. *Kwantitatieve Methoden*, 1986, 7, 109-129.
- Oud, J. H. L. & M. J. C. Mommers, Longitudinale computerondersteunde onderkenning van lees- en spellingmoeilijkheden: een toepassing van het Kalman-filter in de onderwijspraktijk. *Tijdschrift voor Onderwijsresearch*, 1988, 13, 31-50.

- Scheerens, J., School effectiveness research and the development of process indicators of school functioning. *School effectiveness and school improvement*, 1990, 1, 61-80.
- Schoot, F. van der, Schoolevaluatie in het basisonderwijs. *School en Begeleiding*, 1991, 8, (29), 47-51.
- Schreuder, R. & W. van Bon, Phonemic analysis: Effects of word properties. *Journal of Research in Reading*, 1989, 12, 59-78.
- Stiggings, R. J., N. F. Conklin & N. . Bridgeford, Classroom assessment: a key to effective education. *Educational Measurement: Issues and Practice*, 1986 (summer), 5-17.
- United States Departement of Education, *What works. Research about teaching and learning*. Washington D.C.: 1986.
- Verhelst, N. D. & T. Eggen, *Psychometrische en statistische aspecten van peilingsonderzoek*, PPO-rapport, 4. Arnhem: Cito, 1989.
- Verhoeven, L., *Handleiding woorddictee 1*. Arnhem: Cito, 1980.
- Verhoeven, L., Foutenclassificatie bij beginnende spellers. *Tijdschrift voor taalbeheersing*, 1983, 5, 131-138.
- Verhoeven, L., *Ethnic minority children acquiring literacy*. Dordrecht: 1987.
- Verstralen, H., Rasch scales for extended curricula. *Tijdschrift voor Onderwijsresearch*, 1984, 9, 149-160.
- Voeten, M. J. M., Longitudinaal onderzoek van de leesontwikkeling. In: C. Aarnoutse & M. Voeten (red.), *Gaat en onderwijst; Liber amicorum voor dr. M. J. C. Mommers bij gelegenheid van zijn afscheid van de Katholieke Universiteit Nijmegen*. Tilburg: 1990.
- Warm, A. W., Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 1989, 54, 427-450.
- Wijnstra, J. M., De samenstelling van de schoolbevolking in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 1987, 12, 99-103.
- Wijnstra, J. M. (red.), *Balans van het rekenonderwijs in de basisschool; uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs*; PPO-rapport, nr. 1. Arnhem: Cito, 1988.
- Willis, J. A., Standardized classroom testing in West-Virginia. *Educational Measurement: issues and practice*, 1990, 9, (no 2), 11-14.
- Wolf, J. C. van der, Hoe beoordeelt de leerkracht? De leerkracht beoordeeld. *Tijdschrift voor Orthopedagogiek*, 1986, 25, 390-404.
- Zwarts, M., Toetsanalyse. De voordelen van de itemresponsstheorie. *De psycholoog*, 1990, 25, 514-519.

### Curricula vitae

P. Gillijns (1959) werkte ruim acht jaar in het basisonderwijs en studeerde daarnaast M.O. – Pedagogiek (specialisatie onderwijskunde); eerst in Utrecht en later in Nijmegen. Daarna volgde hij de doctoraal-opleiding Interdisciplinaire Onderwijskunde aan de universiteit van Nijmegen, waar hij in 1989 cum laude afstudeerde. Sinds twee jaar is hij als onderwijskundige werkzaam op de afdeling Basis- en Speciaal Onderwijs van het Instituut voor Toetsontwikkeling (Cito) in Arnhem als medewerker bij het project Leerlingvolgsysteem.

Adres: Cito, Instituut voor Toetsontwikkeling, Postbus 1034, 6801 MG Arnhem

L. Verhoeven studeerde ontwikkelingspsychologie (doctoraal 1978) en orthopedagogiek (doctoraal 1979) aan de Katholieke Universiteit Nijmegen. Hij promoveerde in 1987 aan de Katholieke Universiteit Brabant op het proefschrift getiteld 'Ethnic minority children acquiring literacy'. Van 1978-1981 was hij als medewerker, en in de periode erna als adviseur, binnen het Instituut voor Toetsontwikkeling betrokken bij de ontwikkeling van evaluatie-instrumenten voor het basisonderwijs. Vanaf 1982 is hij werkzaam als universitair (hoofd)docent aan de Letterenfaculteit van de KUB. Zijn onderzoeksbelangstelling richt zich met name op individuele verschillen in mondelinge en schriftelijke moedertaal- en tweedetaalverwerving.

Adres: Letterenfaculteit Katholieke Universiteit Brabant, Postbus 90153, 5000 LE Tilburg

Manuscript aanvaard 7-5-'91

### Summary

Gillijns, P. & L. Verhoeven. 'Towards a longitudinal assessment system at primary school.' *Pedagogische Studiën*, 1991, 68, 216-230.

In the present article the outline for a longitudinal assessment system at primary school is presented. After a discussion on the feasibility of systems monitoring the educational outcomes of children during the course of primary school the basic characteristics of such systems are highlighted. Arguments are given for item response theory as the optimal measurement procedure. In addition, a longitudinal assessment system, designed at the Dutch Institute for Educational Measurement, is introduced. This system aims at monitoring children's learning results in oral and written language, maths and social sciences in an adaptive way. An illustration is given for the unit of spelling. Finally, the use of the proposed assessment system in educational practice is discussed.