

Beoordelen in het teken van een leven lang leren

Dr. D.M.A. Sluijsmans
(dominique.sluijsmans@han.nl)
is als lector werkzaam bij de
Faculteit Educatie van de
Hogeschool van Arnhem en
Nijmegen. Haar lectoraat is
getiteld 'Duurzaam beoordelen in
vraaggestuurd onderwijs'.

Door de opkomst van competentiegericht leren is het thema 'beoordelen' sinds enkele jaren nadrukkelijker op de innovatie-agenda gezet. Gezien de sturende kracht die beoordelen heeft op het studeergedrag van studenten, hebben opleidingen leren en beoordelen dichter bij elkaar gebracht. Dit heeft geresulteerd in een generieke toepassing van portfolio's en andere proeven van bekwaamheid, met het doel studenten actief te betrekken bij het aantonen van hun competentieontwikkeling. De implementatie van deze nieuwe benadering van beoordelen verloopt echter niet zonder slag of stoot. Wanneer we kijken naar de dagelijkse onderwijspraktijk, is nog niet echt sprake van een volledige integratie van leren en beoordelen. Sterker nog, studenten zien toch nog steeds beoordelen – zelfs het portfolio – als het sluitstuk van onderwijs, in plaats van een rijke bron voor verder leren. Als gevolg van de redenering 'De voldoende is binnen, dus het leren is klaar' is het beklijven van kennis en kunde op lange termijn een illusie. In dit artikel betoog ik dat het roer om moet als het gaat om de functie van beoordelen. Ik introduceer 'duurzaam beoordelen' als dé wijze van beoordelen die past bij leren voor de toekomst. Een goed ontwerp van beoordelen, een heldere communicatie over beoordelen en een zelfsturende rol van de student in beoordelen zie ik daarbij als de drie belangrijkste speerpunten. Met deze bijdrage hoop ik bij opleidingen een herbezinning op het 'waarom' van beoordelen te stimuleren.

INLEIDING

Beoordelen is een fascinerend fenomeen. Toen ik eens op een middag mijn 'beoordelingsverleden' op een rijtje zette, kwam ik tot de schrikbarende conclusie dat ik in mijn leven ongeveer 4200 keer ben beoordeeld. Een ruwe maar realistische schatting van het aantal overhoringen, proefwerken, (her)tentamens, rijexamens, Cito-toetsen en muziekexamens. De schrik zat hem niet zozeer in de hoogte van het aantal, maar vooral in het feit dat deze beoordelingen in mijn beleving nauwelijks van meerwaarde zijn geweest voor mijn leren of hebben geleid tot het beklijven van kennis of kunde. Sterker nog, het gevoel overheerste dat ik met al die beoordelingen vooral mijn tijd aan het verdoen was. Ook vond 90% van deze beoordelingen plaats tussen mijn 4^e en 23^e levensjaar; een indicatie dat beoordelen een piek heeft in je formele onderwijstijd en daarna beperkt blijft tot bijvoorbeeld een resultaat- en ontwikkelingsgesprek.

Tot op de dag van vandaag lijken leren en beoordelen weinig met elkaar van doen te hebben. Uit onderzoek en praktijk blijkt dat beoordelen in de perceptie van studenten

niks te maken heeft met (langetermijn)leren (Boud & Falchikov, 2006). Beoordelen wordt vooral gezien als het sluitstuk van een onderwijsfase, in plaats van een rijke bron voor verder leren. Huidige beoordelingen, zowel summatief als formatief, zijn nog vaak gericht op de korte termijn, waardoor in onvoldoende mate kan worden vastgesteld of de beoordeling de student bagage meegeeft voor de toekomst. Om beoordelen meer te benutten in het kader van leren op de lange termijn, introduceer ik graag het duurzaam beoordelen, dat in het buitenland bekendstaat onder de noemer *sustainable assessment* (Boud, 2000). Duurzaam beoordelen betekent drie dingen. Op de eerste plaats gaat het om beoordelen dat het beklijven van kennis en kunde waarborgt. Ten tweede is bij duurzaam beoordelen een centrale rol weggelegd voor de student, van wie steeds meer wordt verwacht dat hij/zij zelfsturend leert. Ten derde worden in duurzaam beoordelen de huidige leervragen van de student erkend door verder leren te stimuleren op individueel niveau. Behalve dat duurzaam beoordelen goed past in een maatschappij waarin duurzaamheid het toverwoord lijkt te zijn, sluit het ook goed aan bij 'een leven lang leren', een thema dat zowel nationaal als internationaal een prominente plaats inneemt op de maatschappelijke agenda. In deze bijdrage geef ik een eerste invulling aan duurzaam beoordelen aan de hand van drie speerpunten:

1. het ontwerpen van beoordelen;
2. het communiceren over beoordelen;
3. het zelfsturen bij beoordelen.

