

Een schets voor een adaptieve toets voor het funderend onderwijs

Lex Borghans en Trudie Schils

Kaans, Universiteit Maastricht

Maart 2011

Kaans: een voorbeeld van een schooloverstijgende toets in het VO

Kaans is een initiatief van de Universiteit Maastricht in samenwerking met leraren, scholen, schoolbesturen en de lokale en landelijke overheid. Doel is bevordering van de schoolprestaties van leerlingen door gegevens te verzamelen over hun achtergrond, de inzet in het onderwijsproces en de ontwikkeling van leerlingen van de vroege kinderjaren tot het einde van de middelbare school. De gegevens worden gebruikt als feedback voor leerlingen en leraren om daarmee de prestaties te verhogen. Ze worden ook gebruikt door het management van de school om de prestaties van de school te volgen als basis voor hun beleid. Tot slot worden de gegevens benut door wetenschappers en beleidsmakers om de effecten van onderwijsbeleid en onderwijsmethodes te meten.

In dat kader zijn er in 2010 een aantal toetsen gehouden onder leerlingen in het derde jaar van het voortgezet onderwijs waaraan vrijwel alle scholen in Limburg deel hebben genomen (circa 8.000 leerlingen in het onderzoek). Deze toetsen maken het mogelijk om een uitgebreid beeld te krijgen van het niveau in begrijpend lezen en wiskunde. Scholen kunnen zich onderling vergelijken en er ontstaat een beeld over de situatie in deze doorstroomvakken in de provincie. Door de toetsresultaten te relateren aan informatie uit achtergrondsurveys bij leerlingen en ouders en door ze te relateren aan informatie van de basisschool kunnen de gegevens worden benut om determinanten van studiesucces vast te stellen. Via rapportages aan schoolbesturen en scholen wordt bijgedragen aan een proces van opbrengstgericht werken in het onderwijs. Doel van Kaans is ook om de manier waarop de gegevens door betrokkenen worden benut te volgen zodat duidelijk wordt waar eventuele knelpunten liggen bij opbrengstgericht werken en waar nodig de informatieverzameling kan worden verbeterd. Een kritische houding ten aanzien van de afgenomen toets en hoe deze te verbeteren past bij dat doel.

Deze toetsen zijn in onze ogen een eerste stap naar een hedendaags vormgegeven uniforme meting in het onderwijs die op allerlei wijzen kan bijdragen aan de kwaliteitsverbetering van het onderwijs. Vaak wordt over een toets nagedacht alsof deze maar één specifieke functie heeft (bij de CITO-eindtoets bijvoorbeeld de doorverwijzing naar het VO). In de praktijk breidt het aantal gebruiksdoelen zich meestal gaandeweg uit, maar het zou goed zijn al bij het ontwerp rekening te houden met deze brede bruikbaarheid. Juist doordat een goede toets op vele fronten belangrijke feedback aan het onderwijssysteem kan leveren, kan de waarde er van zeer groot zijn. Een toets voor het VO laat leerlingen zien waar ze staan – waarbij aan de ene kant iedereen moet kunnen zien in hoeverre hij aan bepaalde normen voldoet, maar ook kennis en inzicht dat ver uitstijgt boven deze minimumnormen erkend moet worden –, en maakt een vergelijking tussen scholen en groepen leerlingen mogelijk. Tevens kan de VO-toets benut worden om de effecten van basisscholen op langere termijn te vergelijken, de doorverwijzing op basis van de CITO-toets te evalueren, onderwijsmethodes met elkaar te vergelijken en innovaties en verbeteringen die op scholen worden doorgevoerd te evalueren.

In deze notitie beschrijven we de toetsen die in 2010 in het kader van Kaans zijn uitgevoerd en geven we eens schets van een ideaal dat ons voor ogen staat.

