

Het Cito-Leerling- volgsysteem

Spelen met oogkleppen op*

P. Vedder, Rijksuniversiteit Leiden

Samenvatting

In het meinummer 1991 van dit tijdschrift presenteerden Gillijns en Verhoeven het Cito-Leerlingvolgsysteem. Dit systeem bestaat uit een reeks gestandaardiseerde unidimensionele tests voor lezen, spellen, rekenen en andere leergebieden. In deze discussiebijdrage worden verschillende problemen beschreven, waaronder de volgende:

- Gillijns en Verhoeven gaan nauwelijks in op de inhoudvaliditeit van hun volgsysteem;
- Ze beschrijven de unidimensionaliteit van hun tests als een positieve kwaliteit en gaan er daarbij aan voorbij dat de meeste schoolse vaardigheden eerder multidimensioneel dan unidimensioneel zijn;
- De informatie die met het leerlingvolgsysteem wordt verzameld is waarschijnlijk nauwelijks bruikbaar voor leerlingen en ouders;
- De suggestie dat met de tests leer-vorderingen kunnen worden gemeten via absolute vergelijking is onterecht.

In het meinummer van Pedagogische Studiën van vorig jaar stond een artikel van Gillijns en Verhoeven getiteld "Naar een leerlingvolgsysteem voor het basisonderwijs". Dit artikel gaat over het ontwerpen van een leerlingvolgsysteem voor de onderwijspraktijk. Ik las dit artikel indertijd vooral vanwege mijn interesse voor problemen rond het meten van de kwaliteit van het onderwijs en stelde teleurgesteld vast dat het artikel daaraan weinig aandacht schonk. Bij wijze van commentaar op het artikel van Gillijns en Verhoeven beschrijf ik

hierna een aantal problemen rond het meten van de kwaliteit van het onderwijs.

De kernboodschap van Gillijns en Verhoeven (1991) is, dat het met de praktijk van het evalueren in het onderwijs niet goed is gesteld, terwijl evalueren van groot belang is voor de kwaliteit van het onderwijs. Ze stellen verder dat het nieuw te ontwikkelen Cito-leerlingvolgsysteem een bijdrage kan zijn aan het verbeteren van de heersende evaluatiepraktijk. Binnen deze evaluatiepraktijk zijn zo'n 50 andere volgsystemen in gebruik of in de maak, maar het gros is psychometrisch en didactisch gezien onder de maat. Naast deze boodschap wordt de kracht van het eigen volgsysteem op een aantal punten verduidelijkt. De spanning tussen het twee keer per jaar registreren van leervorderingen en het belang van het regelmatig en intensief diagnostiseren van leerproblemen wordt beschreven. Er wordt een aanzet gegeven tot vermindering van de spanning door aanvullend op het volgsysteem diagnostische toetsen te ontwikkelen. Er wordt uitgebreid aandacht besteed aan de psychometrische kwaliteit van reeds beschikbare onderdelen van het systeem. Het gaat om een reeks gestandaardiseerde, genormeerde toetsen. Het probleem van de gebrekkige aansluiting van de informatie verkregen met gestandaardiseerde toetsen en de didactische praktijk op school wordt aangestipt. Er wordt gewezen op de concurrentie van methode-gebonden toetsen die beter aansluiten bij een specifiek curriculum dat op school wordt gebruikt en bij de mogelijkheden en behoeften van de leerkrachten. Gillijns en Verhoeven stellen terecht dat er een hechte integratie van evaluatie, instructie en curriculum moet worden nagestreefd, om de kwaliteit van het onderwijs te verbeteren. Daaraan wordt deels een bijdrage geleverd via het Cito-leerlingvolgsysteem. Onderdelen van dit volgsysteem met betrekking tot het aanvankelijk lezen zijn opgenomen in de nieuwste versie van het onderwijsleerpakket "Veilig leren lezen".

Uiteraard is het interessant dat Gillijns en Verhoeven problemen en oplossingen beschrijven, maar ik word hier en daar toch wat kriegel van al die problemen die niet tot volle bloei kunnen komen omdat ze worden bedekt met een sneeuw van oplossingen. Zonder iets af te willen doen aan de waarde van het initiatief

* Met dank aan L. van der Kamp voor zijn commentaar op een eerdere versie.

tot integratie van delen van het volgsysteem in een leesmethode, is het bijvoorbeeld voor toekomstige pogingen om tot een dergelijke integratie te komen van belang even stil te staan bij een specifieke omstandigheid waaronder het tot deze integratie kwam. De methode "Veilig leren lezen" wordt gebruikt op meer dan 80% van alle basisscholen. Hierdoor doet zich nauwelijks een probleem voor wat betreft de aansluiting van de inhoud van de toetsen en de inhoud van het onderwijsprogramma. Voor elke andere poging zullen zich denkkelijk meer aansluitingsproblemen voordoen.

