

OVER STANDAARDEN EN TOETSEN

Kritische stemmen uit het buitenland

Ontwikkelingen in het onderwijs die zich hier voordoen, vinden eerder vaak elders plaats. Daar zouden we van kunnen profiteren, ervan leren, maar dat is echter vaak niet het geval, helaas. In dit artikel worden inzichtjes gegeven in de kritiek op de invoering van leerstandaarden en de daarmee verbonden centrale toetsing elders. Niet dat dit bij ons al het geval is, maar er zijn sterke krachten die dit ook hier willen invoeren.

Een uitgebreide versie van deze tekst is opgenomen in de brochure met discussie- en werkmateriaal van de actie 'Streep door de eindtoets'. Een gewaarschuwd mens telt voor twee. En bovendien is het goed om eens in deze 'spiegel' te kijken.

Echo van elders

Het onderwijsbeleid in ons land en de discussie daar omheen zijn vaak voor een deel een echo van wat ook elders speelt. Er zijn internationale trends zichtbaar. De tendens van het formuleren van 'leerstandaarden' en het invoeren van een verplicht systeem van toetsen dat daarmee verbonden is, is zo'n internationale trend.

Merkwaardig genoeg hoor en lees je bij ons heel weinig over kritiek op trends in landen die daarmee al veel langer bezig zijn, zoals de Verenigde Staten en het Verenigd Koninkrijk. Zo hebben, om bij het laatste te blijven, in 2001 Wales en Noord-Ierland het publiceren van toetsresultaten in de krant afgeschaft, vanwege de negatieve bijwerkingen die dit heeft. We hebben hierover in Nederland nergens iets gehoord en gelezen.

In dit artikel laten we kritische stemmen aan het woord. Het gaat daarbij om gerenommeerde auteurs, die zorgvuldig documenteren wat ze beweren.

Eerst geven we kort aandacht aan recente waarnemingen van 'Hollandse toestanden' door een Duitse pedagoo. Het kan verkeren: bij ons wordt wel gesproken over 'de Engelse ziekte' als het om het evaluatiesysteem in Engeland gaat, maar die ziekte lijkt besmettelijk te zijn.

"Hollandse toestanden"

Klaus-Jürgen Tillmann, hoogleraar in Bielefeld, beschrijft de tendens naar

internationale vergelijkende studies van de prestaties van leerlingen op terreinen als wiskunde, natuurwetenschappen, lezen en andere facetten van taalonderwijs, vreemde talen. De uitkomsten daarvan leiden vaak tot grote opwindings- en – als een land slecht scoort – soms ingrijpende beleidsmaatregelen om tot verbetering te komen. Dergelijke vergelijkende studies zullen – zo is de gangbare redenering – kunnen bijdragen aan de kwaliteitsverbetering in scholen. Maar dan moet er wel een koppeling gemaakt worden met de evaluatiegegevens van afzonderlijke scholen en zelfs klassen binnen deze scholen. Maar willen we dat wel? Want wat zijn de (onbedoelde) neveneffecten ervan? Wat zijn de valkuilen?

De ontwikkeling in Nederland gaat sterk in deze richting en wordt daarom aan een nadere beschouwing onderworpen. 'Nederland is in Duitsland vooral bekend vanwege het hoge aandeel aan "bijzondere" ("freie") scholen en geldt dan ook als de "schöne Welt" van de schoolautonomie. Andere tendensen zijn veel minder bekend, met name vormen van centrale toetsing, verbonden met het schooltoezicht (inspectie) en de gevolgen daarvan voor het aanscherpen van de kwaliteitscontrole.' Hij beschrijft dan drie elementen van dit systeem (daarbij aansluitend bij een publicatie van de Nederlandse inspectie), zoals 'ranking' van scholen, het IST met de eis van het bepalen van de 'opbrengst' van het onderwijs door middel van erkende, metho-

deonafhankelijke toetsen, de grote invloed van de inspectie op inhoud, didactiek en organisatie van een school, als het gaat om 'verbeteringsplannen'. Hij schrijft dan: 'Dit model wordt door zijn Nederlandse voorstanders gezien als een goed gelukte koppeling van het objectief meten van schoolprestaties, het waarborgen van kwaliteit op schoolniveau en het afleggen van rekenschap aan overheid, ouders, e.a. Van mijn – Duitse – perspectief uit ziet het er eerder uit als een technocratisch controlemodel, waarvan de gevolgen en neveneffecten kritisch gezien moeten worden' (zie voor een vergelijkbare visie Standaard, 2001):

- De koppeling van periodieke centrale toetsing met de evaluatie van individuele leervorderingen en – beoordeling (middels leerlingvolgsystemen) dwingt leraren om hun onderwijs af te stemmen op de toets, waarschijnlijk zelfs tot specifieke trainingen voor de toets. Het is een moderne variant van de 'Nürnberger Trichter' – het 'overbrengen' en 'ingieten' van kennis en vaardigheden.
- De toetsmakers krijgen een ongeloflijk sterke machtspositie.
- 'Ranking' van scholen – in hun grote verscheidenheid – aan de hand van een enkel statistisch kengetal – leidt er via de ouders toe, dat sociale verschillen zich verscherpen. 'Goede scholen' krijgen kinderen met meer kansen, 'slechte' zullen verder 'achterblijven'.

