

# Een andere kijk op toetsen

Frans J.G. Janssens

Universiteit Twente en Inspectie van het Onderwijs<sup>1</sup>

## Inleiding

De Kwaliteitsagenda PO richt zich in het bijzonder op de verbetering van de reken- en taalprestaties van leerlingen in het primair onderwijs. Aangenomen wordt dat de toetspraktijk, 'data-driven teaching' en het beter gebruik maken van toetsgegevens door leraren, een belangrijke succesfactor kan zijn voor de verbetering van de leerprestaties. Onder toetspraktijk wordt in dit verband verstaan het gebruik van professioneel ontwikkelde schoolvorderingentoetsen door scholen, zoals de toetsen uit het Leerlingvolgsysteem, de Entree- en de Eindtoets basisonderwijs van het Cito. Professioneel ontwikkelde toetsen vormen inderdaad een belangrijk onderdeel van de huidige toetspraktijk in het basisonderwijs. Maar op scholen worden ook andere activiteiten ondernomen die gericht zijn op het vaststellen van de schoolvorderingen: het aanpassen van het aanbod en de instructie aan de vorderingen van leerlingen en het beoordelen van de kwaliteit van het onderwijs die bereikt is bij een of meerdere leerjaren. Daarin spelen ook methodegebonden toetsen, proefwerken, overhoringen, de beoordeling van schriftelijk werk, de beoordeling van spreekbeurten, portfolio's, zelf-evaluaties van leerlingen, etc. een belangrijke rol. Het ligt daarom voor de hand eerder te spreken van een evaluatiepraktijk dan van een toetspraktijk<sup>1</sup>.

Als we de evaluatiepraktijk op een betekenisvolle manier willen koppelen aan prestatieverbetering van leerlingen, dan moeten we misschien met een andere blik naar de toetspraktijk in het PO kijken. Dan komen ook die andere activiteiten die gericht in beeld die gericht zijn op het verzamelen van gegevens om beslissingen te nemen over het onderwijs. Evalueren in het onderwijs dient vele doelen, maar zijn terug te brengen tot twee basale functies. Enerzijds wordt er met verschillende middelen en technieken geëvalueerd om de stand van zaken op te maken (= terugkijken, ook wel summatieve evaluatie genoemd). Anderzijds wordt er in de onderwijspraktijk geëvalueerd om beslissingen te nemen over de voortgang van het onderwijs (= vooruitkijken, formatieve evaluatie genoemd). Beide functies vinden plaats op leerling-, groeps- en schoolniveau<sup>2</sup>.

Misschien moet niet de vraag gesteld worden of en hoe het toetsgebruik kan worden verbeterd, maar moet er eerder een antwoord worden gezocht op de vraag of er in het basisonderwijs sprake is van een goede balans in het gebruik van toetsen en andere evaluatietechnieken op en tussen de drie genoemde niveaus? Zonder antwoord op deze vraag zal er geen sprake zijn van de gewenste prestatieverbetering. Ik zal uitleggen waarom dat zo is.

De VS zijn een natie die geobsedeerd is door het geloof dat de weg naar schoolverbetering is

---

<sup>1</sup> De auteur is zowel coördinerend inspecteur bij de Inspectie van het Onderwijs als bijzonder hoogleraar Onderwijstoezicht aan de Universiteit Twente