Ik ga hierna in op een aantal andere problemen rond meetpraktijken in het onderwijs. Eerst richt ik me daarbij op inhoudelijke aspecten van evaluatiepraktijken, hun gebruik en de belanghebbenden, daarna ga ik in op problemen rond veranderingsmetingen.

Wat te meten en voor wie

Gillijns en Verhoeven omschrijven de functie van een leerlingvolgsysteem in aansluiting bij de ARBO als het signaleren van het achterblijven van de vorderingen van leerlingen ten opzichte van door de school nagestreefde doelen. Enkele malen per jaar worden de leervorderingen geregistreerd. Zoals de auteurs ook aangeven staat het registreren van de leervorderingen in functie van het bewaken en zo mogelijk verbeteren van de kwaliteit van het onderwijs. De vraag die dit oproept is: welke kwaliteit en wiens kwaliteit? Broadfoot (1992) wijst op het gevaar dat de kwaliteit van het onderwijs wordt gelijkgesteld aan leerprestaties die met gestandaardiseerde toetsen worden gemeten. Zij merkt op dat een eenzijdig gebruik van dergelijke toetsen om de kwaliteit van het onderwijs vast te stellen uiteindelijk bijdraagt tot een verschraling van de onderwijsinhoud. Tot dusver worden in het onderwijs vooral toetsen gebruikt met betrekking tot relatief eenvoudig te meten vaardigheden en kennis. Goede toetsen waarmee de invloed van onderwijs op complexe algemene en vakspecifieke probleemoplossingsvaardigheden, motivatie, coöperatieve vaardigheden, kritische zin en dergelijke worden gemeten zijn nog niet voldoende voorhanden (vgl. Clifford, 1990; Raven, 1981).

Daarnaast zijn er ook onderwijsdoelstellingen die niet meetbaar zijn, terwijl ze wel belangrijke aspecten van de kwaliteit van het onderwijs betreffen (vgl. Hofstee, 1992).

Meer direct gericht op de beschrijving van het volgsysteem door Gillijns en Verhoeven kan ook opgemerkt worden dat ze weinig aandacht schenken aan de inhoudsvaliditeit van het systeem, en de daarin gebruikte toetsen. Opdat informatie verkregen met toetsen betekenisvol is voor de verbetering van het onderwijs is het noodzakelijk dat de items zijn gekozen op basis van een curriculumanalyse waarbij is nagegaan wat belangrijke elementen in het curriculum zijn. Voor de bepaling van de inhoudsvaliditeit van de toetsen dient de betekenis van de gerepresenteerde curriculaire eenheden te worden onderzocht (zie Van Tilborg, 1987; Plomp & Loxley, 1992).

Mommers (1988) gaf indertijd in zijn commentaar bij de plannen van PPO aan dat er bij testconstructie soms afbreuk wordt gedaan aan de inhoudsvaliditeit. Hij geeft het voorbeeld van items met een lage p-waarde. Dergelijke items worden doorgaans uitgesloten omdat ze de betrouwbaarheid van een test drukken of instabiel maken. De reden voor de lage p-waarde kan echter slecht onderwijs zijn voor het betreffende domein. Uitsluiting van dergelijke items betekent dan dat daarover geen informatie wordt verzameld en dat het slechte onderwijs ongemerkt kan worden voortgezet. Mommers refereert bij wijze van voorbeeld aan instructie in het basisonderwijs over het metrieke stelsel dat nogal eens onbevredigende resultaten oplevert, en dat gezien de maatschappelijke relevantie eigenlijk beter verdient.

