

HAN geeft je de ruimte

L E C T O R A A T



DUURZAAM BEOORDELEN IN
VRAAGGESTUURD LEREN

Betrokken bij beoordelen

Dr. Dominique Sluismans

Lectoraat Duurzaam beoordelen
in vraaggestuurd leren

Betrokken bij beoordelen

Intreerede, in verkorte vorm uitgesproken bij de aanvaarding van het ambt van lector
'Duurzaam beoordelen in vraaggestuurd leren' aan de Faculteit Educatie
van de Hogeschool van Arnhem en Nijmegen op 6 juni 2008 door

Dr. Dominique Slujsmans

COLOFON

Hogeschool van Arnhem en Nijmegen
Faculteit Educatie
Postbus 30011, 6503 HN Nijmegen

E dominique.sluijsmans@han.nl
T (+31) 24 353 00 22
F (+31) 24 353 00 05
I www.han.nl

ISBN: 978-90-74762-15-1

Aan de inhoud van deze uitgave kunnen geen rechten worden ontleend

INHOUDSOPGAVE

Over deze rede	5
Speerpunt 1: Ontwerpen van beoordelen	8
Drie problemen	8
<i>De beperkte functie van huidige beoordelingsmodellen</i>	8
<i>Het ontbreken van heldere beoordelingscriteria en standaarden</i>	12
<i>De gebrekkige aansluiting tussen leren en beoordelen</i>	13
Naar het ontwerpen van duurzaam beoordelen in vraaggestuurd leren	14
<i>Een ipsatief beoordelingsmodel</i>	15
<i>Het definiëren van heldere beoordelingscriteria en standaarden</i>	16
<i>Het ontwerpen van betekenisvolle beoordelingstaken</i>	17
Kwaliteit van leren: Onderzoeksvragen	20
Speerpunt 2: Communiceren over beoordelen	21
Drie problemen	21
<i>De onheldere communicatie over het beoordelingsmodel in scholen</i>	21
<i>De 'black-box' van het beoordelen</i>	23
<i>De informatie-armheid in beoordelen</i>	25
Naar rijke communicatie over beoordelen	27
<i>Transparantie over het beoordelingsmodel</i>	27
<i>Het openbreken van de 'black-box'</i>	28
<i>Naar diversiteit in de communicatie over beoordelingen</i>	29
Kwaliteit van leren: Onderzoeksvragen	31
Speerpunt 3: Zelfsturen bij beoordelen	32
Drie problemen	32
<i>Het geringe zelfsturende vermogen van leerlingen</i>	33
<i>De lage betrokkenheid van leerlingen bij beoordelingen</i>	34
<i>Het negeren van motivatie, toetsangst en concepties van leerlingen</i>	35

Naar meer zelfsturing bij beoordelen door de leerling	37
<i>Het integreren van zelfsturing in het ontwerp van beoordelen</i>	37
<i>Het verhogen van de betrokkenheid van de leerling bij het beoordelen</i>	39
<i>Het erkennen van individuele verschillen</i>	41
Kwaliteit van leren: Onderzoeksvragen	43
Een voorbeeld	44
Kernactiviteiten van het lectoraat	45
Slotwoord	47
Dankwoord	49
Referenties	51

OVER DEZE REDE

Beste aanwezigen,

Van harte welkom bij de officiële start van het lectoraat ‘*Duurzaam beoordelen in vraaggestuurd leren*’ aan de Faculteit Educatie van de Hogeschool van Arnhem en Nijmegen. Zowel beoordelen als vraaggestuurd leren zijn begrippen die sterk in de belangstelling staan, ook in de Faculteit Educatie. Ze roepen tegelijkertijd veel vragen op door het ontbreken van helderheid over de praktische implementatie. Ik besef dan ook dat ik mij met een titel als deze op glad ijs begeef in een tijd waarin het nut van onderwijsvernieuwing zwaar onder vuur ligt. Geen dag ging het afgelopen jaar voorbij of er werd in kranten, tijdschriften, en zelfs met creatieve filmpjes op YouTube pittige kritiek geuit op het huidige onderwijs. Tijdens de voorbereiding van deze rede heb ik, worstelend door de berg aan informatie inclusief het rapport ‘Tijd voor onderwijs’ van de Parlementaire Commissie Onderwijsvernieuwingen, getracht enige samenhang in al deze bronnen te ontdekken om te kunnen komen tot een ‘waarheid’ over de kwaliteit en zin van onderwijsvernieuwingen. En ik zal eerlijk zijn: ik ben er snel mee gestopt. De tegenstrijdigheden in de berichtgeving en de vaak ontbrekende wetenschappelijke onderbouwing boden weinig soelaas voor mijn zoektocht.

De boodschap die ik vandaag over beoordelen in vraaggestuurd leren wil overbrengen is in die zin ‘vernieuwingsneutraal’. Mijn doel is wel om in het licht van de maatschappelijke ontwikkelingen en theoretische inzichten een beeld te schetsen van het belang van vraaggestuurd leren en een passende visie op beoordelen. Ik richt mij daarbij niet alleen op complex leren, zoals dat bijvoorbeeld bij competenties het geval is, maar op het leren van alle vormen van kennis en kunde.

Een voor de hand liggende vraag is hoe de begrippen ‘vraaggestuurd leren’ en ‘beoordelen’ in dit lectoraat worden gedefinieerd. Een stevige basis voor de invulling van vraaggestuurd leren is reeds gelegd in de lectorale rede van Bregje de Vries ‘*Vrijheid, blijheid? Over vraaggestuurd leren en flexibele scholen*’ (2007). Zij definieert vraagsturing als ‘*differentiatie in werkvormen, leermiddelen en leerstof naar niveau, tempo, timing, belangstelling en werkwijze die recht doet aan de zelfverantwoordelijke zelfbepaling van lerenden op enig moment*’ (p. 31). Met dit lectoraat zou ik differentiatie in beoordelen aan deze definitie willen toevoegen. Vraaggestuurd leren vraagt immers om een vraaggestuurde vorm van beoordelen willen we van optimale vraagsturing kunnen spreken.

Om handen en voeten te geven aan het concept beoordelen is het zinvol de vraag te stellen waarom wordt beoordeeld. Op deze vraag zijn verschillende antwoorden mogelijk. Meestal vervult beoordelen de functie van certificering en selectie, ook wel *summatief beoordelen* genoemd. Een tweede functie van beoordelen, die tegenwoordig sterk in de

belangstelling staat, is de leerlingen voorzien van zinvolle informatie over hun prestaties om zo het leerpad verder vorm te kunnen geven. Dit wordt doorgaans gedefinieerd als *formatief beoordelen*. Naast deze twee belangrijke functies van beoordelen, wil ik u vandaag kennis laten maken met een derde functie van beoordelen, namelijk het *duurzaam beoordelen*¹. Deze functie van beoordelen erkent de huidige leerbehoeften van de leerling en de leervragen die de leerling nog in de toekomst zal hebben. Duurzaam beoordelen vereist een andere kijk op beoordelen dan we gewend zijn. Was beoordelen tot nu toe vooral het sluitstuk van onderwijs, bij duurzaam beoordelen gaat het om beoordelen dat met het oog op de toekomst beklijving van (het kunnen ontwikkelen van) kennis en kunde stimuleert, waarbij een centrale rol is weggelegd voor de leerling. In dit perspectief past duurzaam beoordelen bij het idee van een leven lang leren. Het thema ‘een leven lang leren’ neemt zowel nationaal als internationaal een prominente plaats in op de maatschappelijke agenda. Het beeld van een paar jaar naar school en daarna een baan voor het leven hoort definitief tot het verleden. Omdat kennisgebieden in hoog tempo verouderen en werkwijzen en technieken voortdurend veranderen, is het belangrijk leerlingen te leren hoe zij voortdurend nieuwe kennis en vaardigheden eigen kunnen maken.

In deze rede zal ik u proberen duidelijk te maken wat duurzaam beoordelen is en waarom duurzaam beoordelen goed zou kunnen passen bij vraaggestuurd leren en een leven lang leren. Duurzaam beoordelen heeft nogal wat consequenties voor de inrichting van het onderwijs en de rol van scholen en opleidingen. De in mijn ogen belangrijkste consequenties heb ik voor u in de volgende drie speerpunten samengevat:

- 1) Het *ontwerpen* van leeromgevingen waarin duurzaam beoordelen en vraaggestuurd leren centraal staan,
- 2) Het *communiceren* over beoordelen, en
- 3) Het *zelfsturen* bij beoordelen door de leerling.

Binnen elk speerpunt zal ik vervolgens drie vragen aan de orde stellen:

- a) Hoe wordt op dit moment duurzaam beoordelen belemmerd?
- b) Welke wetenschappelijke en praktische inzichten bieden bruikbare handvatten voor het vormgeven van duurzaam beoordelen?
- c) Welke onderzoeks – en professionaliseringsvragen zijn in het programma ‘Kwaliteit van leren’ zinvol voor scholen en opleidingen om duurzaam beoordelen vorm te geven?

Ik hoop dat deze rede voor collega’s aan de HAN en daarbuiten en voor scholen een eerste inspiratiebron zal zijn voor een gezamenlijke invulling van het lectoraat. Om hiermee een begin te maken nodig ik u van harte uit na de rede mee te discussiëren tijdens het

¹ Het begrip ‘duurzaam beoordelen’ is een vertaling van ‘sustainable assessment’ (Boud, 2007), een begrip dat in Nederland nog niet of nauwelijks aandacht krijgt.

forum. In dit forum zullen experts op het gebied van onderzoek naar onderwijsontwerp en beoordelen in dialoog treden met betrokkenen uit de Faculteit Educatie over de inhoud van mijn betoog.

In deze rede zal ik spreken over leerlingen en leraren. Met leerlingen bedoel ik de individuen die een leertraject of onderwijsprogramma *volgen*. Met leraren refereer ik naar de individuen die leerlingen *onderwijzen, begeleiden en beoordelen*.

SPEERPUNT 1

ONTWERPEN VAN BEOORDELEN



Drie problemen

Het ontwerpen 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. In de afgelopen decennia is fors geïnvesteerd in onderwijsontwerp, blijkend uit prachtige onderwijsmethoden, overzichten van kerndoelen, schoolwerkplannen en het werk van bijvoorbeeld het SLO (Stichting Leerplanontwikkeling) dat zich in opdracht van de overheid bezighoudt met curriculumontwerp. En dankzij de Stichting Beroepskwaliteit Leraren is er sprake van een gemeenschappelijke taal in de onderwijsprogramma's van de lerarenopleidingen. Het is echter jammer dat, als het gaat om het ontwerpen van onderwijs, het ontwerpen van beoordelen lijkt te worden vergeten. Als eerste speerpunt wil ik het belang van het ontwerpen van duurzaam beoordelen benadrukken. Ik zal dat doen door het duiden van drie problemen die duurzaam beoordelen belemmeren, namelijk 1) de beperkte functie van huidige modellen van beoordelen, 2) het ontbreken van heldere beoordelingscriteria en standaarden en 3) de gebrekkige aansluiting tussen leren en beoordelen. Vervolgens zal ik enkele voorstellen doen hoe we in dit lectoraat deze problemen zouden kunnen aanpakken.

Probleem 1. De beperkte functie van huidige beoordelingsmodellen

Beoordelen kent een lange geschiedenis². Als we bijvoorbeeld kijken naar de Verenigde Staten in de 19e eeuw, was het de leraar die mondeling de voortgang van de leerling rapporteerde aan de ouders. Pas toen aan het einde van die eeuw de populatie van leerlingen sterk toenam begon men klassen te formeren op basis van leeftijd, het zogenaamde leerstofjaarclassensysteem. Er verschenen nieuwe methoden van beoordelen, zoals formele voortgangsrapportages. Leraren schreven op welke vaardigheden de leerlingen al beheersten en welke nog ontwikkeld moesten worden. Leerlingen mochten pas verder naar een volgend niveau als het huidige gewenste niveau was bereikt. Deze vorm van beoordelen was de eerste vorm van een zogenaamd 'narratief rapport'. Door de leerplichtwet rond de start van de twintigste eeuw groeide de leerling-populatie enorm, met als gevolg dat de onderwijshouden specifiek en de populatie van leerlingen meer divers werden. Terwijl de lagere scholen nog de narratieve rapporten aanhielden om voortgang in leren te documenteren, werkten leraren in het voortgezet onderwijs al snel met percentages. Hier zien we het begin van de beoordelingssystemen zoals we die tot op de dag van vandaag nog kennen. Want hoe verloopt de discussie over beoordelen anno 2008?

In 2007 stond naar aanleiding van uitlatingen van onze minister-president op enig moment

² Het voert te ver voor deze rede om deze geschiedenis uiteen te zetten, voor mooie overzichten verwijs ik naar publicaties van Ben Wilbrink, zie <http://www.benwilbrink.nl/>

de zogenaamde ‘zesjescultuur’ in de belangstelling. Premier Balkenende verweet leerlingen middelmatigheid omdat zij te snel tevreden zouden zijn met een cijfer ‘6’. Tegelijkertijd bepleitte de Onderwijsraad (2007) dat iedere leerling een voldoende (dus een 6!) moet hebben behaald voor wiskunde, Engels en Nederlands om aan het hoger onderwijs te mogen deelnemen. Toch wel een vreemd advies: de consequentie is namelijk dat een deel ‘fout’ gescoord mag worden. Over welk percentage ‘goed’ en ‘fout’ hebben we het dan? Betekent dat je een deel van de vaardigheid of kennis niet hoeft te beheersen? In dit opzicht is ook het advies van de Parlementaire Commissie Onderwijsvernieuwingen (2008) om voor zowel het schoolexamen als het eindexamen in het voortgezet onderwijs een voldoende te eisen beperkt. Saillant gegeven is dat we een zesjescultuur ook terug zien in de beoordeling van instellingen voor hoger onderwijs: de meest opleidingen ontstijgen nauwelijks een zesje (zie Figuur 1).

DE BEST BEOORDEELDE HOGESCHOLEN

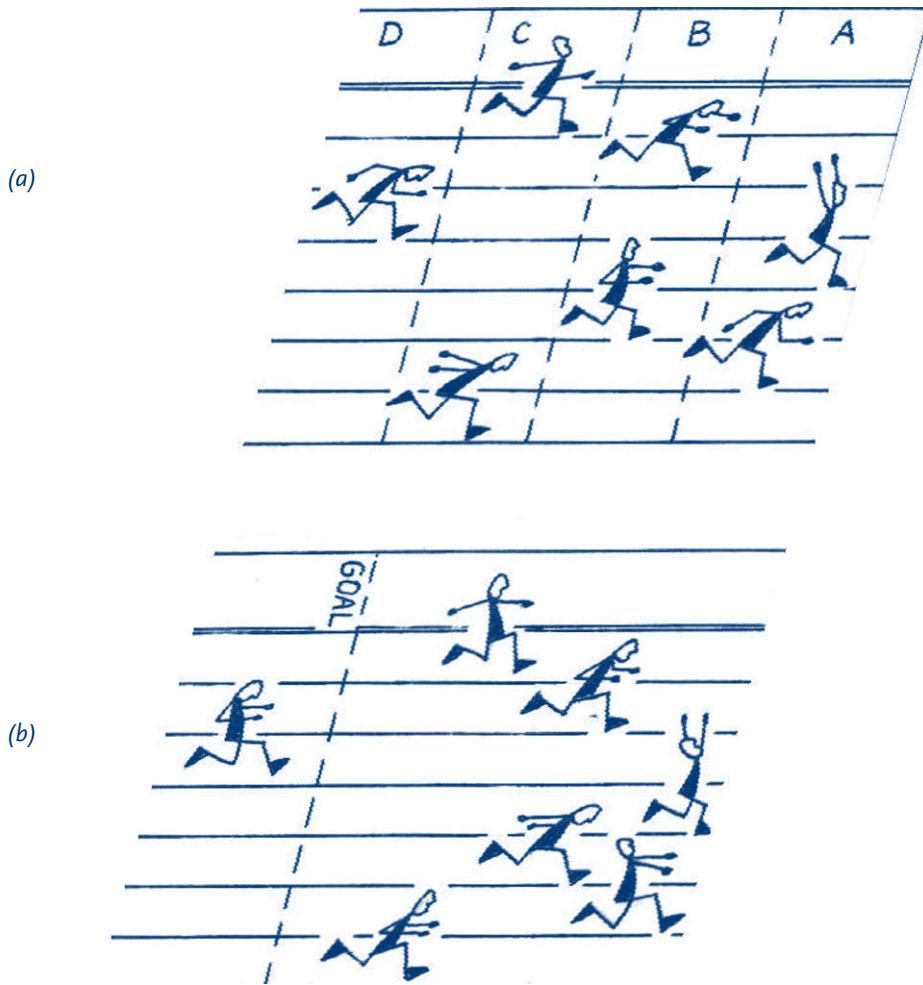
INSTELLING	BELANGRIJKSTE LOCATIES	BEOORDEELDE STUDIES	OORDEEL VAN STUDENTEN
GROTE HOGESCHOLEN			
Hogeschool Zuyd	Heerlen, Maastricht, Sittard	42	6,85
Haagse Hogesch	Den Haag, Rijswijk	32	6,84
Noordelijke Hogesch	Leeuwarden	30	6,75
Hs van A'dam	Amsterdam, Almere	49	6,73
Hs van Rotterdam	Rotterdam, Dordrecht	47	6,71
Saxion Hogeschool	Deventer, Enschede	49	6,69
Avans Hogeschool	Den Bosch, Breda, Tilburg	49	6,68
HAN	Arnhem, Nijmegen	48	6,68
Hanzehogeschool	Groningen	44	6,66
Fontys	Eindhoven, Sittard, Tilburg, Venlo	54	6,65
Hogeschool Utrecht	Utrecht, Amersfoort	53	6,63
InHolland	A'dam, Alkmaar, Delft, Den Haag, Diemen, Haarlem, Rotterdam	84	6,46

Figuur 1. Studenten beoordelen Nederlandse hogescholen: zesjescultuur?³

³ Uit Keuzegids Hoger Onderwijs, <http://www.keuzegids.org/>

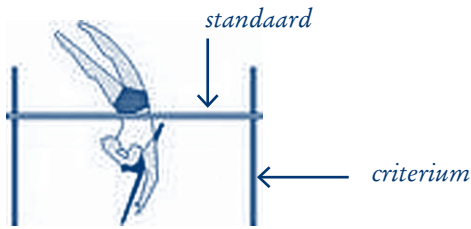
Een voorzichtige conclusie is dat ruim veertig jaar na het verschijnen van het standaardwerk *‘Vijven en zessen’* van Adriaan de Groot in 1966 nog maar bitter weinig is veranderd in het huidige beoordelingssysteem. De Groot illustreerde al met sprekende voorbeelden dat we het cijfersysteem op zeer discutabele manieren inzetten en daarbij de illusie hebben dat er goed wordt beoordeeld. Hoewel hij – gezien de willekeurige scheidslijn tussen voldoende en onvoldoende – in zijn werk vooral pleitte voor meer standaardisatie, wees De Groot op de tekortkomingen van twee beoordelingsmodellen die aan het cijfersysteem ten grondslag liggen en continue individuele ontwikkeling en dus vraaggestuurd leren belemmeren. Een constatering die mij – in het licht van mijn lectorale opdracht – buitengewoon boeide.