Inventaar: Toetsen in 3VO

Om er voor te zorgen dat alle leerlingen vragen krijgen die passen bij hun niveau varieert de wiskundetoets naar schooltype (de leestoets is vooralsnog gelijk voor alle niveaus). Er is gekozen voor de volgende onderverdeling:

1. Toets voor de niveaus VMBO-BL en -KL
2. Toets voor de niveaus VMBO-GL en -TL
3. Toets voor de niveaus HAVO en VWO

De vragen zijn vervolgens gedeeltelijk overlappend op de verschillende schooltypen aangeboden. De overlap van vragen maakt het mogelijk om leerlingen van de verschillende schooltypen met elkaar te vergelijken. Deze informatie is belangrijk in het kader van het opsporen van onder- of overpresterende leerlingen. Door de overlap wordt het duidelijk hoe de beste VMBO-leerlingen zich verhouden tot de laagstscorende Havisten.

De toetsen zijn volledig digitaal afgenomen. Dit heeft een aantal voordelen: tijd- en kostenbesparing omdat niet alle gegevens nadien ingevoerd dienen te worden, foutenreductie omdat er bij handmatige, maar ook bij computergestuurde overname van schriftelijke toetsen fouten gemaakt kunnen worden en milieubesparing omdat niet alle toetsen op papier gedrukt hoeven te worden. Een veelgehoord nadeel van de digitale afname is dat deze afhankelijk wordt van de infrastructuur op de scholen. Toch blijkt dit probleem in de praktijk sterk mee te vallen. Slechts een enkele school heeft het onderzoek moeten laten schieten omdat het netwerk niet voldeed. In andere gevallen was er sprake van onvoldoende computers met daardoor roostertechnische problemen. Dit werd echter opgelost door de toetsen over een langere periode te spreiden. Een ander veelgehoord nadeel van een digitale toets is dat de leerlingen alles van het scherm moeten lezen, maar leerlingen in het voortgezet onderwijs zijn hier inmiddels aan gewend.

De vragen die in de toetsen zijn opgenomen komen uit:

- het nationale COOL-onderzoek 2007/2008 (Cohort Onderzoek Onderwijs Loopbanen)¹;
- het internationale PISA-onderzoek 2003 (OECD Programme for International Student Assessment)²;
- het Vlaamse SFS-onderzoek 2006/2010 van de Katholieke Universiteit Leuven³.

De PISA vragen zijn daarbij overigens niet gedifferentieerd naar de verschillende schooltypen, de vragen uit de andere bronnen wel. De keuze voor de vragen uit de drie verschillende bronnen maakt het mogelijk de resultaten met deze drie onderzoeken te vergelijken en te zien waar de Limburgse leerlingen staan ten opzichte van de Nederlandse, Vlaamse en internationale leerlingen en hoe de verschillende accenten die in de verschillende contexten gelegd worden op wat belangrijk is bij wiskunde, van invloed zijn op de score.

Er wordt momenteel gewerkt aan de rapportage over de toetsresultaten, een evaluatie van de vormgeving en de bruikbaarheid van de toetsen voor de praktijk. De belangrijkste eerste bevindingen ten aanzien van de prestaties van de leerlingen in relatie tot de bron van de vragen zijn:

- Een vergelijking tussen de Limburgse en Nederlandse leerlingen op basis van de COOL- en PISA-vragen laat zien dat de Limburgse leerlingen in het VMBO-BL het gemiddeld iets beter doen op de wiskundetoets dan ‘de Nederlandse leerlingen als geheel’. Daarentegen doen de Limburgse leerlingen

¹ Voor meer informatie, zie <http://www.cool5-18.nl>.

² Voor meer informatie, zie <http://www.pisa.oecd.org>.

³ Voor meer informatie, zie <http://ppw.kuleuven.be/schoolfeedback/>

in het HAVO en VWO het over het algemeen iets slechter dan de Nederlandse leerlingen. Interessant is dat het per vraag verschilt of de Limburgse leerlingen beter of slechter scoren dan ‘de Nederlandse leerlingen als geheel’. Dit verdient nader onderzoek.