Daar komt nog bij dat de bepaling van de inhoudsvaliditeit geen eenmalige zaak is. Toetsontwikkelaars zouden eigenlijk moeten aangeven hoe ze de validiteit op peil houden. Immers, onderwijsinhouden of de mate waarin leerlingen in aanraking komen met inhouden veranderen over tijd. Overigens wordt in het volgsysteem gebruik gemaakt van unidimensionele toetsen. Het werken met unidimensionele toetsen in een onderwijscontext is een omstreden zaak. Van der Kamp (1992) merkt op dat het empirisch weliswaar mogelijk blijkt voor een veelheid van schoolse vaardigheden unidimensionele schalen te construeren, maar

dat de relevantie van die schalen vermoedelijk in vele gevallen gering is. Reeksen opgaven voor bijvoorbeeld begrijpend lezen of rekenen die in onderwijsprogramma's aan de orde komen hebben eerder betrekking op een complex van verschillende vaardigheden dan op een specifieke vaardigheid. Kortom, ze zijn eerder multidimensioneel dan unidimensioneel (vgl. Goldstein & Wood, 1989; Shavelson & Baxter, 1992).

Gillijns en Verhoeven schenken aan deze inhoudelijke problemen en de bijbehorende discussies geen aandacht. De gedachte, zoals vertolkt door Gillijns en Verhoeven, dat een leerlingvolgsysteem vanwege het systematiseren van de evaluatiepraktijk op school zal bijdragen aan de verbetering van de kwaliteit van het onderwijs is veel te simpel.

Broadfoot (1992) gaat ook in op de vraag voor wie de kwaliteit van het onderwijs gemeten wordt. Zij stelt dat de kwaliteit van het onderwijs geen zaak is van slechts de leerkrachten, schoolleiders en schoolbesturen, maar vooral van ouders en leerlingen. Wat betreft de wensen van ouders over evaluatiepraktijken op school in Nederland is weinig bekend. Broadfoot beschrijft dat voor Engelse ouders geldt dat ze regelmatig geïnformeerd willen worden over leerprestaties, zowel schriftelijk als via persoonlijk contact met de leerkrachten. Daarbij willen ze zowel iets vernemen over de zwakke als de sterke kanten van de leerlingen. Problemen moeten vroegtijdig worden vastgesteld en over die problemen moet ook vroegtijdig met de ouders worden gesproken. Ze willen naast informatie over leerprestaties ook informatie over leerattitudes, over de sociale en emotionele ontwikkeling, het gedrag en de sociale redzaamheid ten aanzien van alledaagse zaken. Ze willen bovendien dat de informatie die via evaluatie is verkregen wordt gewaardeerd in het licht van toekomstige leerdoelen. Verder geven ouders te kennen dat ze het van belang achten dat de kinderen bij de beoordeling betrokken zijn en dat ze de waarde van de gehanteerde evaluatiecriteria kunnen begrijpen in relatie tot het onderwijsprogramma en hun eigen leren. In relatie tot een studie van Crooks (1989) merkt Broadfoot op dat dit laatste stellig van belang is voor leerlingen, maar

ook voor leerkrachten zelf. Ook leerkrachten moeten de evaluatie kunnen betrekken op het curriculum en het leren van de leerling, want daarmee wordt een duidelijke basis gelegd voor aansluitend didactisch handelen.

Indien de beschrijving van de functies van het volgsysteem zoals gepresenteerd door Gillijns en Verhoeven wordt vergeleken met de hier beschreven verlangens van ouders, dan lijkt het gebruik van hun volgsysteem geen echt gewenste bijdrage te kunnen leveren. Ook op de brede vraag voor wie de kwaliteit van het onderwijs moet worden gemeten gaan Gillijns en Verhoeven niet in.

Veranderingsmetingen in het onderwijs

In discussies over de kwaliteit van het onderwijs is juist het meten van veranderingen van groot belang. Nog preciezer uitgedrukt gaat het om het meten van de veranderingseffecten van onderwijs. Het is weinig zinvol om over de kwaliteit van het onderwijs te praten in relatie tot slechts eenmalig gemeten prestaties van leerlingen of groepen leerlingen. In dergelijke discussies treedt onvermijdelijk een verwarring op van de kwaliteit van het onderwijs en de ontwikkelingsstand van leerlingen. Wat wordt beoordeeld als goed onderwijs kan even zo goed onderwijs zijn waaraan wordt deelgenomen door kinderen die betrekkelijk onafhankelijk van het onderwijs goed presteren. Leerlingen komen op school om te leren, en leren is veranderen. De veranderingen die worden geschapen en ondersteund door de activiteiten en mogelijkheden die door de school worden geboden dienen het belangrijkste studieobject te zijn van personen die de kwaliteit van het onderwijs willen bewaken of verbeteren (vgl. Nuttall, 1986, p. 157; Willett, 1988/89, p. 346). Niet elke benadering van veranderingsmeting kan evenwel als kansrijk worden aangemerkt (vgl. Van der Kamp, 1992).