Zijn advies: 'Een stevige discussie is nodig over de vraag of we dit (in Duitsland,) wel willen'.

Algemene negatieve effecten

Hier worden enkele voorbeelden gegeven van omschrijvingen van negatieve effecten van centrale toetsing, (Kohn (2000 en 2001; Meier, 2000; Ohanian 1999 en 2001). Een enkele keer nemen we de vrijheid om daar aan wat toe te voegen.

1. Leraren gaan op den duur de test scores meer vertrouwen dan het eigen professionele oordeel over

kinderen, gebaseerd op langdurige observaties en interacties (Kohn). Dit laatste garandeert op zichzelf nog geen kwalitatief goed oordeel, maar het eerste nog veel minder. Het herinnert aan de kritiek destijds van de Utrechtse onderwijsonderzoeker Karel Stokking op het SAVU-leerlingvolgsysteem, namelijk dat leraren daarvoor 'gedeprofessionaliseerd' worden. In dit verband is bij ons ook regelmatig het argument te horen dat een sterkere druk van buiten de school – met externe evaluatiemethoden – nodig is, omdat de kwaliteit van de leraren ver onder de maat is. Als je deze weg inslaat wordt dit echter een 'self-fulfilling prophecy' (KB). Zie het volgende punt.

2. Goede – creatieve – leraren en directeuren raken gedemotiveerd en verlaten het onderwijs (Kohn).

3. Leraren gaan steeds minder reflecteren op hun werk en leren steeds minder daarvan en van collega's. In plaats hiervan gaan ze zich afschermen en verdedigen en worden ze tevens in feite elkaars concurrenten.

4. Het is evident dat – gezien de grote druk en de ernstige consequenties van laag scoren – fraude uitgelokt wordt.

5. Leraren ontwikkelen in veel gevallen een negatieve houding tegenover kinderen die laag scoren, ook omdat zij erop worden afgerekend.

6. Het gesprek in de school over kinderen en onderwijs wordt ernstig versmald: het gaat niet meer over wat wij als school belangrijk vinden en wat kinderen (vaak verrassend, onverwachts) ontdekt hebben, maar steeds meer over toetsen en hoe die het best gemaakt kunnen worden.

7. In Amerikaanse scholen heeft het systeem van centraal toetsen geleid tot radicale veranderingen in de didactiek, met als resultaat een grotere eenvormigheid, vooral gericht op trainen voor de test. In de weken voor de periodieke toets worden 'niet relevante' vakken verwaarloosd en worden 'inefficiënte' activiteiten zoals excursies, het doen van proefjes, gesprekken over wat je bezighoudt, voorlezen, lezen voor je plezier, etc. opgeofferd (Kohn, Ohanian 1999 en 2001b).

8. Kinderen worden door het systeem van genormeerde toetsen gesorteerd, in plaats van dat vastgesteld wordt wat ze kunnen, hetgeen een pedagogisch ongewenste rivaliteit versterkt (Kohn; Ohanian).

9. Kinderen ontwikkelen een bepaald en zeer eenzijdig, om niet te zeggen vals beeld van wat intelligent gedrag is, namelijk het behalen van goede testresultaten. Hoge testscores duiden eerder op het wennen aan testsituaties, dan op betere leerprestaties. Goede testresultaten geven daarom heel vaak geen valide beeld van wat kinderen kunnen, ze vormen eerder een reflectie van kinderen die gedrild zijn om deze test te kunnen maken.

10. Linda Darling-Hammond geeft dit beeld: stel je voor dat beslist is dat ziekenhuizen beter moeten presteren en de standaard dus verhoogd moet worden. Alle patiënten dienen dan op een gestandaardiseerde wijze – dat wil zeggen op ongeveer hetzelfde moment – getemperatuur te worden. Vlak daarvoor deelt het verplegend personeel hoge doses aspirine en koude dranken uit. De resultaten zijn inderdaad opmerkelijk – niemand heeft koorts! De kwaliteit van dit ziekenhuis is excellent. Uiteraard heeft de werkelijkheid achter deze hoge score weinig te maken met goede medische zorg en het nauwkeurig evalueren daarvan, net zoals het onderwijzen voor de toets volstrekt verschilt van goed onderwijs en het nauwkeurig evalueren daarvan.

11. De gangbare, technocratische, wijze van omgaan met leerstandaarden en van het evalueren daarvan leidt tot standaardisering van inhoud en didactische werkvormen (Darling-Hammond, 1997). Een voorbeeld van het laatste is in Engeland de invoeren van het verplichte 'literacy-hour', met een vast lesschema.