ONTWERPEN VAN BEOORDELEN

Het ontwerp van onderwijs is als de bouwtekening van de architect. Zoals een huis niet gebouwd kan worden zonder tekening, zo kan een onderwijsprogramma niet worden opgetuigd zonder een solide onderwijsontwerp. Als het gaat om het ontwerp van beoordelen, is er zelden sprake van een ontwerp gericht op duurzaamheid. Dit heeft drie redenen.

Een eerste reden is de beperkte functie van de huidige beoordelingsmodellen. In het 'normgerichte beoordelingsmodel' worden studenten afgezet tegen een gemiddelde van de groep. Deze vaak gestandaardiseerde beoordelingen veronderstellen uniforme studenten en het bestaan van een 'gemiddelde student'. Tevens wordt een soort 'eindigheid' in leren gesuggereerd. Voor diegenen die het beste zijn in hun groep, is er geen reden om verder te leren als de norm is behaald. Een ander bezwaar van dit model is dat het niveau van de groep bepalend is voor het individuele resultaat. Zit je als laag presterende student in een hoog presterende groep, dan is de kans op een voldoende klein. In het 'criteriumgerichte beoordelingsmodel', een model dat vooral wordt gehanteerd in het middelbaar en hoger beroepsonderwijs, worden studenten niet zozeer vergeleken met de norm van de groep, maar beoordeeld op een aantal beoordelingscriteria en bijbehorende standaarden. Resultaat van criteriumgerichte beoordelingen is dat sommige studenten de vastgestelde standaarden op de criteria bereiken en andere studenten niet of met moeite. Hoewel het behalen van standaarden natuurlijk prima en noodzakelijk is, staat dit model duurzaam beoordelen in de weg. Ook in dit model geldt

namelijk dat er geen reden is om verder te leren als de standaarden zijn behaald. In zowel het normgerichte als het criteriumgerichte beoordelingsmodel schuilt het gevaar van het bekende toetsgedrag (*wash-back effect*; Alderson & Wall, 1993). De boeken en het portfolio verdwijnen in de kast als de voldoende binnen is. De individuele student met zijn unieke profiel verdwijnt volledig in een groepsgemiddelde of wordt slechts summatief afgeinkt op een lijst van beoordelingscriteria, die niet op het niveau van de individuele student zijn afgestemd.

Een tweede reden is het gebrek aan helderheid over wat en op welk niveau wordt beoordeeld. Helder geformuleerde beoordelingscriteria en standaarden vormen nog te weinig de basis in het onderwijs. Ze worden meestal per vak door de individuele docent bepaald, vaak op een moment dat het onderwijs al 'achter de rug is' en er nog een beoordeling moet plaatsvinden. De beoordelingscriteria en standaarden zitten daarbij vooral in het hoofd van de docent (*tacit knowledge*; Sadler, 1989), waardoor de student niet kan begrijpen waaraan moet worden voldaan en zich dus ook niet adequaat kan voorbereiden. Uit onderzoek blijkt dan ook dat studenten nogal eens verrast worden door de inhoud van een beoordeling, omdat de verwachtingen van de student over de beoordelingscriteria niet overeenkomen met de verwachtingen van de docent (Broekamp & Van Hout-Wolters, 2007). Naast heldere beoordelingscriteria zijn ook heldere standaarden (= het moeilijkheidsniveau) nodig. Zo kan de standaard op bijvoorbeeld het criterium 'begeleiden van een leerling' vrij laag zijn (gemakkelijke leerstof, goed luisterende leerling) of hoog (complexe leerstof, leerling met gedragsproblemen). Standaarden zijn voor de student ook niet altijd helder.