Gillijns en Verhoeven presenteren hun volgsysteem als een instrumentarium waarmee groei van individuele leerlingen en groepen leerlingen, absolute groei, kan worden gemeten. Hun instrumentarium bestaat uit een reeks Rasch homogene schalen die oplopen in moei-

lijkheidsgraad. De vergelijkbaarheid wordt ge-
waarborgd via overlappende items, zoge-
naamde ankeritems. Voor zover ik het begrip
vormt ook de verzameling van alle gebruikte
items van alle spellingsschalen een Rasch-
homogene, unidimensionale schaal. Indien dit
niet het geval zou zijn zou dit aanleiding zijn tot
vergelijkbaarheidsproblemen, aangezien de
verschillende schalen dan niet dezelfde vaar-
digheid zouden meten. Alhoewel Gillijns en
Verhoeven daarover geen informatie geven is
het goed voorstelbaar dat het voor vaardig-
heidsdomeinen als begrijpend lezen, stellen,
informatieverwerking, wereldoriëntatie en re-
kenen moeilijker is om relevante, unidimensio-
nele schalen te ontwikkelen dan voor het vaar-
digheidsdomein dat in het artikel uitvoerig
wordt beschreven, nl. spellen. Overigens ma-
ken Goldstein en Wood (1989) duidelijk dat
zelfs de unidimensionaliteit in termen van de
itemrespons theorie van spellingstoetsen be-
trekkelijk is. Voor een gegeven populatie kan
een dergelijke test unidimensioneel blijken,
terwijl voor een specifieke steekproef uit de po-
pulatie, waarin bijvoorbeeld dyslectische kin-
deren zitten, de test multidimensioneel is (zie
ook Goldstein, 1983).

Om veranderingsmetingen te kunnen interpre-
teren in termen van groei moet de dimensiona-
liteit van de te gebruiken toets stabiel zijn over
de tijd bezien. Deze vereiste levert problemen
op. Die kunnen worden gezien vanuit de toets
of vanuit het in de toets gerepresenteerde item-
domein. Beide zijn uiteraard gerelateerd, maar
voor de duidelijkheid is het handig ze te onder-
scheidend.

Dimensionaliteit is een empirisch vast te
stellen kwaliteit van een verzameling items in
een test. In onderzoek wordt keer op keer dui-
delijk dat de dimensionaliteit van een test tijd-,
context- en situatiespecifiek is (vgl. Goldstein,
1983; Mellenbergh, 1972). Stabiliteit van de
unidimensionaliteit over tijd betekent dat er
technisch gesproken sprake is van invariantie
van itemparameters over verschillende meet-
momenten. Meer beeldend uitgedrukt: de items
moeten in relatie tot de verzameling waar ze
deel van uitmaken dezelfde zeggingskracht, of
betekenis hebben. Ze mogen niet aan relevantie
inboeten. Of dat werkelijk het geval is bij het
gebruik van een test is keer op keer afhankelijk

van de leerlingen die de test maken. Testge-
bruikers kunnen dat uiteraard negeren en er van
uitgaan dat het woord van de testconstructeurs
telt die vanwege de gevolgde constructie- en
analyseprocedures uitspraken doen in termen
van unidimensionaliteit van toetsen en moge-
lijkheden tot absolute meting van leervorderin-
gen. Volgens Nuttall (1986) zou dat onverstan-
dig zijn aangezien de aanname van de stabiliteit
van de unidimensionaliteit in het onderwijs niet
of nauwelijks is waar te maken. Hij schrijft:
"The essence of the process of education is to
achieve change in behavior, change that is (in
practice if not in theory) not identical in each
individual" (p. 157).

