

Blog naar aanleiding van het Dylan Wiliam Event op 3, 4 en 5 september 2018 in Amersfoort

Dominique Sluijsmans | lector Professioneel Beoordelen Zuyd Hogeschool

Wordt in drie delen gepubliceerd op www.toetsrevolutie.nl

Januari 2019

Neem formatief niet te snel voor lief: tien lessen van Dylan Wiliam

In 1998 startte ik bij de Open Universiteit een promotietraject dat zich richtte op het betrekken van studenten bij toetsing door hen te trainen in peer assessment vaardigheden. Eén van de eerste publicaties die ik bestudeerde was [Assessment and Classroom Learning](#) van Paul Black en Dylan Wiliam, een onderhavig literatuuronderzoek naar formatief toetsen in de klassenpraktijk. Sindsdien heeft dit thema en het werk van Wiliam mijn aandacht. In 2008, tien jaar na het verschijnen van deze studie, ontmoette ik deze markante man voor de eerste keer. Het was tijdens het vierde congres van de SIG Assessment van de EARLI, waar hij een [keynote](#) verzorgde. We hadden een onderhoudend gesprek in de bus van de congreslocatie naar het hotel. In september jl., tien jaar later na de eerste ontmoeting, deed de kans zich voor hem weer te zien. In een [driedaags event](#) heeft hij ons deelgenoot gemaakt van al zijn kennis en ervaring over formatief toetsen en curriculumontwerp. Over dit bijzondere event is al eerder geblogd door [Jörgen van Remoortere](#) en [Martje Köhlen](#). De presentaties van Wiliam zijn inmiddels ook te vinden op de [Toetsrevolutie-website](#). In deze bijdrage deel ik de voor mij tien belangrijkste lessen uit Wiliam's seminar. Ik hoop dat deze lessen enige helderheid over formatief toetsen geven, maar vooral ook handvatten bieden voor een goed gesprek.

1. Formatief is een eigenschap van een beslissing, niet van een toets

Formatief toetsen wordt vaak in één adem genoemd met summatief toetsen. Waar formatief toetsen is gericht op beslissingen over verdere stappen in het leren, gaat summatief toetsen over beslissingen als certificering of selectie. In de praktijk blijkt echter dat [formatieve toetsing nog steeds veel eigenschappen heeft van summatieve toetsing](#). Denk bijvoorbeeld aan een summatieve toets die niet meetelt en daarom formatief zou zijn, het werken met dezelfde typen vragen als in een summatieve toets, of formatieve toetsing beschouwen als een toetsinstrument. Toetsing in algemene zin ziet Wiliam echter niet als een set van instrumenten, maar als een procedure die je nodig hebt om zorgvuldige conclusies te trekken over het leren en de voortgang van studenten. Formatief en summatief zijn in deze definitie geen eigenschappen van toetsen, maar eigenschappen van die conclusies. Toetsing gericht op het nemen van formatieve beslissingen ziet Wiliam veeleer als een *proces van onderwijzen*, waarbij de docent een aantal technieken/activiteiten inzet binnen het territorium van formatief toetsen. Binnen dit territorium is een vijftal formatieve strategieën te benoemen die van belang zijn in de fasen van doelen stellen, werken aan taken en het bepalen van de weg om doelen te bereiken (zie afbeelding 1, afkomstig uit een presentatie van Wiliam in Adelaide, Australië): 1) het verhelderen, delen en begrijpen van leerdoelen, 2) het organiseren van discussies, taken en activiteiten om bewijs voor leren te genereren, 3) feedback om de student verder te helpen, 4) studenten activeren als leerbronnen voor elkaar en 5) studenten activeren als verantwoordelijken voor hun eigen leren.