geplaveid met betere en meer gestandaardiseerde toetsen. Het probleem is echter dat deze verplicht gestelde toetsen, die bedoeld zijn om alle leerlingen bij de les te houden, vaak het tegenovergestelde bewerkstelligen. Veel leerlingen geven als gevolg van slechte toetsresultaten de hoop op en scholen besluiten steeds vaker zwakke leerlingen niet meer aan deze toetsen te laten meedoen<sup>3</sup>. Deze tendens is ook in Nederland waarneembaar bij de afname van de Cito-toets<sup>4</sup>. Daar komt nog bij dat een eindtoets, omdat deze terugkijkt, maar betrekkelijk weinig bijdraagt aan prestatieverbetering van zittende leerlingen in het basisonderwijs. Hooguit speelt deze toets achteraf een rol bij de beoordeling van het niveau van het onderwijs zoals gegeven aan groep leerlingen die school inmiddels heeft verlaten. Dat betekent dus dat leraren vooral op andere evaluatie-instrumenten zijn aangewezen om het onderwijs op zittende leerlingen af te stemmen. Ook daar bestaan professioneel ontwikkelde toetsen voor, zoals de Entreetoets en het LVS van het Cito. Het is echter de vraag of deze toetsen en de gangbare evaluatiepraktijk in het algemeen toereikend genoeg zijn om het onderwijs daadwerkelijk af te stemmen op de behoeften van leerlingen. Het is ook nog maar de vraag of leraren voldoende deskundig zijn in de evaluatie van leervorderingen en de consequenties ervan om te zetten in 'passend onderwijs'.

Het is niet de bedoeling een pleidooi te houden tegen de huidige toetspraktijk, tegen het afleggen van verantwoording over de opbrengsten door scholen of tegen de manier waarop de inspectie de opbrengsten gebruikt in het toezicht. De huidige toetspraktijk stelt ons immers in staat te reflecteren op de leerprestaties en op de wijze waarop er toezicht gehouden moet worden op scholen die onderpresteren. Ook de scholen zelf moeten kunnen laten zien wat ze doen en kunnen. Als genormeerde toetsen daaraan een bijdrage kunnen leveren, dan moeten die ook worden gebruikt. Maar willen toetsen bijdragen aan prestatieverbetering, dan moeten we ook beseffen dat de toetspraktijk hierin maar een beperkte rol kan spelen. Ze vormen immers maar een deel van de evaluatiepraktijk en zijn lang niet altijd gericht op directe sturing van leerprocessen bij leerlingen die nog aan het leren zijn. Zo rapporteren onderzoekers<sup>5</sup> dat in de VS het gebruik van gestandaardiseerde toetsen maar tot een beperkte verbetering van de leerprestaties heeft geleid. Tegelijkertijd laten anderen<sup>6</sup> zien dat dergelijke toetsen ook neveneffecten hebben als prestatiedaling en toename van dropouts, vooral onder leerlingen die afkomstig zijn uit etnische minderheidsgroepen<sup>7</sup>.

### **Huidige visie op de toetspraktijk**

De afgelopen jaren heeft het denken over toetsgebruik geleid tot de opvatting dat voor verbetering van de opbrengsten het volgende nodig is (Zie bijvoorbeeld het advies van de Onderwijsraad over leerstandaarden):

1. Ontwikkel leerstandaarden;
2. Pas toetsen toe om deze standaarden te meten;
3. Leg verantwoording af over de resultaten.

De gedachtegang is als volgt: De leerstandaarden geven aan welke resultaten met leerlingen bereikt moeten worden. De toetsen dienen het bewijs te leveren dat de standaarden inderdaad bereikt zijn. De verantwoording door scholen is erop gericht te laten zien of scholen deze standaarden bij hun leerlingen hebben bereikt. Omdat de hoogte van de leerprestaties een rol speelt in het toezicht, neemt de summatieve impact van toetsen toe. Zo is de eindtoets van het Cito maatschappelijk een steeds grotere rol gaan spelen buiten de scholen doordat de inspectie die toets gebruikt om beslissingen te nemen over het toezicht en krijgen, om dezelfde reden, ook op de tussenopbrengsten een summatief accent. In deze context vertrouwen we erop dat leerprestaties zoals gemeten met toetsen, een goede informatiebron zijn van hetgeen leerlingen hebben geleerd, hadden moeten leren en voor de vraag of scholen hun best wel hebben gedaan om alles uit hun leerlingen te halen. Dat

toetsen primair een didactische functie hebben, raakt op deze manier wellicht onderbelicht.