Een derde reden is het gebrek aan aansluiting tussen leren en beoordelen; zichtbaar in de ontworpen verhouding tussen 'onderwijstijd' en 'beoordelingstijd'. Voorbeelden zijn de geclusterde proefwerkweken, (her)tentamens en de halfjaarlijkse portfoliobeoordelingen. Zolang onderwijstijd en beoordelingstijd zwak zijn geïntegreerd, zullen studenten zich vooral inspannen in de periodes vlak voor de beoordeling, met het bekende uitstelgedrag als gevolg (Crooks, 1988).

Naar het ontwerpen van duurzaam beoordelen

Gezien de geschetste oorzaken lukt het nog niet onderwijs zo te ontwerpen dat integratie van leren en duurzaam beoordelen volledig wordt gewaarborgd. Beoordelen wordt vooral als toetskundig probleem gezien, maar is eigenlijk een ontwerpprobleem (Van der Vleuten & Schuwirth, 2005). Om duurzaam beoordelen mogelijk te maken zou ik graag willen refereren aan de door De Vries (2007) aangekaarte noodzaak tot 'gedifferentieerd uitdagen'. Als we stellen dat studenten taken uitvoeren die passen bij hun voorkennis en motivatie, is het vanzelfsprekend om de beoordeling ook daarop aan te laten sluiten. Als we wel in het onderwijzen differentiëren, maar niet in het beoordelen, wordt het adaptieve dat in het onderwijs wordt opgebouwd volledig tenietgedaan door een beoordelingsmodel dat allesbehalve adaptief is. Studenten zullen bij gedifferentieerd leren en beoordelen niet meer op hetzelfde moment met dezelfde beoordelingscriteria en op hetzelfde niveau aan de slag gaan. Ze zullen een individueel leerpad volgen met een daarop aansluitende beoordeling. Om het probleem van de

bepaalde functie van de normgerichte en criterium-/standaardgerichte beoordelingsmodellen aan te pakken, stel ik een ipsatief beoordelingsmodel voor ('ipse' is Latijn voor 'zelf'). In dit model worden beoordelingen van studenten niet zozeer vergeleken met het gemiddelde van hun klas, of met vooraf vastgestelde criteria en standaarden, maar worden individuele beoordelingen, die zijn aangepast aan het niveau van de student, vergeleken in de tijd. Met andere woorden, de student wordt steeds vergeleken met zichzelf. Er is sprake van duurzaam beoordelen, omdat criteria en standaarden flexibel kunnen worden aangepast en de beoordelingen zich niet meer beperken tot een momentopname.

Om helderheid te creëren over beoordelingscriteria en standaarden is voor elk beoordelingsmodel helderheid over de zogenaamde gedragsdoelen noodzakelijk (Van Merriënboer & Kirschner, 2007). Gedragsdoelen worden afgeleid van de gewenste vaardigheden in een beroep of bepaald kennisdomein. Goed gedefinieerde gedragsdoelen beschrijven de condities waaronder een vaardigheid moet worden gedemonstreerd en de beoordelingscriteria en standaarden. De condities waaronder de student een taak uitvoert, bepalen in veel gevallen de standaard van een criterium. Vervolgens is een groot aantal beoordelingen met duidelijke beoordelingscriteria en standaarden als basis essentieel. Alleen dan kunnen resultaten op deze taken met elkaar worden vergeleken en is het ontwerp van een ipsatief beoordelingsmodel mogelijk. Juist voor complexere kennisgebieden en vaardigheden is het cruciaal de relevante beoordelingscriteria optimaal te integreren in betekenisvolle beoordelingstaken. In deze beoordelingstaken – die kunnen worden geschaard onder de noemer 'onderwijs' – worden leren en beoordelen geïntegreerd, omdat ze studenten aanzetten tot het werken aan gedragsdoelen die gericht zijn op het beklijven van kennis, vaardigheden en houdingen. Beoordelingstaken kunnen de student in vele vormen worden aangeboden, zoals opdrachten, situaties, problemen of projecten. Beoordelings- en onderwijstijd zijn door de integratie volledig in balans.