- De vergelijking op basis van de vragen uit de Vlaamse toets laat zien dat het per vraag behoorlijk verschilt of de Limburgse leerlingen het slechter doen dan de Vlaamse leerlingen of niet. Bijvoorbeeld op de vraag “Bepaal de richtingscoëfficiënt van de rechte lijn met vergelijking $2y - (2-x)6 = 0$ ” weet 11 procent van de Vlaamse leerlingen op een niveau vergelijkbaar met ons VWO het goede antwoord, maar slechts 2 procent van onze VWO-ers die deze vraag heeft beantwoord. Een aantal leerlingen merkte op dat de term richtingscoëfficiënt hen volledig onbekend is. Op een tweetal meetkundige vragen (inhoud kubus en lengte zijde driehoek) scoren de Limburgse VWO-ers opvallend beter dan de Vlaamse leerlingen.

Daarnaast komen interessante verschillen aan het licht waaruit blijkt dat bepaalde groepen leerlingen ondanks vergelijkbare CITO-scores toch heel verschillend scoren op de toetsen van Inventaar. Dit hangt samen met de basisschool die men heeft gevolgd, maar ook met persoonskenmerken. Verschillen in scores op de toets hangen overigens niet alleen samen met hoeveel een leerling weet, maar ook met hoe men een toets maakt. Afhankelijk van het ontwerp van een toets en het soort vragen dat wordt gesteld scoren bepaalde leerlingen hoger en andere lager. Uit onze analyse van PISA blijkt bijvoorbeeld dat de prestaties van leerlingen behoorlijk afnemen naarmate de toets vordert⁴. Deze afname verschilt sterk tussen groepen (ook tussen landen) en hangt nauwelijks samen met de kennis en vaardigheid die de test beoogt te meten. In de volgende ronde van Inventaar willen we – door een subtiele opzet van de toetsen – zicht krijgen op deze invloeden. Verder is het streven om in samenspraak met de scholen de volgende toets onderdeel uit te laten maken van de reguliere toetsing door de scholen. Op die wijze stijgt de bruikbaarheid van de informatie voor de scholen, terwijl bovendien bekend is dat leerlingen toetsen op een andere manier maken als de toets er voor hen toe doet. Dit vergroot de validiteit van de toets. De toets in ons onderzoek had, net als toetsen als PISA en COOL, geen direct belang voor de leerling en waarom zou hij zijn best doen? In de volgende ronde zal ook een expliciete link met de referentieniveaus in de toetsen worden opgenomen. Deze toetsen zullen tevens worden ingezet bij de evaluatie van onderwijsexperimenten die we in samenwerking met de scholen uitvoeren.

Om de doelstelling te realiseren van gezamenlijke toetsen die benut kunnen worden voor opbrengstgericht werken van scholen, die leerlingen een beeld geven van hun positie ten opzichte van de referentieniveaus en die gebruikt kunnen worden bij de evaluatie van onderwijsinitiatieven en de verdere ontwikkeling van kwalitatief hoogstaand onderwijs is het streven op termijn te komen tot een adaptieve toets.

Een schets voor een adaptieve toets voor het funderend onderwijs

Inleiding

Toetsen vormen een essentieel onderdeel van het onderwijs, omdat ze alle betrokkenen bij het onderwijs cruciale informatie geven over hoe het er voor staat en ze op basis van deze informatie verdere beslissingen kunnen nemen. Uiteraard wordt er daarom veel getoetst in het onderwijs. Docenten houden met regelmaat een proefwerk, scholen hebben proefwerkweken en ouders overhoren hun kinderen. Landelijke toetsen kunnen een belangrijke aanvulling op de bestaande toetsen zijn. Het is van groot belang

⁴ Borghans en Schils (2011), *The leaning tower of PISA: Disentangling ability and motivation*, ongepubliceerd manuscript, Maastricht University.

dat ook landelijke toetsen voor een leerling ertoe doen en zo goed mogelijk worden geconstrueerd om dat te meten wat men wil meten. In deze notitie gaan we in op de vraag hoe een dergelijke landelijke toets vorm zou kunnen worden gegeven en wat de toegevoegde waarde is in vergelijking met de bestaande toetspraktijk. Daarbij maken we gebruik van de ervaringen die we hebben opgedaan bij de bovenstaand beschreven toets.