	Where the learner is going	Where the learner is	How to get there
Teacher	Clarifying, sharing and understanding learning intentions	Engineering effective discussions, tasks, and activities that elicit evidence of learning	Providing feedback that moves learners forward
Peer		Activating students as learning resources for one another	
Learner		Activating students as owners of their own learning	

Afbeelding 1. Het territorium van formatief toetsen

2. Formatief toetsen staat niet gelijk aan responsive teaching

Dat Wiliam niet altijd gelukkig is met de praktische invulling van zijn werk, blijkt uit een tweet van 23 oktober 2013: *'Example of really big mistake: calling formative assessment formative assessment rather than something like "responsive teaching"*. In een Twitterbericht van 24 maart 2018 nuanceert hij echter deze definitie, door *responsive teaching* te positioneren binnen het bovenstaande territorium van formatief toetsen en de relatie tussen onderwijzen en toetsen te duiden (zie afbeelding).

	Where the learner is going	Where the learner is	How to get there
Teacher	Before you can even begin Clarifying, sharing and understanding learning intentions	Engineering effective discussions, tasks, and activities that elicit evidence of learning	Providing feedback that moves learners forward
Peer		Activating students as learning resources for one another	
Learner		Activating students as owners of their own learning	

Responsive teaching

The learners' role

Afbeelding 2. De plek van responsive teaching in het territorium van formatief toetsen

In deze afbeelding is te zien dat formatief toetsen duidelijk onderscheid maakt tussen de rol van de docent in de fasen vaststellen waar de studenten staan en hoe zij hun doelen kunnen bereiken, en de rol van de studenten zelf. Wiliam stelt dat *responsive teaching* en de rol van studenten beiden belangrijke aspecten zijn van formatief toetsen, maar dat hij formatief toetsen niet gelijkstelt aan *responsive teaching*. Dat zou immers betekenen dat het alleen gaat om de rol van de docent. Formatief toetsen verwijst naar het vormgeven van een (cyclisch) leerproces waarbinnen docenten door *responsive teaching* de brug slaan tussen onderwijzen en leren, en daarbij in staat zijn studenten

actief te betrekken en aan het werk te zetten. Om in het vervolg van deze bijdrage de focus te houden op de formatieve functie van toetsing (het komen tot conclusies met een formatief doel), zal ik in de volgende lessen - in lijn met de eerder verschenen [blog](#) van René Kneyber - spreken over formatief handelen, waarin dus zowel de docent als de student een belangrijke rol vervullen.

3. Kennis over wat leren betekent is voorwaardelijk voor goed formatief handelen

Als we het hebben over onderwijs, is het woord *leren* misschien wel het meest gebruikte woord. Er is binnen docententeams lang niet altijd eenduidigheid over wat onder leren wordt verstaan. Wiliam stelt heel duidelijk: leren is verandering in het langetermijngeheugen, waarbij *storage strength* en *retrieval strength* een belangrijke rol spelen (lees hierover meer in deze [blog](#)). Een ander belangrijke boodschap vanuit de onderwijspsychologie is dat we dienen aan te sluiten bij wat de student al weet (het werk van [David Ausubel](#)): voorkennis is de start van leren. Het is van belang deze voorkennis goed in kaart te brengen en vanuit die informatie het onderwijs in te richten. Bij het introduceren van een nieuw onderwerp aan studenten, bestaat het gevaar dat zij na een les er verschillende beelden over hebben. Het is van belang deze beelden expliciet te krijgen, zodat we snappen wat er in de hoofden van studenten gebeurt. Het bijzondere is dat we juist in de gangbare wijzen van toetsing gericht zijn op kortetermijnleren en niet uitgaan van de voorkennis van studenten. Studenten hollen van toets naar toets, vaak door net voor de de toets de opgegeven stof in het hoofd te proppen (*cramming*). Daisy Christodoulou stelt dat toetsen op deze wijze lijkt op het rennen van teveel marathons, en houdt in haar recente boek *Making good progress* een goed en onderbouwd pleidooi het aantal toetsen op basis waarvan summatieve conclusies worden getrokken sterk te beperken. Zij stelt dat formatief handelen studenten de kans biedt veelvuldig te oefenen (*deliberate practice*); fouten maken mag en moet zelfs om langetermijnleren te versterken. Huidige wijzen van toetsen lijken dit juist te frustreren. Een visie op toetsen dient daarom ALTIJD te volgen op een visie op leren, nooit andersom. Het heeft weinig nut een toetsprogramma te ontwerpen zonder een gedeeld beeld over de doelen van het onderwijs en een gedeelde opvatting over leren.