We zijn er altijd van uitgegaan dat het gebruik van toetsen twee belangrijke doelen diende, namelijk dat ze enerzijds gegevens verschaffen voor het nemen van beslissingen over het onderwijs en dat ze leerlingen motiveren. We zijn opgegroeid in scholen waarin leraren erin geloven dat het gebruik van toetsen, proefwerken, tentamens, etc. leidt tot het maximaliseren van de inspanning van leerlingen om zo goed mogelijk leerresultaat te behalen. Voor heel veel leerlingen werkt dit stimulerende mechanisme ook, maar lang niet voor allemaal.

Een vergelijkbare veronderstelling ligt ook ten grondslag aan het streven om scholen verantwoording af te laten leggen over hun prestaties en deze te betrekken in het onderwijstoezicht. Indien scholen hun leerprestaties transparant maken, of wanneer de inspectie dat doet, zal dit een prikkel zijn om deze prestaties te behouden of te verbeteren<sup>8</sup>. Maar net zoals toetsen niet alle leerlingen prikkelen tot prestatieverbetering, geldt ook voor scholen dat het gebruik van toetsen of de openbaarmaking van hun prestaties niet altijd automatisch leidt tot schoolverbetering.

De belangrijkste beperking van het beschikbare toetsinstrumentarium is echter de mogelijkheid om toetsresultaten te gebruiken voor de sturing van het onderwijs aan zittende leerlingen. Daarvoor is het nodig dat deze toetsen dekkend zijn voor de inhoud van het onderwijs (liever nog: geïntegreerd zijn in het onderwijs), voldoende diagnostisch vermogen hebben en dat ze aanwijzingen geven over de wijze waarop het onderwijs op de leerlingen kan worden afgestemd. De Cito-eindtoets is een afsluitende toets en heeft om die reden geen gevolgen meer voor de leerlingen bij wie deze is in het basisonderwijs is afgenomen<sup>9</sup>. De Entreetoets en het LVS hebben deze mogelijkheden wel maar vanwege het methode-onafhankelijke karakter van beide is de vertaling van de individuele uitkomsten ervan naar de instructie niet vanzelfsprekend. Maar zoals eerder betoogd spelen niet alleen toetsresultaten een rol in de afstemming van het onderwijs op de leerlingen, maar de hele evaluatiepraktijk.

Het is dus zeer de vraag of verbetering van de toetspraktijk de enige manier is om te bevorderen dat de prestaties van leerlingen zullen verbeteren. Daarvoor moeten we misschien toch onze visie op toetsen verbreden en van andere accenten voorzien.

### **Een andere visie**

Er is een andere visie mogelijk op de relatie tussen de toetspraktijk en prestatieverbetering. In de bovenstaande benadering staan toetsen vooral in dienst van het vaststellen van schoolvorderingen en het terugkijken op de resultaten. Er is ook een benadering mogelijk die vooral vooruit kijkt en waarbij evaluatie meer in dienst staat van de ontwikkeling van het onderwijs. In de eerste visie staan toetsen in dienst van resultaatbepaling en van het afleggen van verantwoording, terwijl de tweede visie meer gericht is op het verbeteren van de leerresultaten door middel van de afstemming van het aanbod en de instructie op leerlingniveau. We zouden deze visie 'voortgangsevaluatie' kunnen noemen<sup>10</sup>. Beide visies zijn van belang voor de verbetering van het onderwijs, maar ze zijn thans in de onderwijspraktijk niet in balans. Het verdient aanbeveling meer te investeren in de tweede visie. Er valt simpelweg meer leerwinst te behalen als we erin slagen de evaluatiepraktijk, inclusief het gebruik van toetsen, meer te richten op voortgangsevaluatie op leerlingniveau, waardoor het een krachtig instructiemiddel wordt. Als scholen de taal- en rekenprestaties van hun leerlingen moeten verbeteren, moeten we leraren de evaluatietechnieken bieden waarmee dat kan worden gerealiseerd. Er valt dan meer te verwachten van de op- en herwaardering van de dagelijkse evaluatiepraktijk, dan alleen van stimulering van toetsgebruik. Denk bijvoorbeeld aan de ooit zo populaire zogenaamde foutenanalyse, waarin leraren werden geschoold die in de jaren zestig en

zeventig hun opleiding volgden<sup>11</sup>.