Goede toets

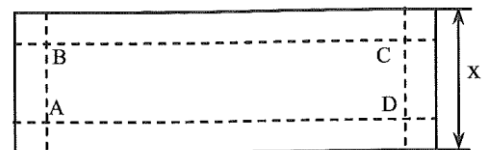
Een goede toets test leerlingen met vragen die op de grens liggen van wat hij kan en wat hij niet kan en maakt een vergelijking tussen eerdere toetsen, andere leerlingen en eerdere cohorten mogelijk. Een ouder die zijn kind overhoort zal altijd doorvragen op onderdelen waarvan hij merkt dat zijn kind er moeite mee heeft en niet eindeloos vragen stellen die het kind toch wel weet of die het kind echt niet zal kunnen beantwoorden. De vraag of het kind beter is geworden dan de vorige keer kan hij op het gevoel beantwoorden. Als verschillende leerlingen van verschillende leraren of verschillende scholen vergeleken moeten worden is dat niet meer mogelijk. Traditionele toetsen zijn zo opgesteld om dit soort vergelijkingen toch te maken. Een traditionele toets stelt daarom vragen die ruwweg het spectrum van de capaciteiten van de leerlingen in de doelgroep afdekken. De toets is in feite gebaseerd op de gemiddelde leerling. Sommige leerlingen krijgen enkele vragen die veel te makkelijk zijn, anderen krijgen vragen die voor hen veel te moeilijk zijn. De exacte positie van de leerling in de leerlingpopulatie kan niet worden bepaald, er kan slechts een schatting worden gemaakt. Bovendien werken de te moeilijke/makkelijke vragen demotiverend voor de leerlingen waardoor de prestaties ook beïnvloed kunnen worden.

Dankzij de mogelijkheden van ICT bieden adaptieve toetsen de mogelijkheid om ook bij formele toetsen leerlingen vragen voor te leggen die passen bij het niveau dat ze hebben, dat niveau ook preciezer te bepalen, en in te zoomen op zwakke plekken in hun kennis. Adaptieve toetsen kunnen daarom zowel voor de zwakke leerlingen en de leerlingen die heel erg goed zijn een stimulerende rol hebben en veel gericht feedback geven op het niveau van de leerling, leraar, school en het onderwijssysteem als geheel over de sterke en de zwakke punten. Daarnaast biedt de informatie die beschikbaar komt met een goed vormgegeven adaptieve toets ongekende mogelijkheden voor verdere ontwikkeling van de diagnostische en voorspellende kracht van de toets en kan deze informatie gebruikt worden voor de verbetering van onderwijsmateriaal en -methodes en onderwijsprogramma's.

Een voorbeeld is de volgende vraag uit een wiskundetoets:

Van een rechthoekig stuk karton, met lengte driemaal de breedte, knippen we uit elke hoek een vierkant met zijde 2 eenheden. Met dit karton maken we een open, balkvormige doos met rechthoek ABCD als grondvlak.

Druk de inhoud I van deze open doos uit in een functie van x .



Slechts 0,2% van alle VWO'ers weet deze vraag goed te beantwoorden. Voor een beoordelende eindtoets waarbij het louter om het al dan niet behalen van een minimumnorm gaat zou het daarom inefficiënt zijn deze vraag op te nemen, maar voor de kleine groep leerlingen die dit niveau wel haalt is de vraag een enorm mooie uitdaging. Als ze deze vraag de ene keer niet weten te beantwoorden, willen ze wellicht de volgende keer wel dit niveau halen. Bovendien brengt deze vraag in kaart welke leerlingen tot deze subgroep behoren.

Een andere voorbeeld is de volgende wiskundevraag :

Vanochtend was het $-3\text{ }^{\circ}\text{C}$. Vanmiddag was het $12\text{ }^{\circ}\text{C}$. Hoeveel is de temperatuur gestegen?