12. Alfie Kohn citeert een test-expert, die voorstelt om de 'league-tables' met toetsgemiddelden van scholen in de krant – als men ze toch beslist wil publiceren – dan maar daar te plaatsen waar ze thuishoren: op de sportpagina's.

13. Deborah Meier (Meier, 2000) benadrukt in haar kritiek de desastreuze gevolgen voor de betrokkenheid van

leraren, ouders en leerlingen en andere mensen en instellingen in de omgeving van de school en het in naam van 'democratische controle' uithollen van democratie in en rond de school. Het begint met het door experts voor het hele onderwijssysteem van een land vaststellen of het wel genoeg 'oplevert', gezien de investeringen, en in vergelijking met andere hoogontwikkelde landen. Daarvoor heb je landelijke normen nodig, die gespecificeerd worden naar schoolsoort, en verder naar verschillende ontwikkelingsniveaus. Dit wordt voorts omgezet in gestandaardiseerde toetsen, waarmee je individuen en scholen 'objectief' kunt vergelijken – regionaal, landelijk, internationaal. Voorts komt er een systeem van sancties voor falende scholen en deze moeten op een 'onpartijdige' wijze vastgesteld worden, dus door instellingen op enige afstand van de scholen. Men meent op deze wijze ook meer rechtvaardigheid te kunnen bereiken – gelijke kansen op een minimum aan goed onderwijs voor alle kinderen. Tenslotte komen er aanwijzingen voor 'effectief' onderwijs die 'verbeteringen' kunnen bewerkstelligen. De stappen in een dergelijk(e) systeem, c.q. denkwijze, bloeien logisch uit elkaar voort. Hieraan liggen aannames ten grondslag over democratie, schoolontwikkeling, kwaliteit en het borgen daarvan, etc.

Meier pleit voor 'alternatieve' aandnames:

- Pluraliteit van scholen als rijkdom – meer definities van wat een 'goede school' is. Waar er een duidelijke hoofdstroom in het onderwijs zichtbaar is, moeten 'alternatieve' visies ook een kans krijgen, binnen ruim gedefinieerde algemene voorwaarden. Dit geeft binnen een onderwijssysteem een gezonde spanning.
- Plaatselijke controle mag niet uitgehouden worden. Naast de oordelen van experts van buiten de locale situatie is democratische zeggenschap ('ownership') van de gemeenschap van direct betrokkenen – leraren, ouders, leerlingen en 'afnemers' van de school

(vervolgsscholen, werkgevers, e.a.) van wezenlijk belang, ook met betrekking tot doelen, inhoud en werkwijzen. Daarin is het ook een belangrijk voorbeeld van democratisch leven voor de kinderen.

Bij de evaluatie moeten verschillende soorten bronnen benut worden, waarbij de nadruk ligt op door de school zelf te formuleren en voor alle betrokkenen heldere ('stevige') standaarden en de plaatselijke toetsing daarvan. Allereerst door systematische interne evaluatie – waarbij de nadruk ligt op het laten zien wat je kunt in een zo echte mogelijke situatie ('performance-evaluatie', ook wel 'authentieke evaluatie' genoemd, zie daarover ook Both, 1995 en Darling-Hammond, c.s., 1995 en 2000), aangevuld door periodieke 'externe audits' door een evaluatiegroep uit de regio (zie hierover ook het slot van 'De Rozentuin'). Evaluatie-activiteiten vanuit de overheid (in de VS de deelstaat) hebben in zo'n situatie het karakter van een aanvullende 'second opinion', binnen flexibel te hanteren landelijke normen.

Goed voor sociaal belemmerde kinderen?

Meier schrijft ook nog over 'maatschappelijke gelijkheid', c.q. sociale rechtvaardigheid. Zij krijgt in het hier gebruikte boek echter repliek van iemand die haar vraagt of ze misschien vergeten is dat het formuleren van leerstandaarden juist bedoeld is om sociaal-economisch bepaalde verschillen aan te pakken. Juist arme gezinnen moeten profiteren van hogere eisen aan de output van het onderwijs. Het onderwijs aan kinderen uit achtergestelde gezinnen is nu slecht tot zeer slecht en dat wordt ook veroorzaakt door lage verwachtingen ten aanzien van deze kinderen, waardoor ze niet voldoende gestimuleerd worden. De reactie is in de trant van: 'Het is toch zo goed bedoeld, en hoe kan dan juist jij, Debbie Meier, daartegen zijn?'