Het gaat dus om meer dan alleen frequenter toetsen of een beter gebruik van toetsresultaten, ofschoon deze elementen wel van belang zijn en blijven, maar om alle evaluatietechnieken en -procedures die in de klas worden gebruikt om de instructie beter af te stemmen op de vorderingen van leerlingen. Uit onderzoek is bekend dat bij deze benadering ook de leerlingen betrokken moeten worden<sup>12</sup>. Het is voor deze werkwijze van belang dat leraren:

- voorafgaand aan het leerproces de leerdoelen voor hun leerlingen vaststellen;
- hun leerlingen vanaf het begin, in hun eigen taal, informeren over de leerdoelen die bereikt moeten worden
- vakkundig worden in het evalueren en daardoor hun verwachtingen kunnen omzetten in evaluatie-activiteiten en scoringsprocedures die nauwkeurig de prestaties van hun leerlingen in beeld brengen;
- evaluatietechnieken gebruiken waarmee leerlingen vertrouwen in zichzelf krijgen en verantwoordelijkheid krijgen in hun eigen leren:
- evaluatie-uitkomsten vertalen naar voorschrijvende feedback aan leerlingen waarmee deze inzicht krijgen hoe de resultaten te verbeteren zijn;
- voortdurend de instructie afstemmen op de uitkomsten van voortgangsevaluaties;
- leerlingen betrekken in regelmatige zelf-evaluaties op basis van vaste standaarden zodat leerlingen zelf hun vorderingen kunnen vaststellen en voelen dat ze de leiding hebben over hun leersucces;
- leerlingen actief betrekken in de communicatie met hun leraar en hun ouders over hun resultaten en de vorderingen

### **Zal het werken?**

Er zal meer aandacht moeten worden besteed aan het richten van de evaluatiepraktijk in het basisonderwijs op de voortgang van het leren van (individuele) leerlingen. Ook verbetering van de toetspraktijk is van belang maar de toetsen die in het primair onderwijs worden gebruikt zijn niet voor alle evaluatiefuncties geschikt. Ze worden veelal gebruikt om de tussen- en eindopbrengsten op groepsniveau te beoordelen en dienen dus de resultaatbepaling. Bovendien bevatten deze toetsen te weinig concrete didactische aanwijzingen om het onderwijs op de toetsresultaten af te stemmen. Te weinig aandacht is uitgegaan naar de verdere ontwikkeling van de evaluatiepraktijk waarmee:

- leraren dagelijks over diagnostische informatie kunnen beschikken om leerprocessen bij te sturen;
- leerlingen zelf inzage krijgen in de wijze waarop ze tot goede leerresultaten kunnen komen en
- ook ouders informatie krijgen over de wijze waarop ze de vorderingen van hun kinderen kunnen ondersteunen.

Daarvoor zijn evaluatietechnieken nodig die leraren informatie geven over de voortgang van leerprocessen en waarmee aanwijzingen worden verkregen om de instructie te sturen.

De hamvraag is natuurlijk of het in balans brengen van de visies ook leidt tot betere leerprestaties? Verschillende studies laten zien dat dit een vruchtbare weg kan zijn.

In 1984 publiceerde Benjamin Bloom<sup>13</sup> een overzicht van onderzoek waarin klassikaal onderwijs (controle-groep) werd vergeleken met twee experimentele condities: beheersingsleren en individueel onderwijs. In deze experimentele condities werd veelvuldig gebruik gemaakt van leerdoelgerichte toetsen om de instructie te ondersteunen. Uit vergelijking van de uitkomsten bleek dat de leerwinst in de experimentele groepen 1 tot 2 standaarddeviaties groter was dan in de controle groep.