- | | |
|--------------------------|-------------------------------|
| <input type="checkbox"/> | $-15\text{ }^{\circ}\text{C}$ |
| <input type="checkbox"/> | $-9\text{ }^{\circ}\text{C}$ |
| <input type="checkbox"/> | $9\text{ }^{\circ}\text{C}$ |
| <input type="checkbox"/> | $15\text{ }^{\circ}\text{C}$ |

Slecht 8% van de VMBO'ers heeft deze vraag fout, maar ook voor deze groep kan een dergelijke vraag een belangrijke functie vervullen. Wat belemmert deze leerlingen in het geven van het goede antwoord? Bij dit voorbeeld kan het bijvoorbeeld zinvol zijn te kijken naar het meest gegeven antwoord door degenen die het fout deden en of daar een patroon te ontdekken is.

Een adaptieve toets bestaat uit een zeer grote verzameling vragen met een uiteenlopende moeilijkheidsgraad en van verschillende typen. De test wordt digitaal afgenomen. In eerste instantie krijgt een leerling vragen met een gemiddelde moeilijkheidsgraad of met een moeilijkheidsgraad die past bij het niveau dat van deze leerling op basis van eerdere toetsen verwacht kan worden. Afhankelijk van zijn antwoorden selecteert de computer vervolgens moeilijkere of juist makkelijkere vragen om zo de moeilijkheidsgraad van de leerling aan te passen aan zijn niveau. Als een leerling op bepaalde onderdelen opvallend slecht scoort – gezien zijn algemene niveau – kunnen op dat terrein extra vragen worden gesteld om vast te stellen of er inderdaad sprake is van een kennislacune. Inzicht in de specifieke type vragen waarin leerlingen goed of juist niet goed zijn, is ook van verder belang voor de toekomst van de leerling. Op het MBO en HBO wordt steeds meer gewerkt met competentieprofielen en het zou leerlingen helpen als ze al vroeg weten dat ze bijvoorbeeld niet goed zijn in creatieve wiskundevragen, maar bijvoorbeeld wel in standaard rekenwerk.

Door het opnemen van een grote hoeveelheid vragen (waarvan iedere leerling er slechts een deel krijgt) ontstaat er op school- en systeemniveau een veel gedetailleerder beeld van hoe goed leerlingen nu precies zijn en waar leerlingen goed in zijn en waarin niet. Leerlingen op de ene school, of leerlingen die een bepaalde methode gebruiken kunnen beter of juist minder goed zijn in een specifiek facet van wiskunde of taal. Dergelijke informatie vormt daarom belangrijke feedback voor scholen en onderwijsmethodes. Ook blijkt uit onderzoek dat leerlingen afhankelijk van hun persoonlijkheid op bepaalde vragen beter scoren dan op andere. Sommige leerlingen zijn goed in routinevragen, anderen zijn vooral goed als er creativiteit wordt gevraagd etcetera. Ook hier geldt dat door een grote hoeveelheid toetsvragen te gebruiken duidelijk kan worden of een school meer de routinekant of juist de creatieve kant van het vak heeft bevorderd.

Om leerlingen van verschillende schooltypen ook onderling goed te kunnen vergelijken zou in iedere toets een aantal vragen gesteld moeten worden uit een groep vragen van gemiddeld niveau. Ook bij leerlingen die heel goed zijn in wiskunde moet getest worden of ze ook hun elementaire rekenvaardigheden op peil hebben.

Momenteel worden toetsvragen vastgesteld door deskundigen die een inschatting maken van de moeilijkheidsgraad of hoogstens voorgelegd aan een kleine groep testleerlingen om daarmee de moeilijkheidsgraad vast te stellen. Het is bekend dat de resultaten van een toets sterk afhangen van de omstandigheden waaronder deze wordt afgenomen. Leerlingen scoren veel beter als er wat op het spel staat bij een toets. De testpanels geven daarom geen goed beeld van de moeilijkheidsgraad van een vraag onder andere omstandigheden. Bij een adaptieve toets kunnen, naast de bestaande vragen, ook nieuwe

vragen worden opgenomen in de reguliere toets. Door deze vragen zorgvuldig te verdelen krijgt elke leerling slechts enkele nieuwe vragen, maar groeit de voorraad vragen waarvan een moeilijkheidsgraad bekend is. Eventueel kunnen de antwoorden op deze vragen achteraf niet worden meegenomen bij de bepaling van de eindscore.