Vanuit het itemdomein bekeken is de eis
van de temporele stabiliteit van de dimensiona-
liteit te omschrijven als de eis van onverander-
lijkheid van het in de toetsen gerepresenteerde
itemdomein. Indien het aangeboden curricu-
lum wordt gebruikt als basis voor de itemkeuze
dan is de aanname van onveranderlijkheid van
het itemdomein nauwelijks gegarandeerd. In
de eerste plaats varieert het aangeboden curri-
culum afhankelijk van groepen leerlingen. In
de tweede plaats is het curriculum aan verande-
ring onderhevig vanwege verschuiving in doel-
stellingen. Verandering van itemdomein met
gelijkblijvende toetsen ondermijnt de waarde
van de toetsscores. McLean (1990) wijst op de
risico's van speciale testprogramma's waarin
gestandaardiseerde tests worden gebruikt.
Over het algemeen geldt volgens hem "educa-
tion is slow to change, but testing is slower"
(p. 11). Dergelijke tests boezemen evenwel
ontzag in, ze worden gezien als een symbool
voor de kwaliteitsgarantie van scholen die ze
gebruiken. Door dergelijke tests als richtsnoer
te gebruiken van een kwaliteitsstandaard,
wordt echter al snel afbreuk gedaan aan de
kwaliteit, omdat de toetsen niet meer passen bij
wenselijke inhoudelijke vernieuwingen.

De conclusie is betrekkelijk simpel, en bepaald
niet door mij bedacht. Veranderingen in de in-
houd van items over tijd en in de relevantie van
items maakt absolute vergelijking over tijd om
leerprogressie vast te stellen onmogelijk. Het
lijkt nodig deze oude boodschap af en toe een
nieuwe stem te geven. Deze conclusie, overig-
ens, doet niets af aan het belang van verande-

ringsmeting dat aan het begin van deze paragraaf is verduidelijkt.

Discussie

In mijn commentaar heb ik samengevat de volgende overwegingen naar voren gebracht:

- Nadruk op het gebruik van gestandaardiseerde toetsen kan een inhoudelijke verschraling van het onderwijs tot gevolg hebben.
- Voor de inhoudsvaliditeit van het leerlingvolgsysteem zou dit moeten aansluiten bij onderzoek naar belangrijke elementen in het curriculum. Gezien de veranderlijkheid van het curriculum dient dit herhaaldelijk te gebeuren, opdat de inhoudsvaliditeit op peil kan worden gehouden.
- Veel vaardigheden die in het onderwijs worden geleerd zijn multidimensioneel van aard. De relevantie van unidimensionele schalen is bij voorbaat twijfelachtig in het licht van deze constatering.
- Bij gebruik van het Cito-leerlingvolgsysteem wordt vermoedelijk weinig rekening gehouden met wensen van ouders en kinderen wat betreft informatie over de voortgang van het leren en over mogelijkheden om die voortgang positief te beïnvloeden.
- Om veranderingsmetingen zoals beoogd met het leerlingvolgsysteem te kunnen interpreteren in termen van groei, moet de dimensionaliteit van de toetsen stabiel zijn en tegelijk moeten de toetsen relevant blijven ten opzichte van het curriculum. Deze twee vereisten leveren zoveel problemen op dat absolute vergelijking over tijd, bedoeld om de groei vast te stellen, in de praktijk onmogelijk is.

Al deze punten bijeengenomen lijkt er geen andere conclusie mogelijk dan dat de waarde van het Cito-leerlingvolgsysteem gering is. Dit oordeel kan echter op punten nog worden genuanceerd.

De vraag of het volgsysteem inhoudelijk voldoende relevantie bezit afgemeten aan wenselijke doelstellingen van het onderwijs, kan wellicht het best worden gezien vanuit de gedachte dat onderwijs een vrije ondernemerssector is.

Dit beeld past goed bij de verdergaande automatisering van onderwijsinstellingen t.o.v. de landelijke overheid. Scholen zullen zich vermoedelijk in de toekomst nadrukkelijker kunnen profileren wat betreft de opvoedingswaarden en onderwijsinhouden die zij belangrijk achten. De keerzijde hiervan is dat verzorgingsinstellingen steeds minder een homogene klantenkring hebben. Niet langer zal gelden dat verzorgingsinstellingen kunnen volstaan met de ontwikkeling van een standaardpakket aan voorzieningen voor alle scholen.