In 1998 publiceerden Paul Black en Dylan Wiliam<sup>14</sup> een internationale overzichtsstudie van de uitkomsten van onderzoek naar het effect van leerlingevaluatie. Ze vroegen zich af formatieve voortgangsevaluatie tot betere leerprestaties leidt dan summatieve evaluaties gericht op resultaatbepaling. Ze analyseerden meer dan 250 studies over dit onderwerp. Enkele tientallen onderzoeken waren direct gericht op de vraag naar de invloed van voortgangsevaluatie op de leerprestaties. Hun analyse toonde aan dat formatieve evaluaties tot meer leerwinst leiden dan het gebruik van summatieve toetsen; de effectgrootte betrof een halve tot een hele standaarddeviatie<sup>15</sup>. Bovendien bleek dat zwakke leerlingen meer gebaat zijn bij formatieve evaluaties dan andere leerlingen. Er was niet alleen sprake van een reductie van prestatieverschillen tussen zwakke en andere leerlingen, maar zelfs van prestatieverhoging bij alle leerlingen.

De studie van Black en Wiliam laat zien dat prestatieverbeteringen te verwachten zijn in situaties waarin leraren de nauwkeurigheid van voortgangsevaluaties verhogen, leerlingen frequent informatieve feedback geven en leerlingen bij dergelijke evaluaties betrekken, de vorderingen bijhouden en aan betrokkenen communiceren. Na 2000 zijn er nog meer studies verschenen met vergelijkbare positieve uitkomsten ten gunste van meer gebalanceerde evaluatiepraktijken<sup>16</sup>.

Een nieuw inzicht in het evalueren op school is de noodzakelijkheid hierbij leerlingen te betrekken<sup>17</sup>. Dit is een belangrijk onderdeel van de voortgangsevaluatie omdat goede feedback over leervorderingen uit drie onderdelen bestaat: bijstelling van leerdoelen, informatie over de beheersing van de leerdoelen en enig begrip van de wijze waarop een eventueel gat kan worden overbrugd tussen de leerdoelen en de feitelijke beheersing ervan. Enige kennis over deze drie onderdelen is van groot belang om de leerprestaties te verbeteren. Dat geldt zowel voor de leraar als de voor de leerling.

Overigens leidde de publicatie van de studie van Black en Wiliam tot wijzigingen in het Engelse onderwijsbeleid op het gebied van evalueren op school, maar ook tot aanvullend onderzoek<sup>18</sup>. Zo leverden McDonald en Boud<sup>19</sup> een stevig bewijs voor het gunstige effect van het trainen van leerlingen in zelfevaluatie en peer-evaluatie op de examenresultaten. Ook werd aangetoond dat er een positief effect uitgaat van het geven van constructieve feedback op basis van formatieve evaluatie door leraren op de leerprestaties<sup>20</sup>. Uit een standaardwerk van Creemers en Kyrikiades<sup>21</sup> uit 2008 over effectief onderwijs blijkt dat voortgangsevaluaties op klassenniveau een belangrijke verklarende factor is voor de leerprestaties.

### **Besluit**

Onderzoek wijst uit dat toetsen en andere evaluatie-activiteiten samen te brengen zijn in een samenhangend evaluatie-systeem en dat dit een positieve bijdrage kan leveren aan de verhoging van de leerprestaties. Er zijn in Nederland voor het basisonderwijs al ruim voldoende gestandaardiseerde toetsen voor rekenen en taal beschikbaar. Als scholen in de gelegenheid worden gesteld hun vaardigheid in het juiste gebruik van deze toetsen te vergroten, dan levert dat een goede bijdrage aan de toetspraktijk. Maar het gaat niet alleen om toetsen, de gehele evaluatiepraktijk is in het geding. Vooral van de verdere ontwikkeling van de voortgangsevaluatie mag worden verwacht dat deze een positieve invloed heeft op de leerprestaties, omdat deze gericht is op de sturing van het onderwijsleerproces. De koppeling van de evaluatieresultaten aan didactische aanwijzingen en remediërende leerstof is hierbij van cruciaal belang.