Het eerste model betreft het *‘normgerichte beoordelen’*, waarbij leerlingen worden afgezet tegen een gemiddelde van de groep. Dit resulteert in een zogenaamde Gausscurve (genoemd naar de Duitse wiskundige Friedrich Gauss, zie Figuur 2a). Deze vaak gestandaardiseerde beoordelingen veronderstellen uniforme leerlingen en het bestaan van een ‘gemiddelde leerling’. 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 beoordelingsmodel is dat het niveau van de groep bepalend is voor het individuele resultaat: zit je als een laag presterende leerling in een hoog presterende groep dan is de kans op een voldoende klein. Individuele resultaten kunnen dus afhankelijk van de groep waarin de leerling zich bevindt anders uitvallen (Wijnen, 1971). Bijgevolg is het moeilijk een uitspraak te doen over de prestaties en mogelijkheden van een individuele leerling omdat dit contextafhankelijk is. Met andere woorden, het wordt lastig op een betrouwbare wijze groei te meten.



Figuur 2. Het normgerichte (a) en criterium/standaardgerichte (b) beoordelingsmodel

Het tweede model gaat uit van ‘criterium/standaardgericht beoordelen’ (zie Figuur 2b), een model dat we zowel nationaal als internationaal vooral terug zien in het middelbaar en hoger beroepsonderwijs en gebaseerd is op competenties en kwalificaties. Leerlingen worden in deze benadering niet zozeer vergeleken met de norm van de groep, maar worden beoordeeld op een aantal beoordelingscriteria en bijbehorende standaarden. In Figuur 3 is het verschil tussen criterium en standaard geïllustreerd met de polsstokhoogspringer: het criterium staat vast terwijl de standaard per sporter wordt bepaald en op een steeds hoger niveau kan worden gelegd.



Figuur 3. Polsstokhoogspringen: Criterium en standaard

Resultaat van criterium/standaardgerichte beoordelingen is dat sommige leerlingen de vastgestelde standaarden op de criteria bereiken en andere leerlingen niet of met moeite. Hoewel het behalen van standaarden natuurlijk prima en noodzakelijk is, staat dit model duurzaam beoordelen in de weg. Want ook in dit model geldt dat er geen reden is om verder te leren als de standaarden zijn behaald. Leerlingen voor wie de standaarden te hoog of te laag liggen zullen bovendien gedemotiveerd raken. Ook hebben leerlingen nog steeds de mogelijkheid uitstelgedrag te vertonen en alleen datgene te doen wat in de beoordeling wordt gevraagd.

In zowel het normgerichte als criterium/standaardgerichte 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 (Frederiksen, 1984). Ook het 'doceren voor de toets' wordt in stand gehouden (Popham, 2001). Hoewel dit op zich niet negatief hoeft te zijn is het wel zinvol te analyseren of het doceren voor de toets er niet voor zorgt dat er een schijn van succes wordt gemeten bij leerlingen, die de stof tijdelijk opslaan maar deze niet eigen maken (Meherns, Popham, & Ryan, 1998). Dit is bijvoorbeeld het geval bij de schoolexamens en de centrale examens: de leraar geeft nog even een extra les waarin de examenstof wordt behandeld.

Van duurzaam beoordelen in vraaggestuurd leren is bij het gebruik van beide beoordelingsmodellen nauwelijks sprake (Shepard, 2000; Smith & Fey, 2000). De individuele leerling met zijn unieke profiel en leerbehoeften verdwijnt volledig in een groepsgemiddelde – waarmee we tevens middelmaat waarderen – of wordt slechts summatief afgevinkt op een lijst van beoordelingscriteria en vooraf vastgestelde standaarden die niet op het niveau van de individuele leerling zijn afgestemd. Normgericht beoordelen zal altijd de minder presterende leerlingen benadelen omdat zij steeds worden vergeleken met het gemiddelde van de groep. Bovenal voeden deze manieren van beoordelen de leerlingen op met conservatieve beelden over de functie van beoordelen. Het is jammer dat met name leerlingen in het voortgezet onderwijs zo gewend zijn geraakt aan het eendimensionale karakter van bijvoorbeeld een cijfersysteem en dat ze er niet of nauwelijks bij stilstaan dat beoordelen ook een functie van leren in zich kan dragen.

Probleem 2. Het ontbreken van heldere beoordelingscriteria en standaarden

Een tweede probleem in het ontwerpen van duurzaam beoordelen is het gebrek aan helderheid *wat* en op *welk niveau* wordt beoordeeld. Om het ontbreken van heldere beoordelingscriteria ('wat') te illustreren refereer ik naar een interessant en gedegen

onderzoek van Broekkamp, van Hout-Wolters, van den Bergh en Rijlaarsdam (2004). Zij onderzochten bij 22 geschiedenisleraren en hun leerlingen uit 5 VWO wat beider verwachtingen waren over een proefwerk. De te bestuderen leerstof betrof een hoofdstuk over Amerika na 1945. De leraren waren overtuigd dat de leerlingen wel wisten wat de belangrijke elementen waren in de leerstof, maar het tegenovergestelde bleek het geval. De verwachtingen van de leerlingen kwamen nauwelijks overeen met de verwachtingen van de leraar. Sterker nog, zelfs de leraren lieten onderling nog grote verschillen zien in wat zij als belangrijk achtten.

Helder geformuleerde beoordelingscriteria en standaarden vormen nog te weinig de basis in het onderwijs. Ze worden meestal per vak door de individuele leraar 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 leraar (*‘tacit knowledge’*; Sadler, 1989), waardoor de leerling niet kan begrijpen waaraan moet worden voldaan en zich dus ook niet adequaat kan voorbereiden. Uit onderzoek blijkt dan ook dat leerlingen nogal eens verrast worden door de inhoud van een beoordeling, omdat de verwachtingen van de leerling over de beoordelingscriteria niet overeenkomen met de verwachtingen van de leraar (Broekkamp & van Hout-Wolters, 2007).

Naast heldere beoordelingscriteria zijn ook heldere standaarden (‘niveau’) nodig. Ook wat betreft de standaarden op criteria is er niet altijd helderheid, blijkend uit reacties van leerlingen als ‘De toets was veel te moeilijk!’. Om leerlingen echter de kans te geven kennis en kunde eigen te maken, is het belangrijk dat de standaard niet te hoog ligt – om demotivatie te voorkomen –, maar zeker ook niet te laag ligt – om verveling tegen te gaan. Duurzaam beoordelen in vraaggestuurd leren heeft pas effect als de beoordeling aansluit bij het niveau en motivatie van de leerling en voldoende uitdagend is. Zolang beoordelingscriteria en standaarden niet helder zijn of het niveau niet realistisch is, zal van duurzaam beoordelen in vraaggestuurd leren geen sprake kunnen zijn.

Probleem 3. De gebrekkige aansluiting tussen leren en beoordelen

Het derde probleem dat duurzaam beoordelen belemmert is de gebrekkige aansluiting tussen leren en beoordelen (Biggs, 1996). Leren en beoordelen worden maar zelden in één adem genoemd. Een terugblik op mijn eigen genoten onderwijs laat zien dat ik bij benadering 4200⁴ keer ‘formeel’ ben beoordeeld. 90% van deze beoordelingen vonden plaats tussen mijn 4^e en 23^e levensjaar. Buiten het feit dat dit een schrikbarend hoog aantal is, ook omdat ik alle informele en onuitgesproken oordelen zelfs nog buiten beschouwing laat, is het vooral zorgelijk dat deze beoordelingen in mijn beleving niet of nauwelijks van meerwaarde zijn geweest voor mijn leren of hebben geleid tot beklijving van kennis of kunde. Sterker nog, het gevoel overheerste dat ik met al die beoordelingen vooral mijn tijd aan het verdoen was.

⁴ Een ruwe maar realistische schatting van het aantal overhoringen, proefwerken, (her)tentamens, rijexamens, CITO-toetsen en graadexamens bij muziek.

Tot op de dag van vandaag worden leren en beoordelen nog als twee gescheiden werelden ervaren, zowel door leerlingen als leraren. Proefwerken, overhoringen en tentamens worden als een geïsoleerde activiteit in het curriculum ingebouwd en als zodanig benaderd. Dit leidt vaak tot frustratie van leraren, omdat het maken en nakijken van dergelijke beoordelingen veel tijd kosten en niet als hun meest uitdagende taak worden beschouwd (Brown, 2004, 2006; Gibbs & Simpson, 2004).

Duurzaam beoordelen zal op de eerste plaats worden belemmerd als beoordelingen en dus ook de beoordelingscriteria en standaarden te ad hoc worden vastgesteld en niet worden gebruikt als uitgangspunt voor het ontwerpen van onderwijs (zie ook probleem 2). Hierdoor ontstaat het risico dat de beoordeling zich richt op andere criteria dan de criteria die in het onderwijs aan de orde zijn geweest. Op de tweede plaats kan duurzaam beoordelen niet plaatsvinden omdat de te leren vaardigheden worden opgedeeld in kleine eenheden, die vervolgens in meerdere vakken door meerdere leraren op bepaalde tijdstippen worden onderwezen (zogenaamde *'fragmentatie van onderwijs'*; Janssen-Noordman & Van Merriënboer, 2002). Omdat de individuele leraar voor zijn eigen onderwijs bepaalt welke criteria van belang zijn, zullen leerlingen niet ervaren dat meerdere taken of vakken een beroep doen op dezelfde beoordelingscriteria. Door het geïsoleerde karakter van beoordelingen, die steeds gericht zijn op andere criteria, is groei over taken heen moeilijk vast te stellen (Straetmans, 2006; Straetmans, Sluijsmans, Bolhuis, & Van Merriënboer, 2003; Sluijsmans, Straetmans, & Van Merriënboer, 2008). Bovenal zullen kennis en kunde niet beklijven als er te weinig oefentaken worden aangeboden die dezelfde beoordelingscriteria benadrukken. De leerling ervaart de beoordeling als een eenmalige gebeurtenis en slaakt een zucht van verlichting als de voldoende binnen is. Omdat er geen brug wordt geslagen tussen een vorige en een nieuwe taak – en ook niet tussen verschillende vakgebieden – is het lastig te bewijzen of de leerling vooruitgang heeft geboekt in zijn leren. Tot slot is het gebrek aan aansluiting tussen leren en beoordelen ook terug te zien in de ontworpen verhouding tussen 'onderwijstijd' en 'beoordelingstijd'. Voorbeelden zijn de geclusterde proefwerkweken, (her)tentamens en de halfjaarlijkse portfolio-beoordelingen in het beroepsonderwijs. Zolang onderwijstijd en beoordelingstijd zwak zijn geïntegreerd, zullen leerlingen zich vooral inspannen in de periodes vlak voor de beoordeling met het bekende uitstelgedrag als gevolg (Crooks, 1988). De mismatch tussen leren en beoordelen, blijkend uit gebrekkige aansluiting tussen oefentaken ('lessen') en beoordelingen ('proefwerken') belemmeren een continu leerproces en het beklijven van kennis en vaardigheden.

Naar het ontwerpen van duurzaam beoordelen in vraaggestuurd leren

Gezien de geschetste problemen lukt het nog niet onderwijs zo te ontwerpen dat integratie van vraaggestuurd leren en duurzaam beoordelen volledig wordt gewaarborgd. De vragen over het 'waarom beoordelen?' (probleem 1), 'wat en op welk niveau beoordelen?' (probleem 2) en 'hoe beoordelen?' (probleem 3) krijgen vanuit een ontwerpperspectief nog

onvoldoende aandacht. Dit is niet zo vreemd, aangezien beoordelen – net als in de opleiding van leraren – ook in de wetenschappelijke theorie over onderwijsontwerp nauwelijks aandacht krijgt. Beoordelen is lange tijd als een gescheiden wetenschapsdomein beschouwd, waarin voornamelijk psychometrici de leidende actoren waren. Omdat psychometrici en onderwijsontwerpers lange tijd hun eigen taal spraken en onderzoek uitvoerden, is beoordelen nog maar kort als een integraal onderdeel van onderwijsontwerp erkend. Mede hierdoor zijn in de huidige ontwerpmodellen nog weinig handvatten te vinden voor het ontwerp van onderwijs waarin leren en beoordelen zijn geïntegreerd. In mijn visie gaat het dus niet om beoordelen als een toetskundig probleem, maar om beoordelen als een *ontwerpprobleem* (zie ook Van der Vleuten & Schuwirth, 2005). Het ‘ontwerpdelen’ kan in dit opzicht voor het lectoraat zinvol zijn de genoemde problemen nader te onderzoeken. Het eerste voorstel is een alternatief beoordelingsmodel waarin de functie van duurzaamheid en vraagsturing centraal staat. Het tweede voorstel betreft het definiëren van heldere beoordelingscriteria en standaarden. Om het probleem van gebrek aan integratie van leren en beoordelen aan te pakken stel ik tevens een aantal ontwerpprincipes voor.

Een ipsatief beoordelingsmodel

In de beschrijving van het eerste probleem heb ik de tekortkomingen van huidige beoordelingsmodellen verduidelijkt. Om duurzaam beoordelen en vraaggestuurd leren mogelijk te maken zou ik graag de door De Vries (2007) aangekaarte noodzaak voor ‘*gedifferentieerd uitdagen*’ (p. 21) in vraaggestuurd leren door willen trekken naar de wijze van beoordelen. Als we stellen dat leerlingen 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, slaan we naar mijn idee de plank mis en wordt het adaptieve, dat in het onderwijs wordt opgebouwd, volledig tenietgedaan door een beoordelingsmodel dat allesbehalve adaptief is. Leerlingen 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 beperkte functie van de normgerichte en criterium/standaardgerichte beoordelingsmodellen voor vraaggestuurd leren aan te pakken, stel ik een *ipsatief beoordelingsmodel* voor. In dit model worden beoordelingen van leerlingen 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 leerling vergeleken in de tijd. Er is sprake van duurzaam beoordelen omdat criteria en standaarden flexibel kunnen worden aangepast en de beoordelingen zich niet meer beperken tot een momentopname. Ze zijn gericht zijn op het beklijven van kennis en vaardigheden waar de leerling een leven lang plezier van kan hebben. Ook wordt tegemoet gekomen aan vraaggestuurd leren: omdat het individuele leerpad van de leerling uitgangspunt is voor de beoordeling, wordt vraagsturing gerealiseerd. Maar wat betekent een ipsatief beoordelingsmodel voor de inrichting van het onderwijs? In de volgende twee voorstellen zal ik hier uitvoeriger op ingaan.

Het definiëren van heldere beoordelingscriteria en standaarden

Om meer 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 (bijvoorbeeld rekenen, taal of geschiedenis). Goed gedefinieerde gedragsdoelen omvatten een *werkwoord* dat aangeeft wat het resultaat moet zijn, de *condities* waaronder een vaardigheid moet worden gedemonstreerd en de *beoordelingscriteria en standaarden* (Van Merriënboer & Kirschner, 2007). De condities waaronder de leerling een taak uitvoert, bepalen in veel gevallen de standaard van een criterium. Het vaststellen van beoordelingscriteria en standaarden gebeurt meestal door een groep van experts. Deze criteria kunnen betrekking hebben op de cognitieve aspecten van de te leren vaardigheid (beheerst de leerling de kennis om de taken te kunnen uitvoeren?), op probleemoplossende aspecten (kan de leerling de kennis zinvol toepassen?), en/of op attitudeaspecten (houdt de leerling zich aan de juiste regels en procedures?). Ik zal – gezien het EK dat morgen losbarst – de voetballer als voorbeeld nemen om typen beoordelingscriteria te illustreren.