Dit zijn evenwel ontwikkelingen waarvan de aard en inhoud nog niet duidelijk is. Hiervoor is wellicht al te zeer gesuggereerd alsof er nu al een scherp oordeel over de inhoudelijke relevantie van het volgsysteem kan worden gegeven. Vanuit het beeld van onderwijs als een vrije ondernemingssector kan de relevantievraag weliswaar via onderzoek worden verduidelijkt, maar niet opgelost. Onderwijsdoelstellingen zijn het domein van gewoonten en op verandering gerichte machten rond en in het onderwijs. Aan de mate waarin de ontwikkelde toetsen op een gegeven moment zullen worden gebruikt, kan de dan geldende waarde van de daarin gerepresenteerde onderwijsdoelen worden afgemeten. Kortom, in plaats van het signaleren van een probleem hadden we in dit geval ook het voordeel van de twijfel kunnen laten gelden. Wat wel kan worden vastgesteld is dat de auteurs zelf weinig hebben geschreven dat de inhoudelijke relevantie van het volgsysteem verduidelijkt.

Bij de vaststelling dat belangen van ouders en leerlingen aandacht verdienen gaat het uiteraard ook om een macht rond en in het onderwijs. Via de keuze van een school en via actieve participatie of zelfs weigering daartoe maken ouders en kinderen duidelijk dat ze bepaalde waarden ondersteunen dan wel afwijzen (zie Raven, 1981). Deze macht van ouders en kinderen heeft belang bij een evaluatiepraktijk. Scholen kunnen zich niet permitteren zich op het standpunt te stellen dat wat dit betreft de macht van ouders en leerlingen nog maar eens moet blijken. Vanuit de gedachte van onderwijs als een vrije ondernemingssector is zonder verder afwachten of nader onderzoek nu al duidelijk dat ouders en leerlingen intensief betrokken moeten worden bij het onderwijsgebeuren.

Scholen moeten aan klantenbinding doen. Ouders en leerlingen moeten niet slechts vanuit onderwijskundige en pedagogische redenen sterker betrokken worden bij het onderwijsgebeuren, maar ook om voldoende klanten te houden en nieuwe te werven. Om dit te realiseren dienen scholen in hun evaluatiepraktijk rekening te houden met wensen van ouders en leerlingen. Het lijkt er op dat het leerlingvolgsysteem niet goed aansluit bij de informatiebehoefte van ouders en leerlingen. Als het volgsysteem desalniettemin gebruikt wordt lijkt het dienstig om het in te kaderen in een bredere evaluatiecontext waarin nadrukkelijk wel aandacht bestaat voor de informatiewensen van ouders en leerlingen.

Blijft wat de inhoudelijke discussie betreft nog over het probleem van het gebruik van unidimensionele toetsen. Deze toetsen zijn niet dekkend te maken voor de sterke tendens in het Nederlandse onderwijs om aandacht te besteden aan complexe probleemoplossingsvaardigheden. Daarvoor moeten nieuwe toetsen en toetsprocedures ontwikkeld worden. Interessante aanzetten daartoe worden onder meer gegeven door Snow en Lohman (1989) die schrijven over de betekenis van de cognitieve psychologie voor onderwijskundige meetpraktijken.

Het geschetste probleem met betrekking tot het meten van leervorderingen via absolute vergelijking is niet op te lossen. Absolute vergelijking in strikte zin is onmogelijk. De oplossing voor het volgsysteem ligt ook niet in het eventuele gebruik voor relatieve vergelijking, waarbij de verandering in relatieve positie van vergelijkingsgroepen over tijd wordt nagegaan. Dergelijke veranderingsmetingen zijn mogelijk, maar daarbij dienen de te gebruiken toetsen per meetmoment opnieuw gestandaardiseerd te worden voor de populatie. Met een bestaand volgsysteem dat gebruikt wordt door scholen is een dergelijke exercitie ondenkbaar, of er zou via een nationaal toetsservicesysteem gewerkt moeten worden, dat dataoverwerkingsmogelijkheden en ijkingsmogelijkheden biedt aan scholen.

Er is echter toch nog een sluiproute. We kunnen met zijn allen, of wellicht is dat slechts voor sommigen weggelegd, besluiten dat het bij ver-

anderingsmeting gaat om een moeilijk spel, met spelregels. Soms zijn de omstandigheden er niet naar om alle spelregels steeds op dezelfde rigide wijze toe te passen. Spelen is tenslotte fijn en als we kunnen voorkomen dat we het spel moeten stoppen dan zouden we al te saai worden als we dat niet deden. Geheel in de geest van deze gedachte kan een toetsontwikkelaar en een toetsuitgever zich afvragen hoe lang een ontwikkelde toets redelijkerwijs gebruikt kan worden zonder dat items en de test aan relevantie inboeten ten opzichte van het gerepresenteerde leerstofdomein. Zoals Goldstein (1983) opmerkt "It is a problem, at least in theory, which confronts every test publisher" (p. 374). Zolang de overtuiging bestaat dat de relevantie nog prima is kan dat een reden zijn de stabiliteit van de dimensionaliteit en itemparameters niet te onderzoeken. Het is spelen met oogkleppen op, maar dat is soms goedkoper en vooral prettiger dan helemaal niet spelen.