COMMUNICEREN OVER BEOORDELEN

Een tweede speerpunt dat voor duurzaam beoordelen essentieel is, richt zich op de communicatie over beoordelen. Ook hier zijn drie redenen te geven waarom duurzaam beoordelen wordt belemmerd.

Een eerste reden betreft de onheldere communicatie over het beoordelingsmodel in opleidingen. Hoewel studenten begrijpen dat zij moeten werken aan de opbouw van een portfolio en moeten werken aan beroepstaken, vinden ze het vaak lastig om te doorzien hoe de relatie is tussen de portfoliobeoordeling en andere vormen van beoordeling zoals kennistoetsen.

Een tweede reden is het gebrek aan zicht op het beoordelingsproces en de mens achter de beoordelaar. Over de feilbaarheid van de menselijke beoordelaar is al veel gepubliceerd (zie bijvoorbeeld De Groot, 1966). Vele studies hebben aangetoond dat docenten

met vergelijkbare deskundigheden dezelfde taak verschillend beoordelen. Tot op heden is het gebrek aan overeenstemming tussen beoordelaars een heikel punt. In het bijzonder als het gaat om certificerende beslissingen wordt de mens als beoordelaar in twijfel getrokken. Iedere docent hanteert zo zijn eigen 'beoordelingsregels' en '-strategieën', maar expliciteert deze helaas niet of nauwelijks (De Groot, 1966). Het beoordelingsproces voltrekt zich als het ware in een *black box*. Omdat beoordelen zich vooral in de hoofden van de individuele docenten voltrekt, is het lastig tot structurele veranderingen in (de visie op) beoordelen te komen.

Een derde reden betreft de informatie-armheid van beoordelingen. In beoordelingen wordt de aanvankelijk rijke informatie van beoordelaars vaak teruggeliterd tot een saai, simpel, nietszeggend cijfer; de student nauwelijks motiverend tot verder leren (Schuwirth, 2007). Ook wordt het cijfer op een moment verstrekt dat de student de beoordelingstaak nog maar nauwelijks voor de geest weet te halen. Niet alleen de beoordelingscriteria waren bij het verstrekken van de opdracht onvoldoende specifiek gecommuniceerd, ook is lang niet altijd sprake van specifieke en constructieve feedback. Hoewel onderzoek al meerdere malen heeft aangetoond dat cijfers de meest zwakke vorm van feedback zijn, zien we deze nog veel in het onderwijs. Omdat het becijferen zich vooral richt op het vaststellen van een prestatie op een bepaald tijdstip, zijn beoordelingen per definitie summatief. Beoordelingen die nu vaak als formatief worden bestempeld, krijgen die benaming vooral als het cijfer 'niet meetelt'. Zolang informatie ontbreekt over de betekenis van de beoordeling en het toekomstige leertraject dat de student zal doorlopen, is er geen sprake van duurzaam beoordelen (Yorke, 2003; Taras, 2005).

Naar rijke communicatie over beoordelen

Het is zinvol dat scholen op een toegankelijke manier helder maken welk beoordelingsmodel of welke mix van beoordelingsmodellen (formatief, summatief, duurzaam) wordt gehanteerd. Als leren en beoordelen worden geïntegreerd, is er geen noodzaak beoordelingsmodellen in separate documenten vast te leggen. Volstaan kan worden met één document, waarin de integratie van leren en beoordelen wordt onderbouwd en daarmee de visie op en kwaliteitsborging van leren. Ook op het niveau van de docenten is het van belang het beoordelingsmodel helder te communiceren, zodat er meer synergie ontstaat tussen de beoordelingsaanpakken van individuele docenten. Een mooie ontwikkeling is dat opleidingen hun visie op leren en beoordelen en de uitvoering van beoordelen al op curriculumniveau transparant communiceren. Tegelijkertijd zijn deze prachtige visiedocumenten vaak abstract en is de relatie met het dagelijkse onderwijs moeilijk te herleiden. Studenten weten vaak niet eens van het bestaan van deze bronnen. Zelfs binnen een onderwijsorganisatie worden visiedocumenten niet altijd op dezelfde wijze geïnterpreteerd.