Een goede voetballer zal op de eerste plaats een aantal technieken moeten beheersen, zoals het nemen van een hoekschop, crosspass, lobje of hakbal. Deze technieken zullen echter alleen maar hun nut bewijzen als ze op het goede moment in de wedstrijd worden ingezet. Alleen hierdoor kan de voetballer laten zien dat hij overzicht heeft, de krachten en zwaktes van zijn medespelers kent en de kansen ziet bepaalde technieken toe te passen. Ten slotte zal een goede voetballer zich bewust zijn van de conventies die binnen de sport gelden. Hij komt op tijd bij trainingen, behandelt zijn collega's met respect en kent de regels van het spel. Pas als een voetballer de drie typen beoordelingscriteria op een geïntegreerde wijze kan laten zien, is er sprake van professioneel handelen.

Een overzicht van beoordelingscriteria voorziet in een gedetailleerde beschrijving van de gewenste leeruitkomsten. Hoewel in vergelijking met een aantal jaren terug veel winst is geboekt in het verkrijgen van een gemeenschappelijk beeld wat leerlingen moeten kennen en kunnen, verloopt de vertaling naar een ontwerp voor een onderwijsprogramma vaak moeizaam. Nu zien leerlingen en leraren in het beroepsonderwijs door de lijsten met ontelbare beoordelingscriteria de essentie van een beroep niet meer terug en wordt de kans op fragmentarisch leren en beoordelen alleen maar groter (Korthagen, 2004). Om duurzaam beoordelen mogelijk te maken zullen ontwerpers ook transparant moeten maken welke criteria eerst geleerd en beoordeeld moeten worden, welke criteria in samenhang geleerd en beoordeeld moeten worden en bij welke criteria de volgorde er niet toe doet. Het is van belang de criteria te benoemen die dusdanig routinematig zijn dat ze

afzonderlijk kunnen worden beoordeeld zonder ingewikkelde beoordelingsmethoden. Beoordelingen op basis van reproductie blijven nodig als het gaat om feiten en procedures die alleen door *'drill and practice'* worden geleerd. Tot slot moet worden vastgesteld welke criteria essentieel zijn voor succesvolle taakuitvoering of te allen tijde op een acceptabel niveau moeten worden gedemonstreerd (bijvoorbeeld het hanteren van medische apparatuur bij operaties).

Om van duurzaam beoordelen te kunnen spreken, met als doel het beklijven van kennis en kunde en het leren verwerken van nieuwe kennis en kunde, zijn vervolgens 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 deze taken (Van Merriënboer & Kirschner, 2007). Maar hoe kunnen deze taken er dan uit zien?

Het ontwerp van betekenisvolle beoordelingstaken

In het derde probleem verwees ik naar de gebrekkige aansluiting tussen leren en beoordelen. Om integratie van leren en beoordelen te waarborgen is – naast een ipsatief beoordelingsmodel en een set van heldere beoordelingscriteria – een aantal beoordelingen nodig waarin de leerling zijn beheersing van criteria op bepaalde standaarden kan demonstreren. Omdat deze beoordelingen in de visie van duurzaam beoordelen vooral in functie staan van verder leren, wil ik deze beoordelingen benoemen als *betekenisvolle beoordelingstaken*. In deze beoordelingstaken worden leren en beoordelen geïntegreerd, omdat ze leerlingen aanzetten tot het werken aan beoordelingscriteria die gericht zijn op beklijving van kennis, vaardigheden en houdingen. Beoordelingstaken kunnen de leerling in dit perspectief in vele vormen worden aangeboden zoals opdrachten, situaties, problemen of projecten. Om duurzaam beoordelen in vraaggestuurd leren te realiseren zijn twee ontwerpprincipes van belang: de moeilijkheid van de beoordelingstaak en de mate van geboden ondersteuning bij het uitvoeren van de beoordelingstaken⁵.

De *moeilijkheid* van een beoordelingstaak wordt vastgesteld door het identificeren van de factoren die de beoordelingstaak meer of minder moeilijk maken. Voor een leraar in opleiding bijvoorbeeld kunnen dit factoren zijn als het aantal leerlingen in de klas, een aantal leerlingen met een gedragsprobleem, de moeilijkheid van het inhoudsdomen dat moet worden onderwezen, of de didactische methode. Ook kan de moeilijkheid van een taak samenhangen met de mate van authenticiteit (Gulikers, Bastiaens, & Kirschner, 2004).

⁵ Ik zal mij in de beschrijving van deze beoordelingstaken grotendeels baseren op de uitgangspunten van het 'Vier Componenten-onderwijsontwerpmodel' (kortweg 4C/ID-model), oorspronkelijk ontwikkeld door Van Merriënboer, Jelsma en Paas (1992). Het complete ontwerpmodel en de psychologische achtergronden staan beschreven in het boek *Training Complex Cognitive Skills* (van Merriënboer, 1997) en *Ten Steps to Complex Learning* (van Merriënboer & Kirschner, 2007). Een praktische uitwerking die is gericht op de onderwijspraktijk wordt beschreven in het boekje *Innovatief Onderwijs Ontwerpen* (Janssen-Noordman & van Merriënboer, 2002).

Beoordelingstaken die vergelijkbaar zijn in moeilijkheid kunnen worden gebundeld, in een zogenaamde taakklasse. Voor de vaardigheid ‘begeleiden van leerlingen’ kunnen de taken in een eerste taakklasse bijvoorbeeld alleen gericht zijn op een leren begeleiden van een groepje van maximaal drie leerlingen, die geen gedrags- of leerproblemen hebben, en er sprake is van een rustige leeromgeving. In de daaropvolgende taakklassen worden de condities complexer (meer leerlingen, moeilijke onderwysinhoud, didactische methoden die veel organisatie vereisen, enzovoort). Binnen een set van beoordelingstaken in eenzelfde taakklasse is het mogelijk bepaalde beoordelingscriteria meer dan andere te benadrukken. Dit stelt leraren en leerlingen in staat de aandacht te richten op bepaalde aspecten van de taak zonder de hele taak uit het oog te verliezen. Een voorbeeld van deze nadrukmanipulatie per taakklasse voor het leren ‘lesgeven’ zou kunnen zijn (Van Merriënboer & Kirschner, 2007):

Taakklasse 1: Lesgeven, met nadruk op de te onderwezen inhoud en het geven van uitleg (Is de onderwysinhoud correct? Worden de juiste didactische methoden toegepast?)

Taakklasse 2: Lesgeven, met nadruk op de interactie met leerlingen (Hoe verloopt de interactie? Worden de juiste vragen gesteld?)

Taakklasse 3: Lesgeven, met nadruk op alle bovengenoemde aspecten.

Het tweede belangrijke principe in het ontwerp van beoordelingstaken is de mate van geboden *ondersteuning* aan de leerling. De mate van ondersteuning wordt bepaald door vier elementen op basis waarvan elke beoordelingstaak kan worden beschreven: de gegeven situatie waarmee de leerling wordt geconfronteerd (bijvoorbeeld de vraag een staartdeling te maken), het doel (leren een staartdeling te maken), een oplossing (de uitkomst van de staartdeling) en het probleemoplosproces (welke stappen zijn ondernomen door de leerling om van de vraag naar het doel te komen). Vervolgens kunnen op basis van deze vier elementen taken worden ontworpen. Zo zal de leerling aanvankelijk beoordelingstaken uitvoeren die kunnen worden bestempeld als ‘uitgewerkte voorbeelden’, waarin de leerling een uitwerking krijgt van zowel de gegeven situatie, het probleemoplosproces en het doel. Leerlingen krijgen door het bestuderen van deze uitgewerkte voorbeelden een goed beeld van de onderliggende beoordelingscriteria. Een voorbeeld van een beoordelingstaak met afnemende ondersteuning is een aanvulprobleem, waarin een gedeelte van het oplosproces en het doel ontbreekt. De leerling voert zelf een deel van de beoordelingstaak uit. Tot slot zijn er beoordelingstaken waarin de ondersteuning volledig ontbreekt en alleen een gegeven situatie wordt gepresenteerd. De leerling zal zelf het probleemoplosproces en het doel moeten invullen. Binnen een reeks van beoordelingstaken met dezelfde moeilijkheid zal de leerling aanvankelijk veel observeren en vervolgens

geleidelijk steeds zelfstandiger deze taken uitvoeren. Dit proces van ‘*scaffolding*’ wordt als cruciaal gezien in het ontwerpen van vraaggestuurd leren (Van Merriënboer, Clark, & de Croock, 2002).

Als samenvatting heb ik een mogelijke ordening van beoordelingstaken die ontworpen zijn volgens de principes van moeilijkheid en ondersteuning weergegeven in Figuur 4. In dit figuur zijn de beoordelingstaken (gerepresenteerd als bolletjes) binnen de gestippelde omlijnningen even moeilijk. De mate van ondersteuning is gerepresenteerd als de vulling van de bolletjes.



Figuur 4. Moeilijkheid, ondersteuning en beoordelingscriteria in het ontwerpen en ordenen van beoordelingstaken

Uit Figuur 4 is tevens af te lezen dat beoordelingscriteria in leren niet meer per taak worden vastgesteld, maar voor een reeks van taken die variëren in moeilijkheid en ondersteuning. De set van beoordelingscriteria zal in een individueel leerpad steeds verder worden uitgebreid, zodat de leerling uiteindelijk aan alle criteria kan voldoen. Door aan een (cluster van) beoordelingscriteria standaarden te koppelen die worden vastgesteld door de moeilijkheid van de taak, ontstaat een ‘scoringsrubriek’. Door het vormgeven van scoringsrubrieken wordt elk criterium in context geplaatst, waardoor het makkelijker wordt tot een oordeel te komen en de leerling te voorzien van adequate feedback (Sadler, 1989). Zo heeft Stichting Leerplanontwikkeling (2006) voor het vmbo fraaie scoringsrubrieken ontwikkeld⁶ voor de vaardigheden communiceren, informatie verwerven en verwerken, presenteren, samenwerken en plannen en organiseren.

Een groot aantal beoordelingstaken met een set aan heldere beoordelingscriteria als fundament maakt het mogelijk dat leerlingen afhankelijk van hun leerbehoefte, voorkennis en motivatie beoordelingstaken selecteren, die in moeilijkheid en ondersteuning aansluiten bij deze leerbehoefte (Van Merriënboer, Sluijsmans, Corbalan, Kalyuga, Paas, & Tattersall, 2006). Hoewel in Figuur 4 de ordening van taken heel lineair is geordend, zal deze ordening in vraaggestuurd leren voor elke leerling anders zijn. Omdat de leerling op basis van zijn leerbehoefte een reeks van taken kan selecteren die een beroep doen op dezelfde criteria, volgt hij een uniek leerpad en is het mogelijk groei vast te stellen en ipsatief te beoordelen. Omdat de leerling werkt aan een unieke set van beoordelingstaken met onderliggende criteria en standaarden, bouwt elke leerling een individueel portfolio op dat

⁶ Zie <http://www.slo.nl/themas/00057/00009/Rubrics/> voor diverse scoringsrubrieken

transparant maakt op welke beoordelingscriteria en onder welke condities vooruitgang wordt geboekt. Deze ipsatieve beoordeling is gericht op *toekomstig leren* met als leidende vraag: Hoe kan de leerling verder groeien? (Ward, Gruppen, & Regehr, 2002). Omdat de beoordeling niet meer alleen wordt afgezet tegen een groep leerlingen of een criterium is dit beoordelingsmodel ook motiverend voor de laag presterende leerlingen (Crooks, 1988). Begeleiding en beoordeling zullen in het ipsatief beoordelingsmodel volledig samenvallen (Brouwer, 2007).

Onderzoek van Slavin (1980) laat zien dat een ipsatief beoordelingsmodel tot beter leren leidt dan een normgericht beoordelingsmodel. Bovendien blijken leerlingen die werken volgens individueel vastgestelde leerdoelen realistischere strategieën in te zetten dan leerlingen in een normgericht beoordelingsmodel. Recent onderzoek heeft aangetoond dat wanneer de moeilijkheid en ondersteuning wordt aangepast aan het niveau van de leerling dit ook tot beter leren leidt (Corbalan-Perez, Kester, & Van Merriënboer, 2006).

Kwaliteit van leren: Onderzoeksvragen

Met de drie voorstellen gericht op het ontwerp van een ipsatief beoordelingsmodel met een set van beoordelingscriteria en een variatie aan beoordelingstaken als basis, geef ik een eerste invulling aan duurzaam beoordelen. Het lectoraat beoogt het ipsatief beoordelingsmodel verder uit te werken en te onderzoeken. Zo kan het zinvol zijn om met focusgroepen waarin leraren en leerlingen participeren de huidige beelden over leren en beoordelen boven tafel te krijgen. Ook vragen rondom kwaliteit zijn in dit kader interessant. In duurzaam beoordelen zijn criteria als validiteit en betrouwbaarheid niet voldoende om de kwaliteit van beoordelen te bepalen. Kwaliteitscriteria als bruikbaarheid, betekenisvolheid, eerlijkheid, vergelijkbaarheid en de consequenties van beoordelingen worden sterker benadrukt (Baartman, 2008; Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006). Omdat beoordelen veel meer wordt geïntegreerd in het onderwijs zal een kwaliteitsoordeel zich niet meer alleen beperken tot een uitspraak over de kwaliteit van het beoordelingsinstrument, maar veel meer gericht zijn op de aansluiting tussen beoordelen en vraaggestuurd leren. Samenvattend kunnen de volgende vragen een eerste aanzet zijn voor verdere uitwerking in projecten:

- Hoe zijn huidige beoordelingsmodellen vormgegeven en welke functies vervullen deze modellen (summatief, formatief, duurzaam)?
- Hoe ontwerpen leraren beoordelingstaken en variëren deze voldoende in moeilijkheid en ondersteuning?
- Welke taken in het curriculum doen een beroep op dezelfde beoordelingscriteria?
- Aan welke kwaliteiten moet een vraaggestuurd beoordelingskader voldoen?
- Hoe kunnen de effecten van duurzaam beoordelen (een ipsatief beoordelingsmodel) worden onderzocht?

SPEERPUNT 2

COMMUNICEREN OVER BEOORDELEN



Drie problemen

In het eerste hoofdstuk heb ik het belang van een ipsatief beoordelingsmodel met een geïntegreerd onderwijsontwerp van leren en beoordelen als basis geïntroduceerd. Belangrijkste uitgangspunten in deze ontwerpaanpak zijn het formuleren van heldere beoordelingscriteria en standaarden en het waarborgen van een veelheid aan beoordelingstaken die voldoende variëren in moeilijkheid en ondersteuning. Met de introductie van het ipsatief beoordelingsmodel heb ik een eerste invulling gegeven aan mijn visie op duurzaam beoordelen. Maar dit is pas een eerste stap. Duurzaam beoordelen heeft pas meerwaarde als leerlingen begrijpen wat een beoordeling betekent voor het nemen van een vervolgstap in het leerpad. Pas als dit begrip er is kan vorming, groei en motivatie worden bevorderd. Het tweede speerpunt van deze rede richt zich op het communiceren over beoordelen. Ook hier zal ik vertrekken vanuit drie problemen die duurzaam beoordelen in vraaggestuurd leren op dit moment nog belemmeren. Deze zijn 1) de onheldere communicatie over het beoordelingsmodel in scholen, 2) de ‘black-box’ van het beoordelen en 3) de informatie-armheid van beoordelingen. Vervolgens stel ik enkele wegen voor die we in het lectoraat zouden kunnen bewandelen om deze problemen aan te pakken.

Probleem 1. De onheldere communicatie over het beoordelingsmodel in scholen

Om het eerste probleem te illustreren, leg ik het volgende voorbeeld aan u voor dat stamt uit mijn eigen verleden.

Een zomerse dag eind juni 1991. Ik weet het nog goed. Die dag zou ik te horen krijgen of ik naar de zesde klas van het Gymnasium mocht. Heerlijk, naar de eindexamenklas, samen met de groep vrienden met wie ik zojuist een fantastische reis naar Rome had gemaakt. Ik keek er naar uit. Maar helaas...een 5 teveel, voor biologie. Een herexamen over genetica volgde. Het mocht niet baten. Met een eerste serieuze verkering van net een maand oud, bleek de leerstof over genetica, hoe gek het misschien ook mag klinken, ver van motiverend. Telefoontje, rector aan de lijn, het vijfde jaar moest over. Wat een dreun! Wie had dat ooit gedacht. Een paar dagen later zat ik met mijn vader bij de rector en nog drie andere leraren. In alle rust legde mijn vader alle eindrapporten van klas 1 t/m 5 naast elkaar. Prijkte er op de eerste drie nog prachtige cijfers, in jaar 4 ging het behoorlijk neerwaarts met de bekende uitslag in jaar 5. In alle ernst vroeg mijn vader hoe de school dit verklaarde en waarom hij niet eerder een signaal had ontvangen dat het misschien niet de goede kant opging. Een antwoord bleef uit...