Meier gelooft echter niet dat standaardisering en centrale bureaucratie

sche controle van buiten de plaatselijke gemeenschap hiervoor het goede middel is. 'Het laatste wat we moeten doen is juist het vergroten van de centralisering en standaardisering, dat immers altijd al de scholen voor de armen heeft gedomineerd. Het is juist een deel van het probleem'. Zij wil dat wat in de VS het voorrecht is voor welvarende privé-scholen naar de gewone scholen in achterstandswijken brengen: meer beslissingsvrijheid om deze kinderen en hun gemeenschappen te dienen. Zij wijst ook op het succes dat zij had met de door haar opgebouwde en geleide scholen in East Harlem in New York, de in de VS en daarbuiten beroemde 'Central Park East Elementary and Secondary School', een jenaplanachtige school (Meier, 1996). Deze school wordt gekenmerkt door een sterke betrokkenheid van ouders bij de school, team-teaching van leraren, participatie van leerlingen in de vormgeving van het schoolleven en het onderwijs, kleinschaligheid en gemeenschap (scholen mogen niet te groot worden, of er moeten 'scholen in scholen' gevormd worden), hoge intellectuele en andere standaarden, die helder zijn voor iedereen en waarover ook in het openbaar verantwoording afgelegd wordt (zie boven). Een technocratische verandering van bovenaf dreigt zulke experimenten juist onmogelijk en kapot te maken. In de voorgaande paragrafen is al een en ander opgemerkt over de negatieve effecten voor met name de sociaal-economisch achtergestelde kinderen. Het hangt ook samen met de meestal gebruikte landelijk genormeerde toetsen. Het gaat hier immers om een relatieve maat – de score geeft aan welke plek ('percentielscore') een kind inneemt in het totaal van de populatie die deze test maakt. Als de kinderen bijvoorbeeld veel beter scoren op de rekentoets, dan zullen er toch 'slechte rekenaars' blijven, n.l. de kinderen in de laagst scorende sector (25%), hoezeer je ook je best doet als kind, leraar of school. Het is altijd een systeem van winnaars en verliezers, iemand (of een school) die goed scoort drukt anderen naar beneden. De negatieve effecten van gestan-

daardiseerde toetsen voor met name de kinderen uit arme gezinnen worden zeer gedetailleerd beschreven in het boek van McNeil (2000). De conclusie luidt: '*Standaardiseren reduceert de kwaliteit en kwantiteit van wat in scholen onderwezen en geleerd wordt*'. Dit onmiddellijke negatieve effect van standaardisering is de overheersende conclusie van een studie van scholen, waar de dwang van gestandaardiseerde controles de breedte en kwaliteit van de onderwijsinhouden reduceerde, de rol van de leraren verzwakte en actief leren door de leerlingen vrijwel onmogelijk maakte'.

De effecten op lange termijn zijn nog schadelijker: '*Op lange termijn bezien schept de standaardisering onrechtvaardige ongelijkheid en verbreedt de kloof tussen de onderwijskwaliteit voor jongeren uit arme gezinnen en minderheden enerzijds en die van meer bevoorrechte leerlingen anderzijds hoe langer de gestandaardiseerde controles werkzaam zijn, des te breder wordt deze kloof als het systeem van toetsen en het voorbereiden op de toets in scholen voor kinderen van minderheden in de plaats komt van het curriculum dat meer bevoorrechte leerlingen aangeboden krijgen. Deze nieuwe structuren van discriminatie worden voortgebracht door controles die begonnen in de scholen die in dit onderzoek onder de loep genomen werden en in de loop der jaren het dominante model geworden zijn voor het onderwijs in een van de grootste en qua bevolking meest pluriforme staten van de USA: Texas..... Het documenteert de groeiende macht van deze externe controles om de scholing van alle kinderen om zeep te helpen, in het bijzonder van kinderen met een Afrikaans-Amerikaanse en Latino achtergrond*'.

Greep willen hebben op het onderwijs en de grenzen daarvan

Elliott Eisner, internationaal bekend onderwijskundige en kunstzinnige vormer, graaft dieper en zoekt naar de grondhouding die de basis vormt van

het hele systeem van leerstandaarden en de controle daarvan. (Eisner, 2001). Hij benoemt deze grondhouding als het rationeel willen beheersen van onze omgeving, inclusief de ontwikkeling van mensen en de processen tussen mensen. Het is de opvatting dat kennen = kunnen voorspellen = kunnen beheersen. Zo wil men ook de uitkomst van onderwijsprocessen kunnen voorspellen en de bijdrage die onderwijs levert aan de ontwikkeling van kinderen. Je formuleert daarom zo nauwkeurig mogelijk de beoogde uitkomsten, zoveel mogelijk in meetbare vorm op korte en langere termijn (van lessen, leerjaren, enkele jaren, het basisonderwijs), stelt de beginsituatie van kinderen vast en gaat na afloop na of de beoogde uitkomsten bereikt zijn, c.q. wat het onderwijs aan de ontwikkeling van kinderen bijgedragen heeft: de 'toegevoegde waarde'. 'Kwaliteit' moet uitgedrukt worden in kwantiteit, vanwege de grotere 'objectiviteit', meten is weten. Er wordt gestreefd naar zo universeel mogelijk geldende, 'wetenschappelijk vastgestelde' onderwijsmethoden ('what works'), die de beoogde uitkomsten zouden garanderen en die dan ook algemeen ingevoerd moeten worden. Als 'wetenschappelijk' geldt in dit verband alleen een 'harde' wetenschappelijke benadering als legitiem – alleen meetbare effecten tellen. En zo lijkt de cirkel zich te sluiten:

- a. er wordt van uitgegaan dat 'weten' is kunnen voorspellen en kunnen voorspellen is beheersen; bovendien kan je binnen deze benadering alleen zeker weten door te meten (kwantificering) en op grond van deze beide aannames wordt de beoogde 'opbrengst' zo nauwkeurig mogelijk geformuleerd;
- b. met hulp van een methode die de voorspelde uitkomsten zo goed mogelijk garandeert wordt aan deze doelen gewerkt;
- c. de 'opbrengst' wordt gemeten (diagnose);
- d. voor die kinderen die onvoldoende gescoord hebben wordt de methodiek eventueel aangescherpt.

Deze beheersingsfilosofie leidt gemakkelijk tot allerlei reducties:

- Doelen die niet of moeilijk meetbaar / meetbaar te maken zijn raken uit het zicht. Er is dan bijvoorbeeld onvoldoende geld om voor complexe vaardigheden passende evaluatie-instrumenten te ontwikkelen, of men vindt de beoogde doelen al gauw 'te vaag'.
- De reductie van onderwijsmethoden tot wat eenvoudig meetbare doelen dient.
- De nadruk op 'hapklare', 'overzichtbare' modules, met korte leercycli, ten koste van langer lopende leerprocessen die per definitie niet te modulariseren zijn.
- Het schoolklimaat wordt steeds meer gericht op extrinsieke motivatie.

Eisner beweert niet dat 'beheersingle- ren' en 'directe instructie' met een meetbare feedback op zichzelf verwerpelijk zijn, integendeel. Zij kunnen heel functioneel zijn, ook binnen 'zelfgestuurd' leren. Hij waarschuwt echter – op grond van wat hij in talloze scholen waarneemt – voor een verre-gaande inhoudelijke en didactische verschraving van het onderwijs. Het merkwaardige is dat momenteel in de onderwijswetenschappen tegelijker-tijd sterker benadrukt wordt dat leren allereerst is, c.q. moet zijn: persoonlijk betekenis geven aan wat je leert, mede in interactie met anderen. En verder: waar blijven de verrassing, de verbeeldingskracht, het improviseren? Is onderwijs niet steeds ook: in-teractie, met alle onvoorspelbaarheid daarvan? (zie ook: Eisner, 1992 en 1998). Maak je met de sterke nadruk op het willen beheersen essentiële di-mensies van het onderwijs niet kapot? Ik wijs hier op soortgelijke opvattingen die Peter Petersen in zijn 'Onderwijspedagogiek' al beschreef onder het kopje 'pedagogische situ-atie'.

Stel je eens voor – zegt Eisner - dat we voor de komende vijf jaar een mo-ratorium op landelijk afgenomen ge-standaardiseerde toetsen zouden afkondigen. Je kunt dan dus bij het beoordelen van scholen niet meer af-gaan op hogere testcores. Waar zou

je (als ouders, inspectie, etc.) wel op kunnen letten? Eisner geeft dan een hele lijst van kwaliteitsindicatoren, waaruit we er enkele noemen:

- De soorten problemen en activitei-ten waarin leerlingen zich engage-ren. Worden ze gestimuleerd tot verwondering, tot het stellen van vragen over wat ze bestudeerd hebben? Let wel: hun eigen vra-gen en niet zozeer de vragen die wij hen stellen? De belangrijkste intellectuele prestatie die je kunt leveren is immers niet zozeer het oplossen van problemen, maar het formuleren ervan, de vragen die je kunt stellen.
- Worden de leerlingen geconfron-teerd met meer en verschillende perspectieven? Kunnen ze daar-mee omgaan, zonder in de war te raken?
- Kennismaken met verschillende manieren van kennen van de wer-kelijkheid, die elkaar niet uitslui-ten, maar elkaar juist aanvullen: wetenschap, kunst, filosofie, e.a.?
- Worden er relaties gelegd tussen wat op school geleerd wordt en de wereld buiten de school, kun-nen de leerlingen toepassen wat zij geleerd hebben? 'De werkelijk belangrijke afhankelijke variabelen in het onderwijs zijn niet gelokali-seerd in klaslokalen, noch in schoolgebouwen, maar buiten de school. Onze evaluatiepraktijken komen zelfs nog niet in de buurt daarvan. De echte uitkomst van het onderwijs is daar te vinden waar leerlingen kunnen doen wat ze zelf willen doen'.