Literatuur

- Broadfoot, P. (1992). Measuring the quality of education for parents and students. In P. Vedder (Ed.) *Measuring the quality of education*. Amsterdam: Swets & Zeitlinger.
- Clifford, M. (1990). Student motivation and academic achievement. In P. Vedder (Ed.) *Fundamental studies in educational research*. Amsterdam: Swets & Zeitlinger.
- Crooks, T. (1989). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Gillijns, P., & Verhoeven, L. (1991). Naar een leerlingvolgsysteem in het basisonderwijs. *Pedagogische Studiën*, 68, 216-230.
- Goldstein, H. (1983). Measuring changes in educational attainment over time: problems and possibilities. *Journal of Educational Measurement*, 20, 369-377.
- Goldstein, H. & Wood, R. (1989). Five decades of item-response modelling. *British Journal of Mathematical and Statistical Psychology*, 42, 139-167.
- Hofstee, W. K. B. (1992). Independent educational assessment in the interest of educational quality. In P. Vedder (Ed.) *Measuring the quality of education*. Amsterdam: Swets & Zeitlinger.

- Kamp, L.J.Th. van der (1992). Some issues and problems in measuring the quality of education. In P. Vedder (Ed.) *Measuring the quality of education*. Amsterdam: Swets & Zeitlinger.
- McLean, L. (1990). Let's call a halt to pointless testing. *Education Canada*, 30(3), 10-14.
- Mellenbergh, G. J. (1972). Applicability of the Rasch model in two cultures. In L.J. Cronbach & P. Drenth (Eds.) *Mental tests and cultural adaptation*. The Hague: Mouton.
- Mommers, M. J. C. (1988). Beoordeling "concept meerjarenplan 1988-1991 en werkplan 1988" inzake PPO. *Ongepubliceerd verslag*, K.U. Nijmegen.
- Nuttall, D. L. (1986). Problems in the measurement of change. In D. L. Nuttall (Ed.) *Assessing educational achievement*. London: Falmer Press.
- Plomp, T., & Loxley, W. (1992). IEA and the quality of education in developing countries. In P. Vedder (Ed.) *Measuring the quality of education*. Amsterdam: Swets & Zeitlinger.
- Raven, J. (1981). The most important problem in education is to come to terms with values. *Oxford Review of Education*, 7, 253-272.
- Shavelson, R., & Baxter, G. (1992). On measuring the quality of mathematics and science education: achievement indicators. In P. Vedder (Ed.) *Measuring the quality of education*. Amsterdam: Swets & Zeitlinger.
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.) *Educational measurement* (3rd ed.). New York: Macmillan.
- Tilborg, I. A. J. van (1987). *De betekenis van het arbeidersgezin voor het leerniveau en de schoolloopbaan van het kind*. Den Haag: SVO.
- Willett, J. (1988/89). Questions and answers in the measurement of change. *Review of Research in Education*, 15, 345-422.

Manuscript aanvaard 18-5-1992

Auteur

P. Vedder werkt bij de sectie interculturele pedagogiek van de Rijksuniversiteit Leiden.

Adres: Wassenaarseweg 52, Postbus 9555, 2300 RB Leiden

Abstract

The Cito monitoring system; playing blinked

P. Vedder. *Pedagogische Studiën*, 1992, 69, 284-290.

In number 5 (volume 68, 1991) of this journal Gillijns and Verhoeven presented a student monitoring system developed by the Dutch educational testing service, Cito. The monitoring system contains a series of standardized unidimensional tests for reading, spelling, arithmetics and other subject domains.

This comment addresses amongst others the following problems: a) The developers hardly address the content validity of the system; b) They describe the unidimensionality of the tests as a positive quality neglecting that many skills that are learned in schools are likely to be multidimensional in nature; c) Use of the system will hardly result in information that may be of help to parents and students; d) The suggestion that the tests may be used for measuring progress in learning through absolute comparison over time is an erroneous exaggeration. The tests cannot be used for this purpose.