Dit voorbeeld maakt verschillende dingen duidelijk. Op de eerste plaats kan de school onvoldoende beargumenteren hoe deze situatie heeft kunnen ontstaan en bleek het lastig het beoordelingsmodel van de school transparant te maken. De regel van het aantal toegestane onvoldoendes om over te gaan of te doubleren is helder, maar daar houdt het dan ook op. Het voorbeeld illustreert op de tweede plaats dat te weinig over de voortgang van de leerling wordt gecommuniceerd. Buiten de periodieke rapporten is er nauwelijks ruimte om tijdig aan de bel te trekken. Ook ouders nemen misschien (rapport)cijfers te snel voor lief. Wat zegt het eigenlijk als je kind thuis komt met een 4? Is dit structureel? Had hij of zij niet gestudeerd? Is er sprake van toetsangst? In mijn geval waren niet zozeer de cognitieve aspecten, maar vooral de motivationele aspecten van invloed op de voortgang. Zelden worden deze aspecten meegenomen in de beoordeling of wordt hierover gecommuniceerd. Onlangs stond in de Volkskrant⁷ een bericht van een bezorgde moeder. Haar zeventienjarige zoon die werktuigbouwkunde mbo niveau 4 studeerde had besloten niet meer naar school te gaan.

(...) Wij bezochten een rapportavond, de zogenaamde tien-minutengesprekjes. De man van wiskunde was de meest ongeïnspireerde werknemer die we ooit zagen (...). Daarna liepen we verbaasd naar onze volgende afspraak. Natuurkunde. Nog erger. Enkele weken later hadden we, op ons initiatief, een gesprek met de mentor. Zij herkende de problemen en gaf aan dat niet alleen wij bij haar op de stoep stonden. Daarna nooit meer wat van gehoord, ondanks de toezegging contact te houden. (...)

De beide voorbeelden maken helder dat de consequenties van normgerichte- en criterium/standaardgerichte beoordelingsmodellen onvoldoende worden gecommuniceerd. In de communicatie over beoordelingen die gericht zijn op het gemiddelde van de groep (zoals in normgerichte beoordelingsmodellen) of op een criterium, schuilt het gevaar dat een deel van de groep leerlingen de informatie niet begrijpt. Zo zullen de laag presterende leerlingen informatie ontvangen over beoordelingscriteria die ze nog niet begrijpen, terwijl hoog presterende leerlingen informatie ontvangen over criteria die ze al tijden beheersen. Een jaar doubleren – ten gevolge van het leerstofjaarklassensysteem – en zelfs volledige uitval heeft veel consequenties voor leerlingen en ouders en is lang niet altijd zinvol. Bij zitten-blijven zullen immers alle vakken die op voldoende niveau zijn óók over worden gedaan, met verveling en demotivatie tot gevolg (Rheinberg, 1983). En dan heb ik het nog niet over de sociale aspecten die het doubleren van een klas en uitval met zich meebrengen. Hoewel ook deze ontwikkelingen al decennia lang ter discussie worden gesteld (bijvoorbeeld door Doornbos, 1969), zijn ze nog steeds aan de orde van de dag. Communicatie over het beoordelingsmodel is voor vraaggestuurd leren en duurzaam beoordelen van groot belang. Het stelt leraren, leerlingen en ouders in staat tijdig en specifiek struikelblokken in de voortgang te signaleren en hierop adequaat in te springen.

⁷ 'Niet meer in de jungle', ingezonden artikel in de Volkskrant van 4 april 2008

Probleem 2. De 'black-box' van het beoordelen

Het tweede probleem in de communicatie over beoordelingen illustreer ik met volgend voorbeeld.



Figuur 5. Docenten van de kweekschool anno 1968

Uit: Notulen van een docentvergadering (Figuur 5) op een Jongenskweekschool, anno 1968

Agendapunt C: De leerlingen zouden graag zien dat proefwerken gecorrigeerd worden teruggegeven

Leraar C: Ik doe het niet. Te veel reacties ofschoon de leerlingen niet tot beoordeling in staat zijn. Ik geef zelfs het punt niet door: bij een hoog punt doen ze de volgende keer niets.

Leraar H. Corrigeren betekent voor mij niet alleen fouten aanstrepen, maar ook aantekenen waarom het antwoord fout is. Dit brengt veel werk met zich mee en het beroerde is dat de leerlingen desondanks dat altijd kritiek hebben.

Leraar N. Ik geef de proefwerken – van door het jaar – altijd terug. Jongens hebben steevast kritiek.

Lerares M. Zwakke leerlingen vragen het altijd terug, niet om er wijzer van te worden, maar uitsluitend om kritiek te leveren.

Leraar B. Ik heb nooit moeilijkheden. Ik geef ze terug en bespreek ze.

Leraar H. Ook niet bij proefwerken over theorielessen?

Leraar B. Nee, leerlingen hebben recht op teruggave.

Mister X. Hoeveel proefwerken geeft u?

Leraar B. Per trimester per klas drie proefwerken.

- Leraar C. Ik moet diverse onderdelen geven en hanteer dan ook verschillende beoordelingsmethodieken. Vele vragen hebben bijvoorbeeld slechts een inleidend karakter en ik corrigeer dan ook slechts de antwoorden op vragen die de kern raken. De jongens snappen daar niks van. Alleen zwakke leerlingen geven blijk van wantrouwen. Ik neem dat niet. De directie moet ons daartegen beschermen!*
- Leraar J. Ik geef proefwerken altijd terug. Soms met moeilijkheden. Leerlingen kunnen niet beoordelen. Moeten meer vertrouwen hebben.*
- Leraar B. We moeten ons willen en durven verantwoorden.*
- Directeur: Als we maar vermijden dat de leerlingen het gevoel krijgen: Leraren gooien er met de pet naar.*
- Leraar H. Vanwaar dat argwanende?*
- Directeur: Was vroeger ook zo.*
- Leraar V. Teruggave is positief. Kennis van een methode van beoordelen is leerzaam. (Veel collega's reageren afwijzend, methode nooit prijsgeven!)*
- Leraar H. Op diverse scholen is teruggave verplicht!*
- Onderdir.: Ter toelichting: proefwerken zijn er om op de hoogte te blijven van de prestaties. Voor de leerlingen zijn ze stimulerend.*
- Directeur: Teruggave is een wens van de leerlingen. Over verplichting gaan we het niet hebben.*

Deze authentieke discussie die plaatsvond in een docentenvergadering van een toenmalige Jongenskweekschool (nu de Pedagogische Academie Basisonderwijs) is een veelzeggend voorbeeld als het gaat om een visie op beoordelen.

Het voorbeeld schetst op de eerste plaats de mens achter de beoordelaar met al zijn zorgen en persoonlijke voorkeuren. Over de feilbaarheid van de menselijke beoordelaar is al veel gepubliceerd (De Groot, 1966; Wijnen, 1971). Zo hebben vele studies aangetoond dat leraren met vergelijkbare deskundigheden dezelfde taak verschillend beoordelen. Al in 1912 (!) toonden Starch en Elliot aan dat het cijfer dat leraren Engels van verschillende scholen aan twee identieke essays gaven varieerde van 64 tot 98. Sommige leraren richtten zich voornamelijk op grammatica, stijl, netheid en spelling, terwijl andere leraren veel meer waarde hechtten aan de boodschap van het essay en hoe goed deze door de leerling voor het voetlicht werd gebracht. Een zelfde resultaat is zelfs te zien bij het zo exacte wiskunde (Starch & Elliott, 1912, 1913) en Nederlands (opstelbeoordelingen; Meuffels, 1989). 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.

Het voorbeeld laat ook zien dat iedere leraar zo zijn eigen ‘beoordelingsregels’ en ‘strategieën’⁸ hanteert, maar deze helaas niet of nauwelijks expliciteert (De Groot, 1966). Het beoordelingsproces voltrekt zich als het ware in een ‘black box’. Deze regels en strategieën worden in de praktijk door leraren nauwelijks gedeeld met leerlingen, collega’s, leidinggevenden, werkveld en ouders. Pas tijdens rapportenvergaderingen vernemen leraren hoe een leerling het bij andere vakken doet. Het ontbreken van heldere schoolloopbaandossiers over vakken en jaren heen maakt het moeilijk de groei van leerlingen goed te kunnen communiceren naar de leerlingen, de ouders en collega’s.

De consequentie van uiteenlopende beoordelingsaanpakken van leraren is ook dat leerlingen zich laten leiden door de aanpak van de individuele leraar en dat zij inconsistenties in aanpakken feilloos aanvoelen (Natriello, 1987). Zo kennen we allemaal nog wel de leraren bij wie je goed kon spieken, de leraren van wie de proefwerken elk jaar hetzelfde waren en de leraren die streng dan wel minder streng beoordeelden. Broekkamp (2003) adviseert dat leraren beter communiceren met leerlingen over de beoordelingscriteria waarop een beoordelingstaak zich zal richten.

Het voorbeeld illustreert tot slot dat er nauwelijks sprake is van een eenduidige visie op de functie van beoordelen (en dus het beoordelingsmodel, zie probleem 1). Terwijl de één het belangrijk vindt dat gecorrigeerde proefwerken wel terug moeten worden gegeven omdat ze de lerende van belangrijke informatie voorzien, ziet de ander het nut van teruggave niet in, omdat de lerende er toch niks mee doet. Omdat het beoordelen zich vooral in de hoofden van de individuele leraren voltrekt (en leidt tot zogenaamde ‘ideosyncratische’ strategieën), is het lastig tot structurele veranderingen in (de visie op) beoordelen te komen. Dure professionaliseringsactiviteiten over beoordelen zijn meestal meer gericht op individueel gedrag dan op het verbeteren van samenwerking tussen leraren en kennisdeling over beoordelen (Crow & Pounder, 2000; Fishman, Best, & Tal, 2003). Ook zijn ze niet gericht op het doorbreken van de fragmentarische structuur die in het huidige onderwijs nog vaak bestaat (Crow & Pounder, 2000; zie ook de problemen zoals beschreven bij Speerpunt 1). Om te komen tot een effectieve hervorming in het beoordelen zal een gemeenschappelijk begrip⁹ tussen leraren onderling en tussen leerlingen en leraren moeten ontstaan over de functie van beoordelen. Zeker gezien de impact die beoordelen heeft op het gedrag van leerlingen is het ontbreken van een gemeenschappelijke visie op beoordelen zeer problematisch (Klimoski & Mohammed, 1994).

Probleem 3. De informatie-armheid van beoordelingen

Het derde probleem wordt geïllustreerd door de volgende twee voorbeelden. Voor het eerste voorbeeld refereer ik naar een casus uit mijn familiekring. Het gaat om mijn oudere zus Daniëlle.

⁸ Ik verwijs naar het boek ‘Vijven en zessen’ van Adriaan de Groot (1966) voor een aantal uiteenlopende en sprekende voorbeelden

⁹ Een ‘shared mental model’ (Klimoski & Mohammed, 1994)

Daniëlle is 11 jaar werkzaam als verpleeghuisarts en voert dit vak met hart en ziel uit. Vanuit deze functie is bij haar de wens en leerbehoefte ontstaan zich te specialiseren in de begeleiding van terminale patiënten, de zogenaamde palliatieve zorg. Sinds kort volgt zij hiervoor een pittige opleiding samen met een klein aantal collega-verpleeghuisartsen. Analyse van casussen, reflectie op het beroeps handelen en intervisie zijn centrale activiteiten in deze opleiding. Enkele weken terug is aan ieder in de groep gevraagd een opdracht uit te werken en deze uitwerking in te leveren ter beoordeling. De opdracht richtte zich op het uitbrengen van een advies over de verdere behandeling van een patiënt. Het resultaat van deze opdracht was zeer teleurstellend: een 6. Terwijl de meeste leerlingen wel tevreden zouden zijn met dit cijfer – het is immers een voldoende – hakte dit resultaat er behoorlijk in.

Niet alleen het cijfer viel tegen, vooral de wijze van beoordelen leidde tot frustratie. Niet alleen de beoordelingscriteria waren bij het verstrekken van de opdracht onvoldoende specifiek gecommuniceerd (zie probleem 2, Speerpunt I), ook was er nauwelijks sprake van specifieke en constructieve feedback. Nogal vreemd gezien de complexiteit van de opdracht die diende te worden uitgewerkt. Een tweede voorbeeld:

Drie weken geleden heeft Susanne een proefwerk Engels gehad. Het was een pittig proefwerk. Omdat Duits al niet haar beste vak is, vreest ze de uitslag. De les begint en de stapel ligt al klaar op de tafel van de lerares. Susanne ziet al de grijns op haar gezicht. Op de een of andere manier heeft ze het idee dat ze altijd geniet van dit soort momenten. Susanne weet ook dat ze stevast begint met het noemen van de onvoldoendes. Martijn, 3.4, Joyce, 4.1, Susanne, 4.8.....Tja, dat is wel weer een domper. Na tien namen hoort ze de lerares nog in de verte zeggen ‘en de rest mag zich gelukkig prijzen!’ Naast tien sippe hoofden, klinkt bij de rest een collectieve zucht van opluchting. Het proefwerk wordt niet teruggegeven, de lerares heeft het immers weer nodig voor het daaropvolgende jaar... De les begint.

Ook in het voorbeeld van de lerares Engels wordt een aantal aspecten rondom de communicatie over beoordelen duidelijk. Op de eerste plaats zal het resultaat, waarin aanvankelijk rijke informatie van beoordelaars wordt teruggefilterd tot een saai, simpel, nietszeggend cijfer, de leerling nauwelijks motiveren tot verder leren (Schuwirth, 2007; Schuwirth & Van der Vleuten, 2005). Ook wordt het cijfer op een moment verstrekt dat de leerling het proefwerk nog maar nauwelijks voor de geest weet te halen en in geschetst voorbeeld zelfs niet eens meer terugziet. Hoewel onderzoek al meerdere malen heeft aangetoond dat cij-

fers de meest zwakke vorm van feedback zijn (Black & Wiliam, 1998; Hattie & Timperley, 2007), zien we dit model van beoordelen nog veel terug in het onderwijs. Hoewel in het beroepsgerichte onderwijs deze wijze van beoordelen al minder gebruikelijk is, passen veel leraren ze toch nog toe. Op de tweede plaats blijft ook hier de bespreking van het proefwerk achterwege. Het gevolg is dat de leerling naar huis wordt gestuurd met een 'lege' beoordeling, die geen informatie geeft over welke aspecten verbeterd moeten worden en hoe de leerling dit kan aanpakken. Tot slot worden leerlingen met een voldoende en een onvoldoende nadrukkelijk onderscheiden, dat vervolgens weer een impact kan hebben op motivatie van leerlingen verder te willen leren. In normgerichte- en criterium/standaardgerichte beoordelingsmodellen zullen cijfers vooral voor de laag presterende leerlingen desastreus zijn. Sterke leerlingen echter zullen alleen maar verder worden bevestigd in hun kunnen. De vraag is dan ook of en voor wie deze beoordelingen zinvol zijn. Het houdt een zesjes-cultuur in stand en leidt niet of nauwelijks tot een beter begrip en verhoogd enthousiasme voor de te leren vakinhouden.

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 echter informatie ontbreekt over de betekenis van de beoordeling en de te volgen weg is er geen sprake van vraaggestuurd leren en duurzaam beoordelen (Yorke, 2003; Taras, 2005).

Naar rijke communicatie over beoordelen

Uit de geschetste problemen blijkt dat het *ontwerp* en de functie van het beoordelingsmodel te weinig expliciet worden gecommuniceerd, dat de *uitvoering* van beoordelen zich in een 'black-box' voltrekt en dat de *resultaten* op beoordelingstaken vaak informatie-arm zijn door het cijfersysteem. Ook hier wil ik enkele voorstellen doen die tot een rijkere informatie over beoordelen kunnen leiden. Het eerste voorstel richt zich op het verhogen van de transparantie van het beoordelingsmodel door scholen en opleidingen voor leraren, leerlingen en ouders. In het tweede voorstel pleit ik voor het openbreken van de 'black-box' in beoordelen, door leraren veel duidelijker de uitvoering van beoordelingen te laten communiceren. Tot slot zijn er aanknopingspunten in onderzoek te vinden die tot meer diversiteit in de communicatie over beoordelen kunnen leiden.