De aantrekkingskracht van getallen

In de lijn van Eisner doordenkend kan je je afvragen waarom mensen zoveel waarde hechten aan meten en aan getallen. Er wordt door de mensheid heel wat geteld, gemeten en bere-kend. Voor een deel ligt dat in het ver-lengde van natuurlijke processen – zo wordt er in ons eigen lijf al heel wat geteld, gemeten en gerekend en vin-den op grond van gevonden waarden, (bijvoorbeeld van hormoonspiegels) terugkoppelingen plaats. In de sa-menleving, zeker in onze moderne

samenleving, is het een prominent verschijnsel – zie alleen al de aanduidingen op verpakkingen. Dat is uitermate functioneel, maakt zaken controleerbaar. Het is niet overdreven te zeggen dat onze maatschappij 'loopt op getallen'. We gaan er daarbij als regel maar van uit dat de getallen kloppen. Het uitdrukken van kwaliteiten in getallen is overigens niet zonder problemen. Niet alles laat zich zo 'vangen'.

Pedagoge Selma Wassermann, internationaal bekend van haar bemoeienissen met 'leren denken' en met 'ontwikkelen onderwijs' voor jonge kinderen, stelt ook de vraag waarom mensen zoveel waarde hechten aan getallen. (Wassermann, 2001). Het gaat volgens haar dieper dan 'inter-subjectieve controleerbaarheid' zoals in de wetenschap aan de orde is en ook dieper dan de praktische functie in de samenleving. Het heeft vooral een existentiële dimensie: In een snel veranderende wereld voelen mensen zich onzeker, ervaren stress, moeten talloze beslissingen nemen op grond van onvoldoende zekerheden en waarvan zij de gevolgen ook niet kunnen overzien. In een dergelijke als chaotisch ervaren werkelijkheid is daar dan het getal. 'Niets voelt zekerder of geeft meer zekerheid dan een getal. Het doet er daarbij niet toe of de getallen of het meten voldoende nauwkeurig zijn, we voelen ons eenvoudigweg geruster, als we kunnen geloven dat er iets is dat vastigheid biedt, iets waarop we met zekerheid kunnen rekenen, iets waaraan we ons kunnen vastklemmen als we met grote snelheid onzekere tijden tegemoet gaan.'

Misschien ligt het inderdaad op dit existentiële niveau, waarom ouders en leerkrachten zo graag leerprestaties van kinderen willen uitdrukken in een enkelvoudig getal. Percentielscore 82 op de CITO-toets 'geeft houvast', zo voelt het voor velen tenminste. Onlangs werd van een schoolleider vernomen dat zij elk jaar weer 'gezeur hadden' met veel ouders over de adviezen voor het voortgezet onderwijs. 'We zijn toen begonnen met de CITO-eindtoets en het gezeur was vrijwel over'.

In dergelijke gevallen wordt zelden gevraagd naar de betrouwbaarheid en validiteit van het getal, die zijn soms ronduit bedenkelijk. Wassermann geeft in haar artikel voorbeelden van schijnexactheid in gevallen waar het ging om leven of dood (in de ruimtevaart), van regelrechte fouten in de wereld van natuurwetenschap en techniek, waarbij ook Nobelprijswinnaar Richard Feynman wordt geciteerd, die waarschuwt tegen een blind geloof in getallen.

'Ik denk dat het hard nodig is dat leraren geïnformeerde, evaluatieve observaties doen in hun praktijk. Ja, het is waar dat deze onjuist kunnen zijn; ja, het is waar dat ze subjectief zijn; ja, het is waar dat de oordelen van leraren over de leerprestaties van kinderen verkeerd, misleidend of bevooroordeeld kunnen zijn. Maar in plaats van deze oordelen van leraren te vervangen door het veel minder bruikbare alternatief van gestandaardiseerde toetsen kunnen we wellicht onze energie beter richten op het verbeteren van de door de leraar uitgevoerde evaluaties, door ze bruikbaar en ondersteunend te maken voor leerlingen en ouders en voor het totale leerproces.' Zie in verband hiermee de ontwikkeling rond portfolio's (Castelijns, e.a., 2002).

Heeft de keizer (nog) kleren aan?

Gestandaardiseerde toetsen leunen sterk aan tegen de denk- en werkwijzen van het klassieke intelligentiebegrip. Er wordt – in dit 'psychometrische' model – uitgegaan van de volgende vooronderstellingen:

- Er bestaat bij elk mens zoiets als een redelijk samenhangend en vast te stellen geheel aan denkvaardigheden dat 'intelligentie' genoemd wordt;
- mensen beschikken in meerdere of mindere mate over intelligentie en dat kan gemeten worden en samengevat in een enkelvoudig getal, het IQ;