Zeker als het gaat om competentiegericht leren – waarin niet altijd sprake is van een eenduidig goed antwoord – is inzicht in het beoordelingsproces van belang. Docenten zouden explicieter hun beoordelingsaanpak kunnen communiceren naar studenten, zodat zij beter begrijpen waarop en hoe zij worden beoordeeld. Het expliciteren van de

beoordelingscriteria en standaarden en het geven van voorbeelden kan waardevol zijn om studenten bewust te maken van wat zij moeten leren en waarom (Nicol & Macfarlane-Dick, 2006). Onderzoek waarin studenten werden betrokken bij het interpreteren en definiëren van beoordelingscriteria, heeft aangetoond dat studenten hierdoor ook daadwerkelijk beter presteren (Sluijsmans, Brand-Gruwel & Van Merriënboer, 2002).

Communicatie in duurzaam beoordelen betekent dat niet alleen het resultaat van een beoordeling wordt gecommuniceerd (voldoende, cijfer, geslaagd, goed gedaan), maar dat deze communicatie de student handvatten biedt voor verdere ontwikkeling. Belangrijke vragen hierbij zijn:

1. Welke doelen heeft de student zichzelf gesteld (wat zijn de geselecteerde beoordelingscriteria, standaarden en condities)?
2. Hoe heeft de student de taak uitgevoerd (welke vooruitgang wordt geboekt ten aanzien van de beoordelingscriteria en standaarden)?
3. Hoe gaat de student verder (welke aanpak is nodig om tot groei te komen)?

Deze *feed-up*, *feed-back* en *feed-forward* worden als de belangrijkste componenten gezien in beoordelingen die in het teken staan van toekomstig leren (Black & William, 1998; Taras, 2005). Het overbruggen van de kloof tussen wat de student al beheerst en nog niet beheerst, staat centraal (Sadler, 1989). Studies naar de kracht van *feed-up*, *feed-back* and *feed-forward* zijn waardevolle bronnen voor onderzoek naar de vormgeving van de communicatie over beoordelingen (Hattie & Timperley, 2007). Communicatie over beoordelingen is het meest effectief wanneer studenten informatie krijgen over hun individuele voortgang. Het benadrukken van de individuele vooruitgang bevordert niet alleen het geloof in eigen kunnen (*self-efficacy*; Bandura, 1977), maar reduceert ook de neiging zichzelf continu te vergelijken met medestudenten (Crooks, 1988).

ZELFSTUREN BIJ BEOORDELEN

Duurzaam beoordelen betekent naast een solide ontwerp en heldere communicatie bovenal een actieve betrokkenheid van de student, die verantwoordelijkheid voor het eigen leren zal willen en moeten nemen. Het beeld van een paar jaar studeren en daarna een baan voor het leven hoort definitief tot het verleden. De student zal voldoende bagage moeten meekrijgen om hem in staat te stellen steeds nieuwe situaties, in formele en informele contexten, het hoofd te bieden. In de historie van het beoordelen is de student echter lange tijd niet als volwaardige partner beschouwd. Om duurzaam beoordelen te kunnen realiseren wordt het tijd de student bewust te maken van de rol die hij kan vervullen in beoordelen. Zelfsturing van de student is naast ontwerp en communicatie het derde speerpunt. Ook in dit kader noem ik drie redenen waarom duurzaam beoordelen wordt belemmerd.

Een eerste probleem betreft het geringe zelfsturende vermogen van studenten. Het veronderstelt dat studenten in staat zijn te plannen en zichzelf (en anderen) te beoor-

delen. Er is echter nog maar weinig empirisch bewijs dat studenten in staat zijn hun leervraag goed te definiëren en zichzelf te beoordelen. Zelfsturing is moeilijk. Dit wordt vooral geïllustreerd door onderzoek naar zelfbeoordeling. Dat studenten hun kennen en kunnen over- of onderschatten heeft te maken met een gebrek aan kennis over de beoordelingscriteria (Bjork, 1999; Krueger & Dunning, 1999). Hoe kun je immers jezelf beoordelen op kennis waarvan je de reikwijdte niet weet? Omdat incompetenten studenten 'niet weten wat ze nog niet weten' en zichzelf dus snel overschatten, leidt een zelfbeoordeling van deze studenten niet snel tot frustratie. Dit in tegenstelling tot de sterke studenten, die veel beter beseffen wat je nog allemaal kunt leren en zichzelf daarom juist vaak onderschatten.