Testen testen

Toetsresultaten blijken sterk af te hangen van de manier waarop ze worden afgenomen. Daarbij moet niet alleen gedacht worden aan of de toets ertoe doet voor de leerling, maar bijvoorbeeld ook aan de tijdsduur die voor de toets beschikbaar is. Een adaptieve toets maakt het mogelijk te analyseren hoe de afnamecondities de resultaten beïnvloeden om zo de toets op termijn te optimaliseren. Ook als de resultaten van de toets in eerste instantie af blijken te hangen van de specifieke afnamecondities kan op verantwoorde wijze een vergelijkbaar resultaat voor iedere leerling worden gepresenteerd. Op macroniveau is immers bekend hoe de afnameconditie het resultaat heeft beïnvloedt.

Vergelijkbaarheid

Bij een goed opgezette toets is vergelijkbaarheid een cruciale factor. Resultaten moeten niet alleen een vergelijking geven tussen de leerlingen die in eenzelfde jaar aan de toets hebben meegedaan, maar er moeten ook vergelijkingen mogelijk zijn tussen scholen en onderwijsniveaus. Ook moeten de toetsresultaten over de tijd vergelijkbaar zijn. Het zou bovendien wenselijk zijn als een relatie kan worden gelegd met internationale metingen zoals PISA.

Vergelijkbaarheid in de tijd kent twee grote problemen. In de eerste plaats zal de opzet en de inhoud van een toets in de loop van de tijd altijd wijzigen. Een adaptieve toets maakt het mogelijk deze vergelijkbaarheid te garanderen door bij veranderingen gedurende enkele jaren zowel vragen uit het oude als vragen uit het nieuwe programma op te nemen. Ook curriculumoverstijgende vragen zijn daarom altijd gewenst. Bovendien ontstaat daarmee een beeld van de kennis en vaardigheden die leerlingen buiten het beoogde curriculum hebben weten te vergaren. Door ook – zoals we in Inventaar hebben gedaan – ook vragen uit PISA en de Vlaamse toets op te nemen ontstaat een beeld van de positie van Nederlandse leerlingen ten opzichte van het buitenland.

Toetsresultaten zijn sterk afhankelijk van de leeftijd waarop ze worden afgenomen. Zelfs los van de hoeveelheid onderwijs die leerlingen krijgen gaat men het altijd beter doen als men ouder is. Dit levert een probleem op bij toetsen die in een bepaald schooljaar worden afgenomen. De zittenblijvers in deze klas zijn immers ouder en hebben ook meer onderwijs gehad dan hun klasgenoten. Als gevolg hiervan zijn de scores voor scholen met veel zittenblijvers veel gunstiger dan de scores van scholen met weinig zittenblijvers. Ook voor de vergelijkbaarheid van de prestaties van Nederlandse leerlingen over de tijd, kan een probleem ontstaan als door een stelselwijziging, of door veranderende gebruiken het percentage zittenblijvers toe of af zou nemen. Beleid dat de doorstroom in het onderwijs bevordert zou daardoor ten onrechte tot lagere testcores leiden. Los van de eventuele logistieke problemen zou het daarom de voorkeur hebben om – net zoals dit bij PISA wordt gedaan-, alle leerlingen van een bepaalde leeftijd te toetsen. Het probleem dat hiermee de zittenblijvers geen informatie meer krijgen over waar zij staan aan het eind van de tweede of de derde klas kan worden ondervangen door toe te staan dat scholen deze leerlingen nogmaals laten deelnemen aan de landelijke toets, zonder dat deze gegevens benut worden voor de landelijke analyses.