Transparantie over het beoordelingsmodel

Om duurzaam beoordelen in vraaggestuurd leren te realiseren is op het niveau van de onderwijsorganisatie goede communicatie naar leraren, leerlingen en ouders over de visie op beoordelen van belang. Het is zinvol dat scholen op een toegankelijke manier helder maken welk beoordelingsmodel of mix van beoordelingsmodellen (formatief, summatief, duurzaam) wordt gehanteerd om de functies van beoordelen te verhelderen. Omdat leren

en beoordelen worden geïntegreerd worden deze beoordelingsmodellen bij voorkeur niet in separate documenten gerapporteerd (zoals dat nu nog veel het geval is in bijvoorbeeld toetskaders), maar zijn ze direct gekoppeld aan de visie op leren. Ook op het niveau van de leraren is het van belang het beoordelingsmodel helder te communiceren, zodat er meer synergie ontstaat tussen de beoordelingsaanpakken van individuele leraren. Ook zullen de leerlingen betrokken moeten zijn in de communicatie over het beoordelingsmodel. Dit kan bijvoorbeeld door ze bij een open dag uit te leggen hoe in de opleiding het beoordelen wordt vormgegeven. Een activiteit kan dan het uitvoeren van een beoordelingstaak zijn. Een mooie ontwikkeling bij de HAN is dat in de meeste opleidingen de visie op leren en beoordelen en de uitvoering van beoordelen al op curriculumniveau veel transparanter is dan voorheen. Als fraaie voorbeelden noem ik de documenten van het Domein Educatie (2005a, 2005b, 2007) en het HAN-chassis (Project HAN Onderwijsflexibilisering, 2006). Ook is al veel vooruitgang geboekt in de ontwikkeling van een gemeenschappelijke taal over datgene dat geleerd moet worden, bijvoorbeeld door de formulering van de zeven SBL-competenties en de bijbehorende competentie-indicatoren. Tegelijkertijd zijn deze prachtige visiedocumenten vaak abstract en is de relatie met het dagelijkse onderwijs moeilijk te herleiden. Leerlingen en ouders weten vaak niet eens van het bestaan van deze bronnen. Zelfs binnen een schoolorganisatie worden visiedocumenten niet altijd op dezelfde wijze geïnterpreteerd. Zo heb ik een keer ter voorbereiding van een workshop over beoordelen een aantal passages geselecteerd uit een visiedocument van de hogeschool waar ik de workshop gaf. Deze passages representeerden de visie van de hogeschool op leren en beoordelen, met alle moderne begrippen die daarbij horen zoals competentie, zelfsturing, reflectie, summatief en formatief beoordelen, enzovoort. Tijdens de workshop – waaraan een twintigtal medewerkers deelnam – heb ik deze passages voorgelegd en gevraagd deze individueel te vertalen naar ‘Jip en Janneke taal’. Dit viel niet mee. Bij de nabespreking bleek het niet alleen moeilijk om het in duidelijke taal om te zetten, ook bleken er behoorlijke verschillen te bestaan in de interpretatie van de visie. Het zou interessant zijn een zelfde oefening te doen binnen de context van de Faculteit Educatie. Ook ligt er nog een uitdaging om visiedocumenten te koppelen aan de beoordelingstaken zoals die op het niveau van de leerling op een vraaggestuurde wijze worden uitgevoerd. Tot slot is het van belang dat in de communicatie van het beoordelingsmodel de kwaliteit van duurzaam beoordelen wordt geborgd.

Het openbreken van de ‘black-box’

Zeker als het gaat om complex leren – waarin niet altijd sprake is van een eenduidig goed antwoord – is inzicht in het beoordelingsproces van belang. Nu is het vaak voor de leerling niet meer te achterhalen wat de individuele gezichtspunten van beoordelaars waren om tot een bepaalde uitspraak te komen. Juist deze gezichtspunten kunnen verhelderend en leerrijk zijn voor de leerling (Delandshere & Petrovsky, 1998; Johnston, 2004). Leraren zouden explicieter hun beoordelingsaanpak kunnen communiceren naar leerlingen, zodat

zij beter begrijpen waarop en hoe zij worden beoordeeld. Of zoals Deutsch het treffend verwoordt:

Students are in a bewildering position if a teacher marks them without telling them in sufficient detail the values, rules, and procedures employed in his or her grading. In such a situation, the mark-oriented students are necessarily anxiously dependent on the teacher's approval, since they have no other basis for guiding their behaviour to achieve merit... Where the instructor is explicit in his or her style of grading, the student can be more independent of the teacher. (p.396)

Het expliciteren van de beoordelingscriteria en standaarden en het geven van voorbeelden kan een waardevolle activiteit zijn om leerlingen bewust te maken wat en waarom zij iets moeten leren (Nicol & Macfarlane-Dick, 2006). Onderzoek waarin leerlingen werden betrokken bij het interpreteren en definiëren van beoordelingscriteria heeft aangetoond dat leerlingen hier ook daadwerkelijk beter door gaan presteren (Sluijsmans en anderen, 2002; Sluijsmans en anderen, 2003).

Ook kan ik mij voorstellen dat er sterker wordt geïnvesteerd in voortgangsystemen of –dossiers, die te allen tijde in te zien zijn door ouders, leerlingen en leraren met bijvoorbeeld een inlog-code. Vooral ouders kunnen dan tijdig aan de bel trekken als zij vragen hebben en zullen minder worden ‘overvallen’ bij een periodiek voortgangsgesprek. In dit perspectief zie ik ook een belangrijke verantwoordelijkheid en betrokkenheid voor ouders weggelegd.

Terwijl ik pleit voor het openbreken van de ‘black box’ in het beoordelen, besef ik dat complexe beoordelingsprocessen vaak vanwege onduidelijke redenen niet transparant zijn. De voorbeelden liggen voor het oprapen. Waarom wordt het ene subsidievoorstel wel gehonoreerd en het andere niet? Waarom kiest de internationale turnbond ervoor het laatste startbewijs voor de Olympische Spelen toe te kennen aan Nashwan al-Harazi uit Jemen en niet aan Yuri van Gelder? Waarom bleef dat vriendinnetje van mij dat toch altijd al veel slechtere rapporten had niet zitten en ik wel? Ondanks deze wetenschap ben ik er van overtuigd dat er veel te winnen is als de informatie achter een oordeel bekend is. Zeker als het gaat om *verder leren*.

Naar diversiteit in de communicatie over beoordelingen

Het derde probleem richtte zich op het informatie-arme cijfersysteem dat nog dikwijls de boventoon voert in huidige beoordelingsmodellen. Al een dikke eeuw geleden vonden velen dat het cijfersysteem te ver af stond van het echte leren en onderwijzen. Sommige scholen gingen weer terug naar de meer narratieve rapporten. Anderen maakten alleen nog maar het onderscheid tussen zakken en slagen. Weer anderen stelden het idee van

‘beheersingsleren’ (*mastery learning*) in, waarbij men er van uitgaat dat 95% van de leerling-populatie de leerstof kan beheersen, mits ze maar voldoende tijd krijgt. Formatieve beoordelingen met veel feedbackloops worden hierbij als essentieel gezien (Bloom, Hastings, & Madaus, 1971).

Communicatie in duurzaam beoordelen betekent dat de leraar niet alleen het *resultaat* van een beoordeling communiceert (*Knowledge of Results*; voldoende, cijfer, geslaagd, goed gedaan, enzovoort), maar dat deze communicatie de leerling handvatten biedt voor verdere ontwikkeling (Delandshere & Petrovsky, 1998; Mory, 2004). Deze communicatie kan worden gezien als de lijn die de beoordelingstaken als weergegeven in Figuur 4 met elkaar verbindt. Om duurzaam beoordelen te realiseren zal de communicatie gericht zijn op het beantwoorden van de volgende drie vragen:

1. Waar gaat de leerling naar toe (wat zijn de geselecteerde beoordelingscriteria, standaarden en condities)? Dit kan worden benoemd als de *feed-up*;
2. Hoe heeft de leerling de taak uitgevoerd (welke vooruitgang wordt geboekt ten aanzien van de beoordelingscriteria en standaarden)? Dit kan worden benoemd als de *feed-back*;
3. Hoe gaat de leerling verder (welke aanpak is nodig om tot groei te komen)? Dit kan worden benoemd als de *feed-forward*.

Deze drie vormen van communicatie worden als de belangrijkste componenten gezien in beoordelingen die in functie staan van verder leren (Black & Wiliam, 1998; Taras, 2005). Het overbruggen van de kloof tussen wat de leerling 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 naar en met de leerling (Hattie & Timperley, 2007; Mory, 2004). Om de communicatie te verhelderen kunnen de drie bovenstaande vragen vervolgens betrekking hebben op vier aspecten van de uitvoering van de beoordelingstaak: 1) op de beoordelingscriteria en standaarden (‘je moet meer staartdelingen oefenen met getallen boven de 1000’); 2) op het proces van probleemoplossen (‘Ik zie in de uitwerking van de staartdelingen dat je steeds vergeet de tientallen af te trekken. Ik denk dat het beter werkt als je het volgende probeert...’); 3) op het niveau van het zelfsturingproces (‘je weet hoe je tientallen moet aftrekken, ga eens na in je uitwerking waar dit goed gaat en waar niet’); en op 4) het niveau van de leerling (‘die staartdeling heb je prima opgelost, goed gedaan!’). Communicatie over beoordelingen is het meest effectief wanneer leerlingen 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 medeleerlingen (Crooks, 1988). Uit onderzoek blijkt dat vooral de laag presterende leerlingen baat hebben bij gedifferentieerde, specifieke en geïndividualiseerde manieren van communiceren over resultaten op beoordelingstaken (Oren, 1983).

Ook is het interessant te onderzoeken welke wijze van communicatie over beoordelingen voor welke type leraar en leerling het meeste zinvol en motiverend is. Wanneer de communicatie over beoordelingen niet aansluit op de behoefte van de individuele leerling zal dit een negatieve invloed hebben op verder leren en motivatie. Recent onderzoek van Gerda Geerdink laat bijvoorbeeld zien dat jongens en meisjes op de pabo beoordelen op een hele andere manier benaderen. Jongens willen vooral een hard cijfer, terwijl meisjes een meer narratieve benadering in de communicatie als prettiger ervaren (Geerdink, 2007). Zo zullen er ook verschillen zijn in de manier waarop leraren over beoordelingen kunnen en willen communiceren. De concepties over leren en beoordelen (zie ook Speerpunt 3) zullen in belangrijke mate bepalen welke manieren van communicatie over beoordelen tot verder leren zullen leiden.

Kwaliteit van leren: Onderzoeksvragen

Rijke communicatie over het ontwerp, de uitvoering en de resultaten van beoordelingstaken wordt in het lectoraat als een tweede voorwaarde voor duurzaam beoordelen beschouwd. Als de leerling onvoldoende informatie krijgt over zijn of haar beoordeling, kan dit de beklijving van kennis en kunde in de weg staan. In het lectoraat worden projecten geëntameerd die (het effect van) rijkere vormen van communicatie over beoordelingen tussen leraren en leerlingen, leraren en ouders en leraren onderling nader onderzoeken. Huidige wetenschappelijke inzichten over feedback en effecten op leren zijn daarbij ondersteunend. Mogelijke onderzoeksvragen zijn:

- Hoe worden door scholen en opleidingen huidige beoordelingsmodellen gecommuniceerd naar leraren, leerlingen en ouders?
- Communiceren leraren met leerlingen over hun beoordelingsaanpakken, zo ja, hoe doen zij dat, zo nee, waarom doen ze dat niet?
- Wat ervaren leraren en leerlingen als zinvolle communicatie over beoordelingen?
- Zijn er verschillen tussen opleidingen en scholen als het gaat om communicatie over beoordelingen?
- Hoe begeleiden scholen en opleidingen hun leerlingen in de beoordeling, zowel in voorbereiding, uitvoering en rapportage van de resultaten?

SPEERPUNT 3

ZELFSTUREN BIJ BEOORDELEN



Drie problemen

In de beschrijving van de eerste twee speerpunten heb ik het belang van een geïntegreerd onderwijsontwerp van leren en beoordelen en het belang van communicatie over beoordelingen benadrukt. Maar vraaggestuurd leren in optima forma betekent bovenal een actieve betrokkenheid van de leerling, die verantwoordelijkheid voor het eigen leren zal moeten en willen nemen (De Vries, 2007). Daarnaast zal de leerling in het onderwijs voldoende bagage moeten meekrijgen die hem in staat stelt steeds nieuwe situaties in formele en informele contexten het hoofd te bieden (Boud, 2000). De HAN-collega's Bregje de Vries en Gertjan Schuiling stipten in dit verband ook al Martinus Langeveld aan (zie respectievelijk pagina 20 en 28 van hun redes). Langeveld betoogde dat het doel van opvoeding (thuis en op school) is dat de leerling uiteindelijk op een zelfverantwoordelijke manier beslissingen kan nemen, zowel ten aanzien van zichzelf als de maatschappij. Ook Kees Boeke, die in 1926 de Werkplaats Kindergemeenschap oprichtte, was een pionier als het gaat over de verantwoordelijkheid van de leerling voor het eigen leren.

In de historie van het beoordelen is de leerling lange tijd niet als volwaardige partner beschouwd. Het beoordelen was vooral een zaak van partijen buiten de leerling. Om duurzaam beoordelen in vraaggestuurd leren echter te kunnen realiseren wordt het tijd de leerling bewust te maken van de rol die hij kan vervullen in beoordelen. De verantwoordelijkheid en zelfsturing van de leerling is naast ontwerp en communicatie het derde speerpunt. Ook in dit kader noem ik drie obstakels die duurzaam beoordelen belemmeren. Deze zijn 1) het geringe zelfsturende vermogen van leerlingen, 2) de lage betrokkenheid van leerlingen in de beoordeling, 3) het negeren van motivatie, toetsangst, en concepties van leerlingen. Ook dit deel zal ik afsluiten met enkele voorstellen voor onderzoek binnen het lectoraat.



Figuur 6. Robin van Persie als zelfsturende professional

Probleem 1. Het geringe zelfsturende vermogen van leerlingen

De leerling zal in vraaggestuurd leren meer verantwoordelijkheid en betrokkenheid bij leren en beoordelen moeten tonen dan hij misschien gewend was. Er wordt steeds meer van de leerling verwacht dat hij op eigen kracht zijn leerproces kan sturen. Maar wat betekent zelfsturing? Een antwoord op deze vraag vond ik voor een deel in de enthousiaste typering die de journalist Henk Spaan in het programma Pauw en Witteman¹⁰ gaf van de voetballer Robin van Persie (Figuur 6).

Spaan gaf de volgende reactie op de vraag ‘Wat maakt van Persie tot de meest belangrijke voetballer van dit moment?’

‘Hij is de meest talentvolle. Hij kijkt naar zich zelf en naar zijn medespelers. Hij kan met afstand naar zichzelf kijken. Dat vind ik heel goed van een voetballer. (...) Hij weet wat hij moet verbeteren, hij weet hoe hij nog meer rendement uit zijn spel moet halen. Dat vind ik interessant voor een jongen van 24 jaar.’

Het citaat typeert Robin van Persie als een zelfsturende professional die in staat is zijn eigen prestaties te beoordelen om vervolgens tot nog betere prestaties te komen. Dit wordt door Spaan gezien als iets bijzonders voor een jong iemand. En dat is ook bijzonder. Jezelf kunnen aansturen is moeilijk en een vaardigheid die zelfs volwassenen lang niet altijd beheersen. Zelfsturing veronderstelt dat leerlingen in staat zijn te plannen en zichzelf (en anderen) te beoordelen. Helaas is er nog maar weinig empirisch bewijs dat leerlingen in staat zijn hun leerbehoefte goed te definiëren en een beoordeling van zichzelf kunnen geven (Regehr & Eva, 2006). Zelfsturing is moeilijk voor leerlingen die afhankelijk zijn van extrinsieke vormen van beoordeling of beoordelingsmodellen die aangeven welk norm of criterium behaald moet worden. Uit recent peilingsonderzoek¹¹ blijkt bovendien dat zelfsturend leren, zelfstandig in groepjes werken en via internet leren maar bij weinig leerlingen populair is. De voorkeur gaat uit naar veel lessen met duidelijke opdrachten en veel begeleiding. Cijfers zijn in dit opzicht zaligmakend. De directe waardering voor het nu tilt dan zwaarder dan de investering in leren op de lange termijn.

Dat zelfsturing moeilijk is illustreert vooral de literatuur over zelfbeoordeling. Als het gaat om zelfbeoordeling overschatten leerlingen meestal hun prestaties. Zo denkt bijvoorbeeld ook ieder van ons dat we beter autorijden dan de gemiddelde Nederlander, hoewel dit statistisch gezien nooit zou kunnen kloppen (dit is bekend als het ‘*better than average effect*’; Myers, 1980). En is het feit dat je jezelf bescheiden vindt, niet al een teken van

¹⁰ Zie de uitzending van 4 september 2007 op www.pauwenwitteman.nl. In deze uitzending wordt overigens ook de zesjescultuur nog aangestipt in een discussie met Bas Haring, hoogleraar filosofie aan de Universiteit van Leiden.

¹¹ Zie artikel ‘Studenten willen gewoon meer les’ in Elsevier Thema Studeren, pagina 20, waarin verwezen wordt naar het onderzoek ‘De beste studies 2007’ uitgevoerd door (Elsevier/ResearchNed),

onbescheidenheid? Dat leerlingen hun kennen en kunnen overschatten heeft te maken met een gebrek aan kennis over de beoordelingscriteria (Bjork, 1999; Dunning, 2005; Krueger & Dunning, 1999). Hoe kun je immers jezelf beoordelen op kennis waar je de reikwijdte niet van weet? Krueger en Dunning (1999) verwoorden het als volgt (1999):

'People tend to hold overly favourable views of their abilities in many social and intellectual domains (...) This overestimation occurs, in part, because people who are unskilled in these domains suffer a dual burden: Not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it.' (p. 1121).