Een tweede reden is de lage betrokkenheid van studenten bij de beoordeling. Huidige beoordelingsmethoden zijn overwegend prescriptief van aard. Omdat de opleiding en de docent bepalen wat, hoe en wanneer wordt beoordeeld, hebben zij nog steeds de volledige controle over het onderwijsleerproces van de student (Broekkamp, 2003). Omdat de student nauwelijks betrokken wordt bij de voorbereiding en uitvoering van de beoordeling, zal het effect van de beoordeling op leren minimaal zijn. De laatste jaren zijn veel initiatieven genomen om studenten meer te betrekken bij beoordelen, zoals zelf-, peer- en portfoliobeoordelingen. Maar ook hier gaat het vooral om vormen van beoordelen die de student worden opgelegd. Zo wordt per docent vastgelegd wanneer de reflectierapporten of het portfolio moeten worden geschreven, in hoeveel woorden en volgens welke structuur. Het invoeren van zelfbeoordeling op deze manier demotiveert studenten eerder en leidt tot 'reflectiemoeheid'. Reflecteren wordt een routine en studenten schrijven op wat de docent graag in het reflectieverslag terug wil zien (Sluijsmans, Brand-Gruwel, Van Merriënboer & Bastiaens, 2003).

Een derde reden is het negeren van de motivatie, toetsangst en concepties van studenten. De wijze van beoordelen stuurt vaak het studeergedrag van studenten. Als beoordelen wordt benaderd als een aparte tak van sport, worden studenten verstoord in hun 'flow'. Als de student de kloof tussen wat hij moet leren volgens de opleiding en datgene wat hij al weet en kan vervolgens als te groot ervaart, zal ook de motivatie om te leren verdwijnen (Sadler, 1989; Martens & Boekaerts, 2008). Ook de tegenovergestelde situatie bestaat. Als studenten te lang taken uitvoeren die ze al tijden beheersen, treedt verveling op met desastreuze gevolgen voor motivatie en leren. Een tweede variabele die wordt ontkend in beoordelen, is de invloed van toetsangst. De invloed van toetsangst op de prestatie en motivatie van studenten wordt nog vaak onderschat. Onderzoek heeft aangetoond dat toetsangst het werkgeheugen zodanig beïnvloedt dat de cognitieve processen die nodig zijn voor het uitvoeren van de beoordelingstaak, niet kunnen plaatsvinden (Lee, 1999). Variabelen als motivatie en angst hangen samen met de opvattingen die een student heeft over leren en beoordelen. Deze concepties zijn bepalend voor het gedrag dat studenten vertonen in onderwijssituaties. We weten echter nog weinig over de concepties van studenten over beoordelen.

Naar meer zelfsturing bij beoordelen door de student

Omdat de student de enige is die zichzelf voortdurend kan 'zien' en omdat de kennisbronnen en omgevingen zo wisselend zijn, zal de student in staat moeten zijn het leren te reguleren. Met andere woorden, hij moet leren slim te leren. Graag stel ik enkele wegen voor die zelfsturing en eigen verantwoordelijkheid bij beoordelen kunnen bevorderen.