In een evaluatiepraktijk waarin een goede balans bestaat tussen summatieve en formatieve evaluatie worden verschillende evaluatietechnieken op verschillende niveaus gebruikt en wordt er daardoor

rekening gehouden met de informatiebehoefte van alle betrokkenen. Als het systeem in balans is wordt op basis van dezelfde serie leerstandaarden geëvalueerd, maar wordt er wel onderscheid naar functie gemaakt. Zo verschaffen enerzijds de dagelijkse voortgangsevaluaties een constante stroom van informatie over de mate waarin leerlingen de stof beheersen waardoor op den duur goede prestaties op de leerstandaarden mogen worden verwacht. Anderzijds verschaffen gestandaardiseerde toetsen periodiek informatie over de vraag of de leerstandaarden bereikt zijn en wat er – mocht dat niet het geval zijn - eventueel op school- of groepsniveau moet gebeuren .

De onderwijsinspectie zou op haar beurt niet alleen oog moeten hebben voor de feitelijke tussen- en eindopbrengsten van scholen, maar vooral ook voor de mate waarin en de wijze waarop in scholen de evaluatiepraktijk in dienst staat van de voortgang van het leren, welke evaluatie-instrumenten en – procedures worden gebruikt, hoe 'data-driven' het onderwijs en Zorg & Begeleiding is, hoe de kwaliteit van de evaluatiepraktijk wordt geborgd, welke activiteiten ondernomen voor de verdere professionalisering ervan, etc.

In een gebalanceerd systeem leveren alle evaluatie-activiteiten informatie over de leerprestaties. We kunnen staat maken op de resultaatmetingen van gestandaardiseerde toetsen omdat deze op professionele wijze zijn ontwikkeld. Dat zou eigenlijk ook moeten gelden voor de andere technieken die onderdeel uitmaken van de evaluatiepraktijk. Maar het is dan wel noodzakelijk dat er wordt geïnvesteerd in de professionele ontwikkeling van leraren om beter te kunnen evalueren<sup>22</sup>. Gemiddeld besteedt een leraar ruim een kwart van zijn of haar tijd aan allerlei evaluatieactiviteiten. Als ze deze op een effectieve wijze zouden uitvoeren dan kunnen leerlingen daar alleen maar van profiteren. Doen ze dat niet, dan is het niet ondenkbaar dat dagelijks de vorderingen van leerlingen op een onjuiste wijze worden beoordeeld en dat beslissingen die leraren, scholen en ouders over leerlingen nemen gebaseerd zijn op onjuiste of ontoereikende informatie over de leervorderingen.

Al meer dan 20 jaar worden er wereldwijd pleidooien gehouden om leraren elementaire evaluatievaardigheden bij te brengen, niet alleen in het gebruik van toetsen, maar ook ten behoeve van de dagelijkse evaluatiepraktijk<sup>23</sup>. Al meer dan 20 jaar is ook bekend dat voortgangsevaluatie het beste werkt als deze is ingebed in het onderwijs en dus integraal deel uitmaakt van een systeem van leerstandaarden, bijbehorende (remediërende) leerstof en evaluatie-instrumenten om te signaleren, resultaten te bepalen en te diagnosticeren. De evaluatiepraktijk zal dus in samenhang met de ontwikkelingen in de onderwijspraktijk op reken- en taalgebied nader moeten worden bezien. De uitdaging is nu deze kennis om te zetten in activiteiten voor het reken- en taalonderwijs en in middelen om (aanstaande) leraren in staat te stellen zich evaluatievaardigheden eigen te maken. We hoeven het alleen nog maar te doen.....