Dit IQ is redelijk stabiel en geeft de bovengrens van de intellectuele ontwikkeling van de betreffende persoon aan. Dit psychometrische model wordt steeds sterker bekritiseerd (zie

ook Gipps, 1998 en Gould, 1996). Allereerst bestaat er niet zoiets als 'de' intelligentie van een persoon, maar wordt intelligentie steeds meer gezien als een bundel van veelvormige mogelijkheden, die bij alle mensen in potentie aanwezig zijn, maar bij verschillende mensen in verschillende mate aanwezig zijn, c.q. ontwikkeld kunnen worden., met bij verschillende mensen verschillende accenten. Howard Gardner spreekt over 'meervoudige intelligenties' en beschreef er inmiddels negen. Naast het IQ hebben we nu bijvoorbeeld ook al het EQ, als maat van de emotionele intelligentie. Bovendien blijkt de stabiliteit en voorspelbaarheid van het leerplafond veel minder groot te zijn. Zie het werk van Feuerstein, die weigert van een vooraf vastgelegd plafond aan mogelijkheden uit te gaan en bijvoorbeeld met kinderen met het syndroom van Down verbazingwekkende resultaten bereikte. De huidige praktijk van toetsen pakt vaak slecht uit voor kinderen, met name bij kinderen in sociaal-economische achterstandsituaties. Binnen een project aan de universiteit van Cambridge (Engeland) wordt onderzocht of nieuwe veronderstellingen over intelligentie en cognitieve ontwikkeling een verschil kunnen maken. We volgen de redenering die aan dit project ten grondslag ligt (Hart, 1998).

In Engeland probeert men in het kader van de toetsen bij het National Curriculum en het publiceren daarvan in de krant de vergelijking eerlijker te maken voor scholen met veel 'achterstandskinderen'. Zij scoren bijna altijd laag, gezien hun leefsituatie en scholen met veel van deze kinderen bengelen onderaan de lijst. Het op een verantwoorde wijze corrigeren van de ruwe toetsscores met het oog op sociaal-economische factoren blijkt een ondoenlijke opgave (Standaert, 2001). Daarom worden, om een eerlijker vergelijking te krijgen, aan het begin van de schooltijd, in basis- en voortgezet onderwijs, de cognitieve vaardigheden getoetst. Deze toets moet de basis zijn voor het voorspellen van het uiteindelijk te behalen niveau aan het eind van de betreffende schoolperiode (over x jaar

du), met het verwerken van de te verwachten natuurlijke ontwikkeling van het kind en de toegevoegde waarde van de school. Er wordt van uitgegaan dat je het 'eindniveau' kunt voorspellen. Leerlingen worden op grond van de toetscores – vooral in het VO, minder in het primair onderwijs, – gegroepeerd naar 'ability', met daarbij aanvullende ondersteuning. Kinderen worden in feite vastgelegd op een vooraf bepaald niveau. De verwachtingen van leraren van wat de kinderen kunnen wordt hierop afgestemd, bewust of (vaker) onbewust. Voor dat deel van de kinderen waarvan je kunt veronderstellen dat ze onder het voorspelde niveau zitten kan dat gunstig zijn, zij kunnen voldoende uitgedaagd worden. Maar steeds meer blijkt dat een groot deel van de kinderen in dergelijke situaties te weinig uitgedaagd wordt en meer kan bereiken. Die kinderen worden dus door deze aanpak feitelijk benadeeld en tevens beperkt het in de praktijk leraren om te doen wat in hun vermogen ligt om de leerresultaten te verbeteren. Vanuit een pluriform en niet-deterministische opvatting over intelligentie, c.q. leerpotentie ga je heel anders te werk, doe je meer recht aan de kinderen en hun veelvormige leerpotentie.

Als dit doorzet, dan zal ook blijken dat je met de huidige praktijk van gestandaardiseerde, landelijk genormeerde toetsen niet veel meer kunt aanvangen, omdat de grondslagen ondermijnd zijn. Het wachten is op degene die roept dat de keizer geen kleren meer aan heeft.

Naar een netwerksamenleving

Op korte termijn bezien hebben we echter in ons land en elders nog te maken met een sterke lobby voor het invoeren van leerstandaarden, verbonden met een technocratisch controlesysteem. Er zijn echter steeds meer stemmen, die dit type van onderwijsbeleid beschouwen als naweeën van een maatschappijvorm die aan het verdwijnen is, namelijk de industriële samenleving (Spady, 2001; in ons land Van Aalst, 2000). Metafoor voor deze samenleving is de

lopende band, met massaproductie, een uniforme kwaliteitscontrole, een ver doorgevoerde taakverdeling en rationalisatie (standaardisering) zoals ook Eisner die beschrijft. Ook in de school zijn deze kenmerken terug te vinden. De scholen kunnen, zo dacht en denkt men soms nog steeds, met dit model even efficiënt leerlingen met gespecificeerde kenmerken produceren als Henry Ford T-Fordjes produceerde. Zie de vormen van onderwijsplanning, het leerstofjaarklassensysteem ('grading'), de gefragmentariseerde aanpak van de leerstof, de problemen met 'langzame leerlingen' binnen dit systeem. Veel scholen lopen tegen de grenzen van deze aanpak aan, ook omdat de samenleving, inclusief de kinderen, leraren en ouders aan het veranderen is, langzaam maar zeker. Ook de in de vorige paragraaf beschreven dynamische, niet-deterministische visie op leren en ontwikkeling past in een nieuw maatschappijbeeld, dat getypeerd kan worden als een *netwerkmaatschappij*. Hiërarchische structuren worden in toenemende mate ter discussie gesteld en de samenleving kan steeds minder van bovenaf bestuurd worden. Dit staat haaks op de hierboven beschreven trend. Daar geldt: hoe verder van de school verwijderd, hoe hoger in de hiërarchie, hoe enthousiaster men is over centrale toetsing en omgekeerd: hoe meer kennis van het onderwijs men heeft, dat wil zeggen hoe dichter men bij de schoolpraktijk staat, hoe kritischer en genuanceerder men daarover denkt (Kohn, 2000). Leren in de netwerkmaatschappij is bovendien veel minder lineair, er is meer aandacht voor verbeeldingskracht en creativiteit.