Omdat incompetentere leerlingen 'niet weten wat ze nog niet weten', leidt een zelfbeoordeling van deze leerlingen niet snel tot frustratie. Dit in tegenstelling tot de sterke leerlingen, die veel meer het besef hebben wat je nog allemaal kunt leren en zich daarom zichzelf juist vaak onderschatten (Sluijsmans, Dochy, & Moerkerke, 1999). De goed presterende leerlingen overschatten bijvoorbeeld ook het niveau van medeleerlingen, hoewel deze medeleerlingen in werkelijkheid minder goed zijn ('*false consensus effect*'; Dunning, 2005).

Probleem 2. De lage betrokkenheid van leerlingen bij de beoordeling

Huidige beoordelingsmethoden zijn overwegend prescriptief van aard. Omdat de school en de leraar bepaalt wat, hoe en wanneer wordt beoordeeld, hebben zij nog steeds de volledige controle over het onderwijsleerproces van de leerling (Broekkamp, 2003). Deze situatie staat op gespannen voet met de uitgangspunten van vraaggestuurd leren, waarin de individuele leerbehoeften leidend zijn. Het eerder aangehaalde voorbeeld van de notulen van de docentenvergadering uit 1968 laat zien dat de leerling niet serieus wordt genomen. De citaten als '*ze leveren alleen maar kritiek*' en '*ze kunnen niet beoordelen*' illustreren dat leraren geen hoge pet op hebben van de beoordelingsvaardigheid van leerlingen. De opgelegde vormen van beoordelen belemmeren duurzaam beoordelen en vraaggestuurd leren omdat het beoordelen niet toebehoort aan de leerling, maar iets is dat de leerling ondergaat (Boud, 2000). Omdat de leerling nauwelijks betrokken wordt bij de voorbereiding en uitvoering van de beoordeling zal het effect van de beoordeling op leren minimaal zijn (Dochy & McDowell, 1997; Gibbs & Simpson, 2004). Onderzoek laat zien dat vooral de beoordelingen die gericht zijn op certificering en selectie negatief zijn voor leren, omdat ze leerlingen belemmeren om verantwoordelijkheid te nemen zich te motiveren en te reguleren (Hattie & Timperley, 2007; Ryan & Deci, 2000). Beoordelen op deze wijze leidt tot 'aangeleerde afhankelijkheid' (Yorke, 2003). De leerling wordt een gevoel van eigenaarschap over het leren ontnomen (De Vries, 2007).

De laatste jaren zijn vele initiatieven ondernomen om leerlingen meer te betrekken bij

beoordelen door vormen als zelf-, peer- en portfoliobeoordelingen (Sluijsmans, Dochy, & Moerkerke, 1999; Sluijsmans, Strijbos, & Van de Watering, 2007). Maar helaas gaat het ook hier vooral om vormen van beoordelen die de leerling zijn opgelegd. Zo wordt per leraar vastgelegd wanneer de reflectierapporten of het portfolio moeten worden geschreven, in hoeveel woorden en volgens welke structuur. Ook in de opleidingen van de HAN zien we dat het invoeren van zelfbeoordeling op deze manier eerder leerlingen demotiveert en leidt tot ‘reflectie-moeheid’. Reflecteren wordt een routine en leerlingen schrijven op wat de leraar graag in het reflectieverslag terug wil zien (Sluijsmans, Brand-Gruwel, Van Merriënboer, & Bastiaens, 2003). Bij peer-beoordeling is de betrokkenheid van leerlingen tevens beperkt als het gaat om vraagsturing. Hoewel onderzoek positieve effecten laat zien van training in peer-beoordeling op motivatie en leren (Sluijsmans, 2002; Sluijsmans, Brand-Gruwel, Van Merriënboer, & Martens, 2004; Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002), is nog weinig onderzoek bekend naar peer-beoordelingen die gericht zijn op vraaggestuurd leren. Omdat criteria al op voorhand vastliggen als ook de wijze waarop de peer-beoordeling moet plaatsvinden, is de aanvankelijk gesuggereerde betrokkenheid van leerlingen bij nader inzien beperkt.

Bij portfolio beoordeling is er vaak sprake van een mismatch tussen de *functie* van het portfolio en de *vorm* waarin deze vorm van beoordeling wordt gegoten. Onderzoek heeft aangetoond dat het portfolio vaak teleurstellend leeg blijft totdat de opgelegde deadline zich aandient (Van Tartwijk, Driessen, Van der Vleuten, & Stokking, 2008). Zolang de functie van het portfolio niet overeenkomt met de inhoud en de structuur ervan, leraren niet voldoende ondersteund worden in de begeleiding van portfoliobeoordeling en er geen juiste infrastructuur is, zal ook het portfolio zijn doel missen.

Samenvattend heeft de leerling doorgaans weinig invloed op zowel het ontwerp, de uitvoering als de communicatie van beoordelingen, waardoor zelfsturing en het nemen van verantwoordelijkheid nauwelijks wordt gestimuleerd. Beoordelen zal voor leerlingen pas zinvol zijn voor leren als leerlingen actief zullen worden betrokken in de communicatie over beoordelen (zie ook Speerpunt 2). Ook zal de wijze van beoordelen meer aan kunnen sluiten bij leerbehoeften van leerlingen. Op dit moment zijn beoordelingen overwegend gericht op de verbale dan wel schriftelijke vaardigheden van leerlingen. Omdat de vrijheid van leerlingen in de manier waarop ze beoordeeld worden beperkt is, blijft weinig ruimte over voor de meer creatievere vormen van beoordeling (De Vries, 2007). Zolang leerlingen niet zich eigenaar voelen van het beoordelingsproces zal duurzaam beoordelen niet haalbaar blijken.

Probleem 3. Het negeren van motivatie, toetsangst en concepties van leerlingen

Een ‘opgelegde’ extern gestuurde wijze van beoordelen leidt tot demotivatie en daarmee logischerwijs tot een extrinsiek gestuurde wijze van leren. Zoals we weten stuurt de wijze van beoordelen vaak het studeergedrag van leerlingen. Beoordelen dat wordt benaderd als een aparte tak in het onderwijs, belemmert leerlingen in hun ‘flow’. Flow – als

belangrijk concept in de ‘positieve psychologie’ – wordt gekenmerkt door op de desbetreffende bezigheden gerichte energie, volledige betrokkenheid, en succesvolle uitvoering (Csikszentmihalyi, 1990) of het spelend leren zoals dat in diverse lezingen wordt gezien als belangrijke conditie voor leren (De Vries, 2007; Martens, 2007). Voornamelijk normgerichte beoordelingen hebben een negatieve invloed op motivatie en individuele ontwikkeling. Erkenning van de invloed die beoordelen heeft op de motivatie en zelfwaarde van leerlingen én leraren is cruciaal voor duurzaam beoordelen in vraaggestuurd leren (Black & Wiliam, 1998). Als de leerling de kloof tussen wat hij moet leren volgens de school en datgene wat hij al weet en kan als te groot ervaart zal ook de motivatie om te leren verdwijnen (Sadler, 1989; Martens & Boekaerts, 2008). Ook de tegenovergestelde situatie bestaat: als leerlingen 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. Tegenwoordig zijn we bang voor een film van Wilders, de klimaatverandering of een onderwijsvernieuwing. Vanuit het perspectief van de leerling kan zelfs een vraag als ‘*Hoeveel is 141 – 56*’ dezelfde angst oproepen. Toen ik de eerste keer het Gymnasion binnenliep voor mijn sollicitatiegesprek stond de hal vol met studenten, van wie een aanzienlijk deel duidelijke signalen van nervositeit toonde. Het nog snel inkijken van het boek of aantekeningen, de rode blosjes op de wangen en het zenuwachtig heen en weer geloop. Je zag zonder het te weten dat een tentamen op komst was.

De invloed van toetsangst op de prestatie en motivatie van leerlingen wordt nog vaak onderschat. Nogal schokkend is een studie die aantoonde dat rond examentijd gemiddeld meer familieleden overlijden als gevolg van de ervaren examenstress (Adams, 1999). Daarnaast heeft onderzoek ook 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). Ook blijkt de hartslag van leerlingen met toetsangst toe te nemen naarmate de taken moeilijker worden (Ashcraft & Kirk, 2001). In het wiskundedomein wordt toetsangst erkend, zoals blijkt uit testen als de Mathematics Anxiety Rating Scale (MARS; Richardson & Suinn, 1972), waarmee gemeten kan worden hoe erg het is gesteld met de ‘wiskundeangst’. Angst voor beoordelingen blijkt ook samen te hangen met de manier waarop de leraar de leerlingen ondersteunt tijdens de lessen.

Leraren die vooral gericht zijn op de correctheid van een antwoord, zonder het bieden van motivationele ondersteuning of uitleg, werken angst in plaats van vaardigheid in de hand. Tot slot hangt angst samen met de zelfwaarde van leerlingen. Het blijkt dat vooral normgerichte beoordelingsmodellen minder goed werken voor leerlingen met een lage zelfwaarde dan een ipsatief beoordelingsmodel (Bolocofsky & Mescher, 1984).

Variabelen als motivatie en angst hangen samen met de opvattingen die een leerling heeft over leren en beoordelen. Deze concepties zijn bepalend voor het gedrag dat leerlingen vertonen in onderwijssituaties. Zo zal de leerling die altijd als laatste werd gekozen bij het vormen van teams bij gym, met een andere opvatting over gymnastiekonderwijs de lerarenopleiding binnenstromen dan de leerling die altijd uitblonk in sport.

Helaas weten we nog weinig over de concepties van leerlingen over beoordelen. Onderzoek naar concepties over beoordelen bij leraren richt zich bovenal maar op een beperkt aantal aspecten van beoordelen (Brown, 2004; Gullickson, 1984). Ook ontbreekt het aan instrumenten of kaders die leerlingen en leraren helpen hun concepties over leren en beoordelen te expliciteren. Wel is een en ander bekend over hoe leerlingen zich voorbereiden op een beoordeling (zie Broekkamp & van Hout-Wolters, 2007 voor een interessant overzicht). Het blijkt dat beoordelingen gericht op feiten en kennis bij leerlingen strategieën als reproduceren stimuleren (Crooks, 1988; Scouller, 1998; Scouller & Prosser, 1994; Tang, Lai, Arthur, & Leung, 1999). Deze reproductiegerichte aanpak, waarin de leerling zich voornamelijk richt op ‘stampen’ en verder geen betekenis geeft aan de taak, staat tegenover een aanpak die zich richt op diepteverwerking. Leerlingen die beoordelingstaken voorbereiden volgens deze aanpak verdiepen zich in de taak zonder daarbij afhankelijk te zijn van een beloning. Een studie van Nijhuis en anderen (2008) toonde aan dat de aanpakken van leerlingen bij het voorbereiden en uitvoeren van beoordelingstaken sterk variëren en dat een overgang van een reproductiegerichte aanpak naar een aanpak gericht op diepteverwerking lang niet altijd plaatsvindt. De oorzaak kan ermee te maken hebben dat leerlingen ervaren dat per leraar en dus per beoordelingstaak een andere aanpak wordt verwacht (Nijhuis, Segers, & Gijselaers, 2008).

Naar meer zelfsturing bij beoordelen door de leerling

Een hoog zelfsturend vermogen zie ik naast het ontwerp van beoordelen en de communicatie over beoordelen als derde mogelijke succesfactor voor duurzaam beoordelen in vraaggestuurd leren. Omdat de leerling de enige is die zichzelf ‘meeneemt’ en zichzelf voortdurend kan ‘zien’ en ook de kennisbronnen en omgevingen zo wisselend zijn, zal hij in staat moeten zijn leren te reguleren, met andere woorden: hij moet leren slim te leren (Broekkamp, 2003; Korthagen, 1998). Dit betekent dat hij beoordelingstaken zorgvuldig plant en uitvoert, om vervolgens te kunnen komen tot een nieuwe keuze in beoordelingstaken. Het betekent ook de verantwoordelijkheid nemen in het zoeken naar, in plaats van wachten op feedback en het zelf kunnen vaststellen van beoordelingscriteria en standaarden (Boud, 2000; Kremer-Hayon & Tillema, 1999; Van Merriënboer & Sluijsmans, 2008). Graag stel ik enkele wegen voor die zelfsturing en verantwoordelijkheid bij beoordelen kunnen bevorderen. Deze zijn het integreren van zelfsturing in het ontwerp van beoordelingstaken, het verhogen van de betrokkenheid van de leerling in het beoordelen en het erkennen van individuele verschillen tussen leerlingen.

Het integreren van zelfsturing in het ontwerp van beoordelen

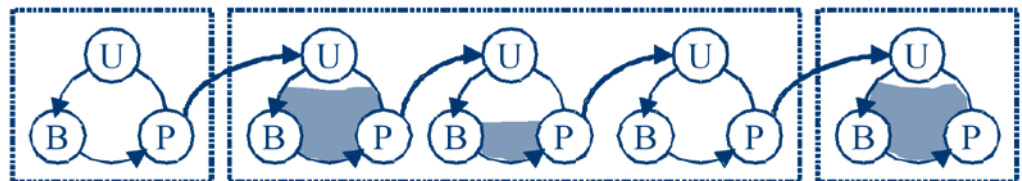
In het eerste probleem schetste ik het gebrek aan zelfsturend vermogen van leerlingen. Enerzijds wordt dit veroorzaakt door een gebrek aan motivatie zelfsturend te willen leren,

anderzijds ontbreekt het aan vaardigheden die zelfsturing mogelijk maken. Het (leren) vormgeven aan het eigen leren is voor duurzaam beoordelen in vraaggestuurd leren echter essentieel. Ook zonder de school zal de leerling in staat moeten zijn verder te leren in formele en informele contexten. Voor leerlingen betekent dit dat zij niet langer meer achteroverleunend het onderwijs consumeren, maar dat zij zich veel meer opstellen als ‘prosumers’, die mede vormgeven aan het leren en beoordelen door op eigen initiatief informatie over hun prestaties op te zoeken (Martens, 2007). Om zelfsturing te integreren in het leren op een wijze die én motiverend is én bijdraagt tot een zelfsturend vermogen, is het zinvol aan te sluiten bij het eerste Speerpunt, 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; Ertmer & Newby, 1996; Kremer-Hayon & Tillema, 1999; Schunk, 1990; Van Merriënboer & Sluijsmans, 2008; Zimmerman, 1990). Terwijl in het ‘reguliere’ onderwijs alleen het uitvoeren van beoordelingstaken aan de leerling wordt overgelaten en de planning en beoordeling aan de leraar (zie Figuur 7a), zal in zelfsturend leren de leerling (mede)verantwoordelijk zijn voor de planning en beoordeling. Het uiteindelijke zelfsturend leren zal steeds verlopen in een cyclisch proces van plannen, uitvoeren en beoordelen van de beoordelingstaak (zie Figuur 7b).



(a)



(b)

Figuur 7. Ordening van taken voor het bevorderen van leren (uitvoeren van beoordelingstaken) (a) en zelfsturend leren (b) (uitvoeren, beoordelen en plannen van beoordelingstaken)

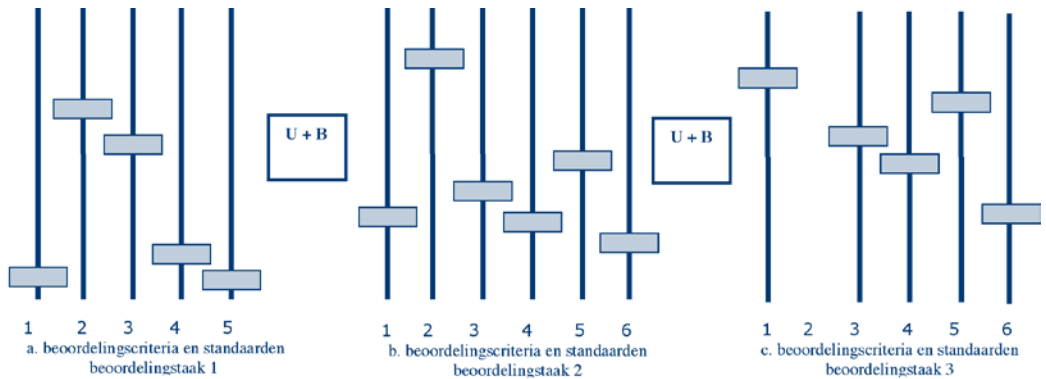
Het is niet realistisch dat leerlingen als vanzelf hun leerproces kunnen reguleren. Een tweede consequentie in het ontwerp van zelfsturend leren is dat de ondersteuning die de leerling wordt geboden zich niet beperkt tot de beoordelingstaak, maar dat ook ondersteuning wordt geboden in het leren plannen en beoordelen. Dezelfde ontwerpprincipes voor het uitvoeren van beoordelingstaken kunnen dan ook worden toegepast op het leren plannen en beoordelen (Van Merriënboer & Sluijsmans, 2008). Zo zal de leraar aanvankelijk uitgewerkte voorbeelden geven van het proces van plannen en beoordelen, maar zal de leerling geleidelijk aan meer controle geven over de planning en beoordeling van taken. Controle over het eigen leren heeft in een aantal onderzoeken in het sportdomein al zijn meerwaarde bewezen (Wulf & Toole, 1999; Chiviacowsky & Wulf, 2002; 2005; Wulf et al., 2005).