Het (leren) vormgeven aan het eigen leren is voor duurzaam beoordelen essentieel. Voor studenten betekent dit dat zij niet langer meer achteroverleunend het onderwijs consumeren, maar dat zij zich veel meer opstellen als *prosumers*, proactieve gebruikers die mede vormgeven aan het beoordelen door op eigen initiatief feedback te zoeken (Martens, 2007). Om zelfsturing te integreren in het leren op een wijze die én motiverend is én bijdraagt aan een zelfsturend vermogen, is het zinvol aan te sluiten bij het ontwerp van beoordelen. Om zelfsturing te bevorderen zijn er voor het ontwerp van beoordelingstaken twee belangrijke consequenties. Een eerste consequentie is dat – naast het uitvoeren van de beoordelingstaak – ook het plannen (selecteren van doelen en taken) en beoordelen in het ontwerp van beoordelingstaken van belang wordt. Plannen en zelfbeoordeling worden als de belangrijkste componenten in zelfsturend leren beschouwd (Butler & Winne, 1995; Schunk, 1990; Zimmerman, 1990). Terwijl in het 'reguliere' onderwijs alleen het uitvoeren van beoordelingstaken aan de student wordt overgelaten en de planning en beoordeling aan de docent, zal in zelfsturend leren de student (mede)verantwoordelijk zijn voor de planning en beoordeling. Het uiteindelijke zelfsturende leren zal steeds verlopen in een cyclisch proces van plannen, uitvoeren en beoordelen van de beoordelingstaak. Het is echter niet realistisch dat studenten als vanzelf hun leerproces kunnen reguleren. Een tweede consequentie in het ontwerp van zelfsturend leren is dan ook dat de ondersteuning die de student wordt geboden, zich niet beperkt tot de beoordelingstaak, maar zich uitstrekt tot plannen en beoordelen. Zo zal de docent aanvankelijk uitgewerkte voorbeelden geven van het proces van plannen en beoordelen, maar zal de student geleidelijk aan meer controle krijgen over de planning en beoordeling van taken. Controle over het eigen leren heeft in een aantal onderzoeken in het sportdomein al zijn meerwaarde bewezen (zie bijvoorbeeld Wulf & Toole, 1999).

Het geven van controle aan studenten over de wijze waarop ze worden beoordeeld, wanneer ze worden beoordeeld en door wie ze worden beoordeeld, kan leiden tot een actievere betrokkenheid en verantwoordelijkheid van studenten. Dit kan vervolgens tot een hogere motivatie en tot meer leren leiden dan wanneer het beoordelen extern wordt aangestuurd (Wulf, Raupach & Pfeiffer, 2005). Dit perspectief sluit tevens goed aan bij motivatietheorieën, zoals de 'zelfdeterminatietheorie' (Ryan & Deci, 2000), waarin competentie, autonomie en sociale verbondenheid de basisbehoeften zijn. Studenten kunnen bijvoorbeeld zelf kiezen op welke criteria en standaarden ze worden beoordeeld. Beoordelingscriteria en standaarden kunnen worden gezien als een mengpaneel: beoordelingscriteria zijn de knopjes die worden aangezet, terwijl de schuifjes kunnen worden gezien als de standaarden. Afhankelijk van de condities waaronder de taak wordt uitgevoerd, worden knopjes en schuifjes aan- of uitgezet.

Tot slot is rekening houden met verschillen tussen studenten een essentiële voorwaarde voor het realiseren van duurzaam beoordelen. Door studenten op basis van hun voorkennis en motivatie meer controle te geven over waarop, wanneer en hoe ze worden beoordeeld, kunnen zij een leerpad volgen dat bij hen past. Ertmer en Newby (1996) bieden een zinvol overzicht van vragen waarmee de student samen met medestudenten en docenten in kaart kan brengen wat de individuele behoeften zijn als het gaat om het plannen, uitvoeren en beoordelen van beoordelingstaken. Onderwijs zou meer gericht moeten zijn op het begeleiden en opleiden van zelfsturende individuen. De mate waarin studenten worden gestimuleerd tot zelfsturend leren, wordt tegenwoordig als een belangrijk kwaliteitscriterium gezien (Baartman, Bastiaens, Kirschner & Van der Vleuten, 2006).