- 1 Janssens, F.J.G. (1986). *De evaluatiepraktijken van leerkrachten: een onderzoek naar het evalueren tijdens het rekenen in het primair onderwijs*. Arnhem: Cito (diss.)
- 2 Dochy, F. & Janssens, S. (2003). Evaluatie en Assessment. In: Verloop, N. & Lowyck, J. (eds). *Onderwijskunde: een kennisbasis voor professionals*. Groningen: Wolters Noordhoff.
- 3 Wolf, I.F. De & Janssens, F.J.G. (2007). Effects and side effects of school inspections and accountability in education: a review of empirical studies. *Oxford Review of Education*. 33(3), 379-396.
- 4 Inspectie van het Onderwijs (2007). *De staat van het onderwijs: Onderwijsverslag 2005/2006*. Utrecht: Inspectie van het Onderwijs, p. 132.
- 5 Margaret E. Raymond & Eric A. Hanushek (2003), High-Stakes Research, *Education Next*, Spring 2003, [www.educationnext.org/20033/48.html](http://www.educationnext.org/20033/48.html). Maar zie ook eerdere studies als: Crooks, T., Kane, M. T. & Cohen, A. S. (1996). Threats to the valid use of assessment, *Assessment in Education*, 3(3), 265–286; Airasian, P. W. & Gregory, K. D. (1997). The Education Reform Act of 1998 (Review essay), *Assessment in Education*, 4(2), 307–314; Firestone, W. A. (1998) A tale of two tests: tensions in assessment policy, *Assessment in Education*, 5(2), 175–191.
- 6 Audrey L. Amrein & David C. Berliner (2002). High-Stakes Testing, Uncertainty, and Student Learning, *Education Policy Analysis Archives*, 10 (18). <http://epaa.asu.edu/epaa/v10n18/>. Zie ook Morrison, K. & Tang Fun Hei, J. (2002) Testing to destruction: a problem in a small state, *Assessment in Education*, 9(3), 289–317; Harlen, W. & Deakin-Crick, R. (2003) Testing and motivation for learning, *Assessment in Education*, 10(2), 169–208; Hong, W.P. & Youngs, P. (2008). *Does high-stakes testing increase cultural capital among low-income students and racial minority students?* Education Policy Analysis Archives. 16(6), 1-28.
- 7 Robert Linn, een autoriteit op het gebied van accountability merkte op: *As someone who has spent his entire career doing research, writing and thinking about educational testing and assessment issues, I would like to conclude by summarizing a compelling case showing that the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in dynamic ways. Unfortunately, that is not my conclusion.* (Robert Linn (2002), Assessments and Accountability. *Educational Researcher*, February 2000, p. 14.).  
Lorrie Shepard merkte in 2000 al op: *The negative effects of high-stakes testing on teaching and learning are well known. Under intense political pressure, testing scores are likely to go up without a corresponding improvement in student learning. In fact, distortions in what and how students are taught may actually decrease students' conceptual understanding.* (Lorrie A. Shepard, (2000). The Role of Assessment in a Learning Culture, *Educational Researcher*, October 2000, p. 9).
- 8 Zie bijvoorbeeld de beleidstheorie die ten grondslag ligt aan de Wet op het Onderwijstoezicht in Ehren, M.C.M. (2006). *Toezicht en schoolverbetering*. Enschede: Universiteit Twente (diss.)
- 9 Overigens zouden de uitkomsten van de Eindtoets Basisonderwijs wel in het voortgezet onderwijs gebruikt kunnen worden voor het bepalen van beginsituatie van individuele leerlingen.
- 10 De 'Assessment Reform Group,' bedacht hiervoor de term: 'Assessment for learning,' als tegenstelling voor 'assessment of learning'. (Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: School of Education, Cambridge University).  
Dochy en Janssens (2003) spreken in dit verband van de testcultuur versus de assessmentcultuur.
- 11 De Hey, W. (1961). *Hoe leer ik mijn kinderen kennen? Praktische handleiding voor de observatie ten behoeve van leerkrachten en jeugdleiders*. Haarlem: Uitgeverij De Toorts.
- 12 Rick Stiggins (2004), New Assessment Beliefs for a New School Mission, *Phi Delta Kappan*, Vol. 86, No. 1, p. 22-27.
- 13 Bloom, B. (1984). The Search for Methods of Group Instruction as Effective as One-on-One Tutoring, *Educational Leadership*, May 1984, pp. 4-17.
- 14 Black, P. & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment, *Phi Delta Kappan*, October 1998, p. 141. Zie ook Black, P. & Wiliam, D. (1998), Assessment and Classroom Learning, *Assessment in Education*, March 1998, pp. 7-74.