Omdat zelfsturing in vraaggestuurd leren gericht is op individuele leervragen, kunnen leerlingen worden gemotiveerd de optimale leeromgeving voor zichzelf en hun medeleerlingen te creëren. Een voorbeeld zijn zelfsturende studieteams, waarin leerlingen relevante taken selecteren en deze taken bestuderen vanuit een reflectieve en samenwerkende onderzoekshouding (Kremer-Hayon & Tillema, 1999).

Het verhogen van de betrokkenheid van de leerling bij het beoordelen

Het tweede probleem betrof de constatering dat leerlingen in het algemeen weinig betrokkenheid bij beoordelen hebben. Het geven van controle aan leerlingen 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 leerlingen. 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 ‘*zelfdeterminatie-theorie*’ (Ryan & Deci, 2000), waarin competentie, autonomie en sociale verbondenheid de basisbehoeften zijn. Competentie betekent het gevoel van zinvol bezig zijn, autonomie verwijst naar de afwezigheid van voortdurende externe sturing, en sociale verbondenheid doelt op het creëren van een vertrouwde leeromgeving. Het ontbreken van één van deze basisbehoeften kan de motivatie dusdanig verstoren dat leren niet plaatsvindt.

Figuur 8 illustreert een mogelijk scenario van vraaggestuurd leren, waarin de leerling in dialoog met de school en/of opleiding op een zelfsturende wijze beoordelingstaken en bijbehorende beoordelingscriteria en standaarden selecteert, taken uitvoert en deze beoordeelt.



Figuur 8 a, b, c. Het kiezen van beoordelingscriteria en het vaststellen van standaarden (planning) op basis van de beoordeling (B) op taakuitvoering (U)

In de eerste beoordelingstaak heeft de leerling vijf criteria gekozen waarbij de standaarden op criterium 1, 4 en 5 aanvankelijk vrij laag worden gezet (Figuur 8a). Op basis van de uitvoering van de taak en beoordeling hiervan past de leerling zijn keuze aan door een zesde criterium toe te voegen en sommige standaarden te verhogen (Figuur 8b). Steeds observeert, analyseert en beoordeelt de leerling zijn eigen gedrag op basis van de gekozen set aan beoordelingscriteria en kan aangeven waarop hij zich wil verbeteren. Hierbij kan de leerling hulp vragen aan zijn omgeving, zoals medeleerlingen, leraren of mentoren in de school en/of opleiding (Brett & Atwater, 2001). Het proces wordt vervolgens herhaald voor de derde beoordelingstaak, waar de leerling besluit het tweede criterium (Figuur 8c) buiten beschouwing te laten, omdat dit criterium bijvoorbeeld al op een voldoende standaard is gedemonstreerd.

Het geschetste scenario dient vooral als een denkkader om leerlingen te betrekken bij de keuze van beoordelingstaken, criteria en standaarden op een manier dat individuele groei en een cyclisch beoordelingsproces wordt gerealiseerd (zoals weergegeven in Figuur 7). 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. Ook kan de leerling zelf beoordelingstaken definiëren, bijvoorbeeld door het inbrengen van stage-ervaringen.

Tiger Woods (Figuur 9) is een voorbeeld van een zelfsturende professional die goed een dergelijk mengpaneel kan bedienen. Toen hij een aantal jaren geleden merkte dat bepaalde technieken niet meer op voldoende niveau waren en hij zelfs nieuwe technieken moest leren, heeft hij een wereldkampioenschap moeten overslaan. Na dit jaar kwam hij sterker dan ooit terug.



Figuur 9. Tiger Woods als zelfsturende professional

Topporters geven het beste voorbeeld als het gaat om het in praktijk brengen van een ipsatief beoordelingsmodel gericht op duurzaam leren waarin zelfsturing centraal staat. Ze laten ons zien dat het mogelijk is de tijd te nemen om te kunnen leren en dat dit ook betekent dat andere zaken moeten wijken (op stap gaan, sociaal leven, andere interesses). Alle deelnemers aan de Olympische Spelen die over een paar weken van start gaan zijn prachtige voorbeelden hoe je door volharding, motivatie en de wens ‘de beste te zijn’ jezelf op een steeds hoger niveau kunt tillen.

Het erkennen van individuele verschillen

Om duurzaam beoordelen en vraaggestuurd leren te kunnen realiseren is rekening houden met verschillen tussen leerlingen een essentiële voorwaarde. Het derde probleem illustreerde een aantal consequenties van het negeren van enkele variabelen zoals motivatie, toetsangst en de concepties die leerlingen hebben over leren en beoordelen. Door leerlingen 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 ze past (Joosten, Sluijsmans, Brand-Gruwel, & Jochems, 2008). Ertmer en Newby (1996) bieden een zinvol overzicht van vragen waarmee de leerling samen met medeleerlingen en leraren in kaart kan brengen wat de individuele behoeften zijn als het gaat om het plannen, uitvoeren en beoordelen van beoordelingstaken. De antwoorden op dergelijke vragen kunnen niet alleen leidend zijn in het maken van keuzes in het leerpad, ze geven de leerling ook het gevoel van betrokkenheid.

Tabel 1. Richtvragen voor de leerling per fase in zelfsturend leren

Fase	Aard	Vraag
Planning	Cognitief	Welke criteria en standaarden stelde ik in mijn vorige taak (opdracht, stageles, project, enz.) en wat wordt mijn volgende doel? Welke aanpak is het meest effectief om dit doel te bereiken? Wat weet ik al over de inhoud van dit gedragsdoel (snap ik de criteria, weet ik hoe ik een standaard kan bepalen?)
	Motivationeel	Zal het werken aan de criteria en standaarden veel tijd en energie kosten? Hoe kijk ik tegen de gestelde criteria en standaarden aan? Lijkt het me leuk hieraan te werken?
	Omgeving	Wat zouden goede leercondities kunnen zijn om de criteria en standaarden te kunnen bereiken (bijvoorbeeld workshops volgen, literatuur lezen, advies vragen aan leraar, medeleerling/ouderejaars) Waar en wanneer leer ik het meest effectief en efficiënt? En hoe verhoudt zich dat tot het gestelde criteria en standaarden?
Uitvoering	Cognitief	Is de aanpak die ik heb gekozen goed gekozen? Is er iets veranderd in de gestelde criteria en standaarden? Begrijp ik wat ik aan het doen ben? Boek ik vooruitgang?
	Motivationeel	Houd ik mijn aandacht erbij? Vind ik het nog interessant genoeg? Heb ik vertrouwen dat het mij gaat lukken?
	Omgeving	Hoe stimulerend is de studeeromgeving? Welke extra materialen/bronnen heb ik nodig? Geef ik mezelf voldoende tijd?
Beoordeling	Cognitief	Blijkt dat de aanpak die ik heb gekozen goed gekozen is? Wat heb ik gedaan als een aanpak niet werkte? Wanneer kan ik de gekozen aanpak nog meer gebruiken? Hoe kan ik mijn aanpak verbeteren? Heb ik mijn doel bereikt? Wat heb ik geleerd over het doel? Welke criteria en standaarden heb ik nu bereikt en onder welke condities?
	Motivationeel	Hoeveel moeite kostte het? Hoe bleef ik gemotiveerd? Hoe kijk ik tegen de uitkomst aan?
	Omgeving	Kwam ik onverwachte obstakels tegen? Hoe ben ik met deze obstakels omgegaan? Hoe heb ik mijn studeeromgeving ingericht (welk onderwijs gevolgd, welke feedback gevraagd, e.d.)? Heb ik het juiste moment gekozen aan deze criteria en standaarden te werken?

Omdat de vragen zich op cognitieve, motivationele en omgevingsaspecten richten, kan voldoende rekening worden gehouden met de unieke behoeften van de leerling.

Kwaliteit van leren: Onderzoeksvragen

In dit derde speerpunt heb ik getracht het belang van zelfsturing en verantwoordelijkheid van de leerling in het licht van duurzaam beoordelen te verduidelijken. Vanaf het moment dat leerlingen er aan deelnemen zou het onderwijs meer gericht moeten zijn op het begeleiden en opleiden van zelfsturende individuen, die zich bewust zijn van hun unieke kwaliteiten en in staat zijn de juiste strategieën toe te passen om verder te leren (Ertmer & Newby, 1996; Zimmermann, 1990). De mate waarin leerlingen worden gestimuleerd tot zelfsturend leren wordt tegenwoordig als een belangrijk kwaliteitscriterium gezien (Baartman et al., 2006). Om een leven lang te kunnen leren moet een leerling immers kennis hebben over zichzelf (wat motiveert mij, hoe leer ik?), over de taak (wat weet ik over deze taak?) en over een scala aan strategieën om taken goed te kunnen uitvoeren (hoe pak ik deze taak aan?).

Omdat leerlingen niet van meet af aan zelfsturend kunnen zijn moedigt het lectoraat projecten aan die de ontwikkeling van volledige sturing vanuit de school en de leraar naar volledige sturing door de leerling onderzoeken. Ook onderzoek naar vraaggestuurde vormen van zelf- en peerbeoordelingen is welkom. Onderzoek dat al een eerste aanzet geeft tot meer flexibiliteit in zelf- en peer-beoordelingen en de wijze waarop deze beoordelingen kunnen worden vormgegeven kunnen hiervoor worden gebruikt (Sluijsmans & Prins, 2006; Sluijsmans, Strijbos, & Van de Watering, 2007; Strijbos, Narciss, & Dünnebiel, 2007). Recente studies over de controle van leerlingen over de selectie van taken (Corbalan-Perez, 2008) zijn zinvol voor het vormgeven van onderzoek naar controle over de selectie van beoordelingscriteria en standaarden. In dit lectoraat wordt verkend in hoeverre het geven van meer controle aan leerlingen in het beoordelen kan worden vormgegeven en welke effecten dit heeft op leren en motivatie. In dit kader is onderzoek naar proceswerkbladen waarin vuistregels voor het proces van plannen, uitvoeren en beoordelen zijn opgenomen interessant. Tot slot is het zinvol om binnen elke fase in zelfsturing specifiek te kijken naar de cognitieve en motivationele aspecten die zelfsturing beïnvloeden. En naar de cognitieve processen die plaatsvinden in het brein van mensen tijdens beoordelingen, met name bij leerlingen met faalangst. Samenvattend kunnen de volgende onderzoeksvragen relevant zijn:

- Hoe kan zelfsturing in het ontwerp van beoordelingstaken worden geïntegreerd?
- Hoe kan een verschuiving worden bewerkstelligd van volledige sturing door de leraar naar volledige zelfsturing?
- Welke ondersteuning is nodig om leerlingen te leren plannen en beoordelen?
- Hoe beïnvloeden cognitieve, motivationele en omgevingsaspecten van de leerling de mate van zelfsturing?
- Welke rol kunnen scholen en opleidingen spelen in het stimuleren en realiseren van zelfsturing?

EEN VOORBEELD

Nu mijn drie centrale speerpunten voor het voetlicht zijn gebracht wil ik graag een voorbeeld presenteren waarin deze speerpunten al op een voorzichtige wijze zijn te herkennen. Als het gaat om de vormgeving van een authentieke leeromgeving waarin leren en beoordelen zijn geïntegreerd, er een heldere communicatie is over beoordelingen en de studenten een hoge mate van verantwoordelijkheid en zelfsturing moeten demonstreren, is het spel van GIMMICS¹² vermeldenswaard. In dit spel, dat oorspronkelijk is uitgewerkt door de afdeling Pharmacie van de Rijksuniversiteit van Groningen, runnen studenten in teams van 5 à 6 personen gedurende vier weken dagelijks van 09.00-17.00 uur hun eigen apotheek in Pildorp (Figuur 10). De apotheek behartigt daarbij ongeveer 2000 patiënten. Zowel kennis als vaardigheden worden continu beoordeeld. Elk facet van het beroep kan aan de orde komen en zowel de moeilijkheid en ondersteuning van taken, in dit geval casussen, kunnen worden aangepast aan de studenten. Wanneer casussen goed worden uitgevoerd krijgt de apotheek er virtuele patiënten bij, anders blijven de patiënten weg. In het concept is bewust gekozen geen directe correctieve feedback te geven om kritische zelfbeoordeling en dus zelfsturend vermogen van de studenten te stimuleren. Omdat het werken aan relevante beoordelingscriteria (gericht op kennis, vaardigheden, attitudes) volledig is ingebed in 'hele taken' wordt integratie van leren en beoordelen gewaarborgd.



Figuur 10. Pildorp: een voorbeeld van een authentieke omgeving voor complex leren¹³

¹² Groningen Institute Model for Management in Care Services, zie <http://www.gimmics.nl/>

¹³ Afbeelding afkomstig van <http://www.gimmics.nl/groningen/index.html>

ONDERZOEK BINNEN HET LECTORAAT

Ik heb u drie speerpunten gepresenteerd die de basis vormen van het lectoraat: het ontwerp van onderwijs waarin vraaggestuurd leren en duurzaam beoordelen worden geïntegreerd, de rijke communicatie over beoordelen en de verantwoordelijkheid en zelfsturing van de leerling. De beschrijving van elk speerpunt heb ik afgesloten met een paragraaf ‘*Kwaliteit van leren*’ waarin ik een aanzet heb gegeven voor een aantal onderzoeksvragen. De inbedding in het gelijknamige onderzoeksprogramma van de Faculteit Educatie is daarmee evident.

Als lector hoop ik de betrokkenheid van scholen, opleidingen en ouders bij beoordelen te verhogen, door meer ruimte en rust te creëren voor reflectie op de huidige beoordelingsmodellen. Omdat ik duurzaam beoordelen vooral benader als een ontwerpvoorbeeld (ontwerpen van beoordelingstaken, ontwerpen van rijke communicatie, ontwerpen van zelfsturing), zullen de onderzoeksactiviteiten vooral een ontwerpgericht karakter hebben. In dit type onderzoek worden leraren en leerlingen in staat gesteld om reflectief en in samenwerking met elkaar in hun dagelijkse leer- en werkomgeving nieuwe ontwerpen te ontwikkelen, uit te proberen in de onderwijspraktijk en vervolgens de effecten te onderzoeken. Op basis van de resultaten kan het ontwerp verder worden aangepast of verbeterd. Ontwerpgericht onderzoek sluit tevens goed aan bij de kerntaken van de leraar, voor wie het ontwerpen van onderwijs in vraaggestuurd leren een steeds belangrijker taak wordt (De Vries, 2007). Omdat echter voor het vergaren van betrouwbare en valide onderzoeksresultaten een mix van onderzoeksmethoden wenselijk is (Schuwirth & Van der Vleuten, 2004), worden niet alleen projecten met een ‘design-based’ karakter aangemoedigd, maar ook projecten die een meer (quasi)experimenteel karakter hebben. Het onderzoek zal een tweeledig doel dienen. Enerzijds zullen projecten worden opgezet die leerlingen en leraren inzicht en instrumenten bieden de speerpunten van deze rede handen en voeten te geven in de dagelijkse praktijk. Anderzijds zullen projecten worden opgezet die gericht zijn op het ontwikkelen van een onderzoekende houding en onderzoeksvaardigheden bij leerlingen en leraren. Om onderzoek en de speerpunten van het lectoraat sterk met elkaar te verbinden kunnen dit projecten zijn waarin leerlingen en leraren leren hoe zij hun ‘leven lang leren’. Zo zijn we bijvoorbeeld gestart met het vormgeven van een professionaliseringslijn bij Pabo Groenewoud in Nijmegen. Een interdisciplinair team zal het komende schooljaar actief participeren in deze professionaliseringslijn en activiteiten definiëren op het niveau van de onderwijsorganisatie, het niveau van de leraar en het niveau van de leerling. Voorbeelden van activiteiten zijn de verdere visievorming over de waarde van duurzaam beoordelen in relatie met opleidingsscholen (onderwijsorganisatie), (leren) doen van eigen onderzoek en reflectie op beoordelingspraktijk (leraar), het begeleiden van leerlingen bij onderzoek doen of het doen van onderzoek door leerlingen (leerling). Het doel van de professionaliseringslijn is zowel onderzoek doen als kenniscon-

structie op het gebied van duurzaam beoordelen. Betrokkenheid van leerlingen, scholen en opleidingen zie ik als een belangrijke randvoorwaarde om de professionaliseringslijn te doen slagen.

De komende tijd hoop ik met meerdere opleidingen binnen de Faculteit de onderzoeksvragen te verfijnen om zo tot een aantal structurele lectoraatactiviteiten te komen. Ik ben mij ervan bewust dat één lectoraatsperiode te kort zal zijn om op alle vragen een antwoord te vinden. Maar als we al komen tot een visie op beoordelen die past bij duurzaamheid in vraaggestuurd leren, ben ik een meer dan tevreden mens.

SLOTWOORD

U hebt gemerkt dat het er mij niet zozeer om gaat een aantal vastomlijnde methoden en technieken voor beoordelen te presenteren. Mijn voornaamste doel was u een aantal vragen en gedachten voor te leggen, die een eerste invulling kunnen geven aan *duurzaam beoordelen in vraaggestuurd leren*. De geschetste problemen binnen elk speerpunt illustreren dat we te lang blijven hangen in een visie op beoordelen die een eindigheid in leren suggereert. Door de ingebrachte voorstellen hoop ik de komende jaren een verandering in deze visie te stimuleren.