Om de voortgang van leerlingen te kunnen volgen zou op termijn een jaarlijkse toetsafname uiteraard het mooiste zijn. Een goed-ontwikkelde adaptieve toets zou voor alle leerjaren geschikt kunnen zijn. Leerlingen kunnen dan perfect hun progressie volgen en juist de excellente leerlingen kunnen kijken hoe ver ze al zijn in vergelijking met leerlingen in volgende leerjaren. In zo'n geval hoeft het onderscheid tussen leeftijd en klas ook niet meer gemaakt te worden.

Innovatie

Gegevens die beschikbaar komen door landelijke toetsen zijn van grote waarde voor talloze betrokkenen bij het onderwijs. Leerlingen zien op basis van de scores hoe ver ze zijn, maar vooral ook waar hun sterke en hun zwakke punten zitten zodat ze op grond hiervan hun studie-inspanningen aan kunnen passen. Ook komen eventueel lacunes aan het licht. Onvoldoende beheersing van bepaalde onderdelen uit eerdere jaren kan remmend werken op de studievoortgang. Ook docenten, scholen en schoolbesturen zien hoe goed ze het doen in vergelijking met andere scholen. Ook hier gaat het weer niet alleen om een algemeen prestatieverschil, maar is ook de vergelijking op onderdelen van belang. Hiervoor is het van belang dat niet – zoals nu vaak het geval is – alleen totaalscores beschikbaar komen, maar de betrokkenen ook inzicht krijgen in welk soort vragen een leerling, docent, school relatief goed of slecht scoort. Als toetsgegevens gekoppeld kunnen worden aan informatie uit eerdere toetsen, het schoolvolgsysteem en gegevens over de achtergrondkenmerken van de leerling, kan deze vergelijking aan kwaliteit winnen.

Ook voor de vergelijking van onderwijsmethodes en de evaluatie van veranderingen in de onderwijsaanpak is de toetsinformatie zeer waardevol. Het is te verwachten dat nieuwe toepassingen worden ontwikkeld als de informatie beschikbaar is voor onderzoekers en ontwikkelaars. Onlangs heeft de commissie de Graaf zijn zorgen uitgesproken over de link tussen onderwijsonderzoek en de onderwijspraktijk. Het beschikbaar stellen van goede gegevens is een eenvoudig instrument om de betrokkenheid van onderzoekers op de praktijk te vergroten. Open innovatie kan veel voor de kwaliteit van het Nederlandse onderwijs betekenen.

Ten slotte hoort ook de toets zelf kritisch te worden bekeken. Daar alle informatie van de toets beschikbaar te stellen voor onderzoekers wordt de manier van toetsen en de mogelijke neveneffecten die dit heeft voor het onderwijs opengesteld voor wetenschappelijke kritiek. Voor degenen die een toets ontwikkelen wordt dit vaak als lastig ervaren maar uiteindelijk profiteert iedereen van zo'n open benadering. Door niet alleen totaalscores maar ook de onderliggende gegevens zoals antwoorden per vragen, denktijd per vraag etcetera beschikbaar te stellen, kan in wetenschappelijke concurrentie gezocht worden naar verbeteringen van de toets en naar betere methodes om op basis van de gegevens de juiste diagnose te stellen.

Nationale Onderwijstoets⁵

Een nationale toets biedt een ideaal moment om de prestaties van het Nederlandse onderwijs publiekelijk onder de aandacht te brengen. Onderwijs is de laatste tijd vaak negatief in het nieuws. Door vlak na de landelijke toets de resultaten naar buiten te brengen en daarbij – met wellicht ieder jaar weer een andere invalshoek – leerlingen of scholen die goed scoren in het zonnetje te zetten kan ook duidelijk worden gemaakt dat ieder jaar weer vele leerlingen met hun docenten een geweldige prestatie neerzetten.

⁵ Zie ook Borghans en anderen (2007), "Parlementair onderzoek onderwijsvernieuwingen: het meten van onderwijskwaliteit en de effecten van recente onderwijsvernieuwingen, Tweede Kamer, vergaderjaar 2007-2008 31 2007 nr. 9.