- 15 Black en Wiliam rapporteren dat zo'n effect, als het zou worden toegepast op de resultaten van de VS op de Third International Mathematics and Science Study (TIMSS), ertoe zou hebben geleid dat de VS van de middelste positie tussen de 42 deelnemende landen naar de top vijf zou zijn verschoven.
- 16 Zie bijvoorbeeld: Samuel Meisels et al.(2003), Creating a System of Accountability: The Impact of Instructional Assessment on Elementary Children's Achievement Scores, *Education Policy Analysis Archives*, vol. 11, no. 9, 2003, <http://epaa.asu.edu/eapp/v11n9>; Michael C. Rodriguez (2004). The Role of Classroom Assessment in Student Performance on TIMSS, *Applied Measurement in Education*, vol. 17, no. 1, 2004, pp.1-24.
- 17 Zie Dochy & Janssens, 2003. Black en Wiliam (1989, p. 143) merken over evaluatie door leerlingen zelf het volgende op: *Thus self-assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment. When anyone is trying to learn, feedback about the effort has three elements: redefinition of the desired goal, evidence about present position, and some understanding of a way to close the gap between the two. All three must be understood to some degree by anyone before he or she can take action to improve learning.*
- 18 Zie Broadfoot, P. & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11(1), p.7-26
- 19 McDonald, B. & Boud, D. (2003) The impact of self assessment on achievement: the effects of self assessment training on performance in external examinations, *Assessment in Education*, 10(2), 209–220.
- 20 Brookhart, S.M. (2004). Classroom assessment: tensions and intersections in theory and practice. *Teachers College Record*, 106 (3), 429-458
- 21 Creemers, B.P.M. & Kyrikiades, L. (2008). *The Dynamics of Educational Effectiveness: a contribution to policy practice and theory in contemporary schools*. Abingdon: Routledge.
- 22 Rick Stiggins (2004), New Assessment Beliefs for a New School Mission, *Phi Delta Kappan*, Vol. 86, No. 1, p. 22-27. Hij merkt op p. 26 op: *Teachers must possess and be ready to apply knowledge of sound classroom assessment practices. The typical teacher will spend one-quarter to one-third of his or her professional time involved in assessment-related activities. If teachers assess accurately and use the results effectively, then students prosper. If they do it poorly, student learning suffers. And it has. Therefore, the new belief must be that, without question, teachers need to know and understand the principles of sound assessment. The evidence of student learning they gather each day influences the most crucial instructional decisions. The remedy to our current situation is to offer targeted, productive professional development to put the available classroom assessment wisdom into the hands of practitioners.*
- 23 Zie onder meer Janssens, 1986; Stiggings, 2004; Broadfoot. & Black, 2004. Om nogmaals op het belang van een effectieve evaluatiepraktijk te wijzen deed the Committee on the Foundation of Assessment of the National research Council in 2001 de volgende aanbeveling: *Recommendation 9: Instruction in how students learn and how learning can be assessed should be a major component of teacher preservice and professional development programs. This training should be linked to actual experience in classrooms in assessing and interpreting the development of student competence. To ensure that this occurs, state and national standards for teacher licensure and program accreditation should include specific requirements focused on the proper integration of learning and assessment in teachers' educational experience.*(James W. Pellegrino, Naomi Chudowsky, and Robert Glaser, eds (2001), *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, D.C.: National Academy Press, 2001, p. 14.