Ik besef dat de praktijk van alledag het zeer lastig maakt om op individueel niveau leerlingen informatie te verschaffen over hun kennen en kunnen op een manier die ook nog eens past bij hun leerbehoefte en persoonlijkheid. Meestal ontbreekt de tijd als de groepen te groot en te heterogeen zijn. Ook zijn organisatorische aspecten als het leerstofjaarklassensysteem, meerdere leraren en leraressen per jaar, niet stimulerend om voldoende ruimte voor communicatie over individuele beoordelingen te realiseren. Ook gezien de dreigende lerarentekorten, gebrek aan tijd, geld, kwaliteit en de prikkelende leefomgeving van leerlingen is het niet eenvoudig duurzaam beoordelen te realiseren. Toch zal het onderwijs – gezien de maatschappelijke ontwikkelingen, zoals die nog recentelijk in diverse oraties aan de orde zijn gesteld (De Vries, 2007; Dieleman 2007; Martens, 2007) – ook in de randvoorwaardelijke sfeer mee moeten veranderen. Ik heb mij vooral gericht op het ontwerpen van beoordelen, het communiceren over beoordelen en het zelfsturen in beoordelen, maar zaken als de groepsgrootte, de fysieke inrichting van de school en de roostering van onderwijsactiviteiten zijn minstens zo belangrijk om duurzaam beoordelen in vraaggestuurd leren te kunnen waarmaken. Het lectoraat heeft daarom alleen kans van slagen in samenwerking met alle betrokkenen uit het onderwijsveld, andere wetenschappelijke disciplines en collega-lectoraten.

Pas als leren en beoordelen sterk op elkaar zijn afgestemd, een rijke communicatie tussen scholen, opleidingen en ouders over de functie van beoordelen plaatsvindt en de leerling wordt gestimuleerd verantwoordelijkheid voor het eigen leren te nemen door zich te ontwikkelen als een zelfsturend individu, mag worden gesproken van een kwalitatief hoogstaand beoordelingsmodel. Ik wil aan het slot van deze rede ervoor pleiten dat deze speerpunten sterker dan nu het geval is worden benadrukt in de criteria die inspecties en accreditatiecommissies hanteren. Ook zij kijken door een bril die beslagen is door productgerichte beelden over beoordelen. In dit opzicht laat het onderzoek van de Parlementaire Commissie Onderwijsvernieuwingen (2008) maar een zeer magere visie zien op beoordelen. Kwaliteit van beoordelen gaat veel verder dan hun pleidooi over verplichte invoering van gestandaardiseerde toetsen en een verplichte voldoende voor schoolexamens en centraal examen. Ik hoop dan ook dat de aanbevelingen op dit punt worden verworpen. Tot slot pleit ik voor nieuwe en rijkere instrumenten die nodig zijn om de kwaliteit van

een beoordelingskader vast te stellen, maar bovenal ook dat alle betrokkenen deze instrumenten goed kunnen hanteren. Heldere communicatie met leerlingen en leraren is hierbij essentieel. Pas dan kan een beoordelingsmodel hét visitekaartje van een onderwijsorganisatie worden.

DANKWOORD

Graag wil ik enkele mensen bedanken die voor mij belangrijk zijn geweest in de voorbereiding van dit lectorschap en dat ook zullen zijn in de verdere vormgeving van het lectoraat.

Deze rede had ik niet kunnen schrijven zonder de ‘meesters’ die mij de afgelopen tien jaar hebben ingewijd in de wereld van onderwijsontwerp en beoordelen. Filip Dochy, Lambert Schuwirth, Mien Segers, Gerard Straetmans, Jeroen van Merriënboer, Cees van der Vleuten en Wynand Wijnen, veel dank voor het delen van jullie onschatbare kennis en ervaring, maar ook voor jullie collegialiteit. Ik hoop nog veel van jullie te mogen leren.

Ik dank het College van Bestuur in de persoon van Ron Bormans en de directie van de Faculteit Educatie in de persoon van Jack Vogels dat ik de mogelijkheid heb gekregen weer in het Nijmeegse te vertoeven. Steef Woldinga, dank voor je kritische blik op mijn lectoraatsplannen. Ilja van Bree, Vera Jansen, Christian Peters en Jeanette Dusschooten, dank jullie wel voor het mede vormgeven van deze dag.

Arjan, Bregje, Frans, Gertjan, Jürg, Marijke en Ruud. Wat een luxe om met zeven lectoren in één faculteit een team te vormen. Ik hoop nog veel met jullie samen te werken in het programma ‘*Kwaliteit van leren*’ maar ook te genieten van vele gezamenlijke uitjes.

Dank aan mijn collega’s bij het Onderwijstechnologisch Expertisecentrum van de Open Universiteit Nederland en daarbuiten, die delen in mijn enthousiasme deze klus aan te gaan. In het bijzonder noem ik Jan van Bruggen, Jos van den Broek, Marieke Dresen, José Janssen, Desirée Joosten, Marcel van der Klink, Jo Knarren, Fleurie Nievelstein, Jan-Willem Strijbos, Iwan Wopereis en Marjo van Zundert.

Anouk, Brenda, José, Marieke, Mieke, Milleke en Susanne, dank voor jullie vriendschap.

Lieve ouders. Dat ik zitten bleef, was voor jullie destijds toch ook best even schrikken. Maar diep van binnen wisten jullie wel dat het weer goed zou komen. Dank voor jullie oprechte steun en betrokkenheid. Ik ben dankbaar dat jullie er zijn.

Roger, precies vijf jaar geleden, op hetzelfde tijdstip, gaven wij elkaar het ja-woord. De afgelopen vijf jaar zijn voorbijgevlogen. Duurzaamheid is zeker op ons van toepassing! Dank voor je geduld, humor en liefde.

Renée en Pascalle, jullie zijn twee heerlijke meiden. Ik geniet dagelijks van jullie spontaniteit en nieuwsgierigheid. Ik hoop dat deze eigenschappen in jullie tijd op school niet verloren zullen gaan.

Tijd om af te sluiten, of beter gezegd, tijd om te beginnen! Ik kijk er naar uit met leerlingen, leraren, lectoren en andere collega's vanuit verschillende perspectieven en disciplines de inhoud van deze rede te vertalen naar concrete activiteiten.

Ik dank u allen hartelijk voor uw blijk van belangstelling.

Dominique Sluijsmans
Nijmegen/Kerkrade, mei 2008

REFERENTIES

- Adams, A. (1999). The dead grandmother/exam syndrome. *Annals of Improbable Research, November/December*, 3-6.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology, 2*, 224-237.
- Baartman, L. K. J. (2008). 'Assessing the assessment': *Development and use of quality criteria for competence assessment programmes*. Unpublished doctoral dissertation, Utrecht University, Utrecht, the Netherlands.
- 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.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.
- 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., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bolocofsky, D. N., & Mescher, S. (1984). Student characteristics: Using student characteristics to develop effective grading practices. *The Directive Teacher, 6*, 11-23.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.
- Brett, J. F., & Atwater, L. (2001). 360 Feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*, 930-942.
- 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.

- Broekkamp, H., Van Hout-Wolters, B. H. A. M., Van den Bergh, H., & Rijlaarsdam, G. (2004). Students' expectations about the processing demands of teacher-made tests. *Studies in Educational Evaluation*, 30, 281-304.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11, 301-318.
- Brown, G. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99, 166-170.
- Brouwer, N. (2007). *De horizon van het leraarschap* (Lectorale rede). Doetinchem: Iselinge Hogeschool.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281.
- Chiviawsky, S., & Wulf, G. (2002). Self-controlled feedback: Does it enhance learning because performers get feedback when they need it? *Research Quarterly for Exercise and Sport*, 73, 408-415.
- Chiviawsky, S., & Wulf, G. (2005). Self-controlled feedback is effective if it is based on the learner's performance. *Research Quarterly for Exercise and Sport*, 76, 42-48.
- Corbalan-Perez, G. (2008). *Shared control over task selection. Helping students to select their own learning tasks*. Unpublished doctoral dissertation, Open Universiteit Nederland, Heerlen, the Netherlands.
- Corbalan-Perez, G., Kester, L., & Van Merriënboer, J. J. G. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science*, 34, 399-422.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Crow, G. M., & Pounder, D. G. (2000). Interdisciplinary teacher teams: Context, design, and process. *Educational Administration Quarterly*, 36, 216-254.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, US: Harper & Row.
- Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, 34, 391-401.
- Delandshere, G., & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27(2), 14-25.
- 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.
- Dieleman, A. (2007). *Pedagogische kwaliteit van de leraar*. Nijmegen: Hogeschool Arnhem en Nijmegen.

- Dochy, F. J. R. C., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23, 279-298.
- Domein Educatie HAN (2005a). *Summatieve competentiebeoordelingen via portfolio-assessment*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- Domein Educatie HAN (2005b). *Summatieve competentiebeoordeling groepsleerkracht Nijmegen*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- Domein Educatie HAN (2005b). *Visie op competentiegericht leren, opleiden en begeleiden*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- Doornbos, K. (1969). *Opstaan tegen het zittenblijven*. Den Haag: Staatsuitgeverij 's Gravenhage.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.
- Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, 24, 1-24.
- Eva, K., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation of the research agenda. *Academic Medicine*, 10, 46-54.
- Fishman, B. J., Marxa, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 19, 643-658.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 3, 193-202.
- Geerdink, G. (2007). *Diversiteit op de pabo. Sekseverschillen in motivatie, curriculum-perceptie en studieresultaten*. Unpublished doctoral dissertation, Nijmegen, Radboud Universiteit Nijmegen.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Gulikers, J. T. M., Bastiaens, Th. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77, 97-128.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Janssen-Noordman, A. M. B., & Van Merriënboer, J. J. G. (2002). *Innovatief onderwijs ontwerpen: Via leertaken naar complexe vaardigheden*. Groningen: Wolters-Noordhoff.
- Johnston, B. (2004) Summative assessment of portfolios: An examination of different approaches to agreement over outcomes. *Studies in Higher Education*, 29, 395-412.
- Joosten-ten Brinke, D., Sluijsmans, D. M. A., Brand-Gruwel, S., & Jochems, W. (2008). The quality of procedures to assess and credit prior learning: Implications for design. *Educational Research Review*, 3, 51-65.

- Klimoski, R., & Mohammed S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20, 403-437.
- Korthagen, F. A. J. (1998). *Leraren leren leren: Realistisch opleidingsonderwijs, geïnspireerd door Ph. A. Kohnstamm* (Inaugurele rede). Amsterdam: Viosiuspers AUP.
- Korthagen, F. A. J. (2004). Zin en onzin van competentiegericht opleiden. *VELON Tijdschrift voor Lerarenleraren*, 25, 13-23.
- Kremer-Hayon, L., & Tillema, H. H. (1999). Self-regulated learning in the context of teacher education. *Teaching and Teacher Education*, 15, 507-522.
- 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.
- Meherns, W. A., Popham, W. J., & Ryan, J. M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice*, 17, 18-22.
- Meuffels, B. (1989). De verguisde beoordelaar: Over de neerlandicus als opstelbeoordelaar. *Tijdschrift voor Taalbeheersing*, 11, 161-176.
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed), *Handbook of Research on Educational Communications and Technology* (2nd ed., pp. 745-783). New York: Lawrence Erlbaum Associates.
- Myers, D. G. (1980). *The inflated self*. New York: Seabury Press.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.
- 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.
- Nijhuis, J., Segers, M., & Gijselaers, W. (2008). The extent of variability in learning strategies and students' perceptions of the learning environment. *Learning and Instruction*, 18, 121-134.
- Onderwijsraad (2007). *Versteving van kennis in het onderwijs II*. Den Haag: Onderwijsraad.
- Oren, D. L. (1983). Evaluation systems and attributional tendencies in the classroom: A sociological approach. *Journal of Educational Research*, 76, 307-312.
- Parlementaire Commissie Onderwijsvernieuwingen (2008). *Tijd voor onderwijs*. Den Haag: Tweede kamer der Staten Generaal.

- Popham, W. J. (2001). Teaching to the test. *Educational Leadership*, 58(6), 16-20.
- Project HAN Onderwijsflexibilisering (HOF) (2006). *HAN-Chassis voor het onderwijs. Onderwijskundig en organisatorisch kader voor de bachelor opleidingen*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- Regehr, G., & Eva, K. (2006). Self-assessment, self-direction, and the self-regulating professional. *Clinical Orthopaedics and Related Research*, 449, 34-38.
- Rheinberg, F. (1983). Achievement evaluation: A fundamental difference and its motivational consequences. *Studies in Educational Evaluation*, 9, 185-194.
- Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19, 551-554.
- 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.
- Schuiling, G. (2007). *Werken in een portfolio van rollen*. Nijmegen: Hogeschool Arnhem en Nijmegen.
- 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.
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2005). A plea for new psychometric models in educational assessment. *Medical Education*, 40, 296-300.
- Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2004). Changing education, changing assessment, changing research? *Medical Education*, 38, 805-812.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472.
- Scouller, K. M., & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education*, 19, 267-279.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Slavin, R. E. (1980). Effects of individual learning expectations on student achievement. *Journal of Educational Psychology*, 72, 520-524.
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment: The training of peer assessment skills*. Unpublished doctoral dissertation, Open Universiteit Nederland, Heerlen, the Netherlands.
- Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International*, 41, 59-78.

- 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.
- Sluijsmans, D. M. A., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self- peer- and co-assessment. *Learning Environments Research*, 1, 293-319.
- Sluijsmans, D. M. A., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, 32, 6-22.
- Sluijsmans, D. M. A., Straetmans, G., & Van Merriënboer, J. (2008). Integrating authentic assessment with competency based learning: The protocol portfolio scoring. *Journal of Vocational Education and Training*, 159-172.
- Sluijsmans, D. M. A., Strijbos, J. W., & Van de Watering, G. (2007, August). Modelling the impact of individual contributions on peer assessment during group work: In search of flexibility. In J. W. Strijbos & D. M. A. Sluijsmans (chairs), *Optimising peer assessment for learning*. Paper presented in a symposium conducted at the 12th biennial EARLI conference, Budapest, Hungary.
- Smith, M. L., & Fey, P. (2000). Validity and accountability of high-stakes testing. *Journal of Teacher Education*, 51, 334-344.
- Starch, D., & Elliot, E. C. (1912). Reliability of the grading of high school work in English. *School Review*, 20, 254-259.
- Starch, D., & Elliot, E. C. (1913). Reliability of grading high school work in mathematics. *School Review*, 21, 442-457.
- Stichting Leerplan Ontwikkeling (2006). *Competentiegericht vmbo hoe-zo?! De vormgeving en organisatie van competentiegericht onderwijs in het vmbo*. Enschede: SLO.
- Straetmans, G. J. J. M. (2006). *Bekwaam beoordelen en beslissen* (Lectorale rede). Deventer: Saxion Hogescholen.
- Straetmans, G., Sluijsmans, D. M. A., Bolhuis, B., & Van Merriënboer, J. J. G. (2003). Integratie van instructie en assessment in competentiegericht onderwijs. *Tijdschrift voor Hoger Onderwijs*, 3, 171-197.
- Strijbos, J. W., Narciss, S., & Dünnebier, K. (2007, August). Peer feedback in academic writing: How do feedback content and writing ability-level of the sender affect feedback perception and performance? In J. W. Strijbos & D. M. A. Sluijsmans (chairs), *Optimising peer assessment for learning*. Paper presented in a symposium conducted at the 12th biennial EARLI conference, Budapest, Hungary.
- Tang, C., Lai, P., Arthur, D., & Leung S. F. (1999). How do students prepare for traditional and portfolio assessment in a problem-based learning curriculum. In J. Conway & A. Williams (Eds.), *Themes and Variation in PBL* (pp. 3-10). Newcastle: Australian Problem Based Learning Network.

- 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. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Merriënboer, J., Clark, R. E., & De Croock, M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology, Research and Development*, 50, 39-64.
- Van Merriënboer, J. J. G., Jelsma, O., & Paas, F. (1992). Training for reflective expertise: A four-component instructional design model for training complex cognitive skills. *Educational Technology, Research and Development*, 40, 23-43.
- 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.
- Van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2008, February). *Toward a synthesis between cognitive load theory and self directed learning*. Paper presented at the 2nd Cognitive Load Theory Conference, Wollongong, Australia.
- Van Merriënboer, J. G., Sluijsmans, D. M. A., Corbalan, G., Kalyuga, S., Paas, F., & Tattersall, C. (2006). Performance assessment and learning task selection in complex learning environments. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 201-220). Amsterdam, The Netherlands: Elsevier.
- Van Tartwijk, J., Driessen, E., van der Vleuten, C., & Stokking, K. M. (2007). Factors influencing the succesful introduction of portfolios. *Quality in Higher Education*, 13, 69-79.
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Science Education*, 7, 63-80.
- Wijnen, W. (1971). *Onder of boven de maat: Een methode voor het bepalen van de grens voldoende/onvoldoende bij studietoetsen*. Unpublished doctoral dissertation, Universiteit Groningen, Groningen, the Netherlands.
- 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.