Over deze zich ontwikkelende netwerkmaatschappij valt veel meer te zeggen (zie met name Van Aalst, 2000). Een dergelijk perspectief kan ons helpen de huidige problemen zelfbewust, met humor en verbeeldingskracht tegemoet te treden en creatief verzet te plegen, netwerken van gelijkgerichte individuen, scholen en organisaties te vormen. ■

BRONNEN

- Aalst, H. van (2000), *Leren in de netwerkmaatschappij*, in: *Mensen-kinderen*, Tijdschrift voor en over Jenaplanonderwijs, jrg. 16, nr. 1 (sept.), p. 10-15
- Both, K. (1999), *Authentiek leren en authentieke evaluatie*, in: *Mensen-kinderen*, jrg. 14, nr. 3, p. 4-7
- Castelijns, J./ B. Kenter/ J. Janssen (2002), *De kracht van portfolio's*, Amersfoort (CPS)
- Darling-Hammond, L./ J. Ancess/ B. Falk (1995), *Authentic Assessment in Action*, New York/ London
- Darling-Hammond, L. (1997), *The Right to Learn. A Blueprint for Creating Schools that Work*, San Francisco
- Darling-Hammond/ J. Ancess/ B. Falk (2000), *Authentieke evaluatie in actie (2)*, in: *Mensen-kinderen*, jrg. 15, nr. 3, p. 12-15
- Eisner, E.W. (1992), *Educational Reform and the Ecology of Schooling*, in: *Teachers College Record*, Vol. 93, nr. 4, p. 610-627
- Eisner, E.W. (1998), *The Kind of Schools We Need. Personal Essays*, Portsmouth (NH)
- Eisner, E.W. (2001), *What does it mean to say a school is doing well?*, in: *Phi Delta Kappan*, Vol. 82, nr. 1, p. 367-372
- Gipps, C.V. (1998), *Beyond Testing: Towards a Theory of Educational Assessment*, London/ Washington, D.C.
- Gould, S.J. (1996), *De mens gemeten. De geschiedenis van de intelligentietest*, Amsterdam/ Antwerpen
- Hart, S. (1998), *A Sorrytail: Ability, Pedagogy and Educational Reform*, in: *British Journal of Educational Studies*, Vol. 46, nr. 2, p. 153-168
- Kohn, A. (2000), *The Case Against Standardized Tests. Raising the Scores, Ruining the Schools*, New York
- Kohn, A. (2001), *Fighting the Tests. A practical guide to rescuing our schools*, in: *Phi Delta Kappan*, Vol. 82, nr. 1, p. 349-357
- cNeil, L.M. (2001), *Contradictions of School Reform. Educational Costs of Standardized Testing*, New York
- Meier, D. (1996), *The Power of Their Ideas. Lessons for America from a Small School in Harlem*, Boston
- Meier, D./ J.Kozol/ J. Cohen (eds.) (2000), *Will Standards Save Public Education?*, Boston
- Ohanian, S. (1999), *One Size Fits Few. The Folly of Educational Standards*, Portsmouth
- Ohanian, S. (2001a), *Caught in the Middle. Nonstandard Kids and a Killing Curriculum*, Portsmouth
- Ohanian, S. (2001b), *De uitdaging van de keuzevrijheid*, in: *Mensen-kinderen*, jrg. 17, nr. 1, p. 7-8
- Spady, W. (2001), *Beyond Counterfeit Reforms. Forging an Authentic Future for All Learners*, Lanham/ Maryland/ London
- Standaert, R. (2001), *The temptation of simplification*, in: H. Bachmann, e.a. *Turning the Perspective. New Outlooks for Education*, Eenschede. Een Nederlandse versie is opgenomen in de ondersteunende brochure van de actie 'Streep door de eindtoets'.
- Tillmann, K.J. (2001), *Leistungsvergleichsstudien und Qualitätssicherung, oder: Auf dem Weg zu holländischen Verhältnissen?*, in: *Neue Sammlung*, Vol. 41, Heft 3, S. 359-368
- Wassermann, S. (2001), *Quantum Theory, the Uncertainty Principle and the Alchemy of Standardized Testing*, in: *Phi Delta Kappan*, sept. 2001, p. 28-40