CONCLUSIE EN DISCUSSIE

Mijn doel was u enkele gedachten voor te leggen die een eerste invulling kunnen geven aan *duurzaam beoordelen*. Ik ben van mening dat de huidige wijzen van beoordelen eindigheid in leren suggereren. Duurzaam beoordelen vraagt van opleidingen een blik over de grenzen van het diploma heen. Het leren houdt immers niet op na een opleiding. Studenten zullen moeten leren hoe ze blijven leren, simpelweg omdat de kennis-economie dat van hen vraagt. Ik besef goed dat de praktijk van alledag het zeer lastig maakt om op individueel niveau studenten informatie te verschaffen over hun kennen en kunnen op een manier die ook nog eens past bij hun leervraag en persoonlijkheid. Meestal ontbreekt de tijd als de groepen te groot en te heterogeen zijn. Ook zijn organisatorische aspecten, als het leerstofjaarclassensysteem en meerdere opleiders per jaar, niet stimulerend om voldoende ruimte voor communicatie over individuele beoordelingen te realiseren. Toch zal het onderwijs gezien de maatschappelijke ontwikkelingen ook in de randvoorwaardelijke sfeer mee moeten veranderen. Pas als leren en beoordelen sterk op elkaar zijn afgestemd, een rijke communicatie over de functie van beoordelen plaatsvindt en de student wordt gestimuleerd zich te ontwikkelen als een zelfsturend individu, mag worden gesproken van een kwalitatief hoogstaand beoordelingsmodel. Ik hoop dat deze speerpunten, sterker dan nu het geval is, worden benadrukt in de criteria die inspecties en accreditatiecommissies hanteren. Ook hun blikveld beperkt zich tot productgerichte beelden over beoordelen. Kwaliteit van beoordelen gaat echter veel verder dan een pleidooi van de Parlementaire Commissie Onderwijsvernieuwingen (2008) voor verplichte invoering van gestandaardiseerde toetsen en een verplichte voldoende voor examens en centraal examen. Nieuwe en rijkere instrumenten zijn nodig om de kwaliteit van een beoordelingskader vast te stellen, maar bovenal is noodzakelijk dat alle betrokkenen deze instrumenten goed kunnen hanteren. Heldere communicatie met studenten en docenten is hierbij essentieel. Pas dan kan een beoordelingsmodel hét visitekaartje van een onderwijsorganisatie worden.

REFERENTIES

- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Baartman, L.K.J., Bastiaens, Th.J., Kirschner, P.A. & Van der Vleuten, C.P.M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programmes. *Studies in Educational Evaluation*, 32, 153-177.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 453-459). Cambridge, MA: MIT Press.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22, 151-167.
- Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31, 4, 399-413.
- Broekkamp, H. (2003). *Task demands and test expectations: Theory and empirical research on students' preparation for a teacher-made test*. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam, the Netherlands.
- Broekkamp, H. & Van Hout-Wolters, B.H.A.M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, 19, 401-428.
- Butler, D.L. & Winne, P.H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281.
- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- De Groot, A. (1966). *Vijven en zessen: Cijfers en beslissingen*. Groningen: Wolters-Noordhoff.
- De Vries, B. (2007). *Vrijheid, blijheid? Over vraaggestuurd leren en flexibele scholen*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- Ertmer, P.A. & Newby, T.J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, 24, 1-24.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Krueger, J. & Dunning, D. (1999). Unskilled, and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 82, 180-188.
- Lee, J.H. (1999). Test anxiety and working memory. *Journal of Experimental Education*, 67, 218-241.
- Martens, R.L. (2007). *Positive learning met multimedia* (Inaugural address). Heerlen/Hilversum: Open Universiteit Nederland/Teleac/NOT.
- Martens, R. & Boekaerts, M. (2008). *Motiveren van studenten in het hoger onderwijs*. Hoger Onderwijs Reeks. Groningen: Wolters-Noordhoff.
- Nicol, D.J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.

- Parlementaire Commissie Onderwijsvernieuwingen (2008). *Tijd voor onderwijs*. Den Haag: Tweede kamer der Staten Generaal.
- Ryan, R. M. & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68-78.
- Sadler, R.D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.
- Schunk, D.H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, *25*, 71-86.
- Schuwirth, L.W.T. (2007). *Meten de maat genomen* (Inaugurele rede). Maastricht, Universiteit Maastricht.
- Sluijsmans, D.M.A., Brand-Gruwel, S. & Van Merriënboer, J. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education*, *27*, 443-454.
- Sluijsmans, D.M.A., Brand-Gruwel, S., Van Merriënboer, J. & Bastiaens, T.R. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, *29*, 23-42.
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, *53*, 466-478.
- Van der Vleuten, C.P.M. & Schuwirth, L. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*, 309-317.
- Van Merriënboer, J.J.G. & Kirschner, P.A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wulf, G., Raupach, M. & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise and Sport*, *76*, 107-111.
- Wulf, G. & Toole, T. (1999). Physical assistance devices in complex motor skill learning: Benefits of a self-controlled practice schedule. *Research Quarterly of Exercise and Sport*, *70*, 265-270.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*, 477-501.
- Zimmermann, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*, 3-17.