4. Het type beslissing die je als docent wilt nemen bepaalt dataverzameling, niet andersom

Formatief handelen richt zich op het verzamelen van informatie die nodig is om een bepaalde beslissing over de vervolgstappen in het leren van studenten te nemen. Er is sprake van doelgericht in plaats van toetsgericht handelen. De verzamelde informatie dient docenten handvatten te geven waar studenten staan en welke vervolgstappen nodig zijn (tijdens een les of over lessen heen). De verzamelde informatie wordt pas bewijs als deze worden gebruikt een bepaalde claim te onderbouwen (bijvoorbeeld: Hebben studenten de juiste voorkennis om taak X of Y te kunnen uitvoeren?). Docenten hebben informatie van hoge kwaliteit nodig om deze beslissingen goed te kunnen nemen. Dit betekent dat zij in staat moeten zijn studenten de juiste stimuli te geven zodat de gewenste response wordt opgeroepen, waarbij de kwaliteit van de stimulus de kwaliteit van de response zal bepalen. De kwaliteit van goede vragen/opdrachten is dus bepalend voor de wijze waarop studenten zullen antwoorden. Vragen zijn bedoeld om studenten aan het denken te zetten en informatie te genereren. Zonder de juiste informatie bestaat het risico dat er onjuiste conclusies worden getrokken ('Studenten die hun hand opstaken gaven het goede antwoord dus ik kan verder met mijn les.'). Wiliam stelt dat in het onderwijs vaak de focus ligt op datagestueerd denken, waarbij de verzamelde data leidend zijn voor beslissingen. Deze wijze van informatieverzameling is vaak normgericht, verzameld op vastgestelde momenten en ook vaak maar gericht op een deel van het onderwezen curriculum. Voorbeelden van datagestueerd werken zijn de formeel georganiseerde proefwerken, gestandaardiseerde toetsen en deficiëntietoetsen. De resultaten van deze toetsen worden vervolgens gebruikt voor het nemen van beslissingen. Om formatief handelen kracht bij te

zetten, pleit Wiliam voor een verschuiving van een focus op *data-driven decision-making* naar *decision-driven data collection*.

5. Formatief handelen betekent ALLE studenten aan boord houden

Wiliam benoemt twee doelen van formatief handelen: ervoor zorgen dat alle studenten betrokken zijn en het versterken van zelfregulatievaardigheden. Betrokkenheid betekent dat alle studenten 'er zijn'. Dit vraagt van docenten dat zij weten wanneer welke formatieve strategieën belangrijk zijn en dat zij doelbewuste keuzes maken welke formatieve technieken/activiteiten zij inzetten om alle studenten bij de les te houden. Voorbeelden van activiteiten gericht op vaststellen waar de studenten staan zijn bijvoorbeeld het inzetten van een 'alle-studenten-geven-antwoord-aanpak', waarbij elke student wordt gevraagd antwoorden te geven op een aantal vragen over eerder behandelde stof. Deze vragen zijn zo geformuleerd dat docenten zicht krijgen op de misconcepties of veel gemaakte fouten. Studenten kunnen vervolgens met elkaar antwoorden bespreken en elkaar uitleggen waarom antwoorden al dan niet goed zijn. Een ander voorbeeld gaat over het leren feedback geven op elkaars werk. Ook hier gaat het er niet om elkaar te beoordelen, maar om als docent informatie te verzamelen waar studenten nog kennis missen of misinterpretaties hebben van de criteria voor succes. Zoals al eerder genoemd, zijn al deze activiteiten gericht op informatie verzamelen die je als docent helpen te bepalen welke vervolgstappen te zetten. Belangrijk is dat studenten in veiligheid en vertrouwen leren worstelen met een taak en niet teveel worden gestimuleerd te doen wat ze fijn vinden te doen of al kunnen. Of zoals Wiliam het stelt: *'We love to be right, but we learn more when we are wrong.'* Formatief handelen biedt kansen de motivatie van studenten te versterken door ze door veel oefening success te laten ervaren. Motivatie volgt na succes, niet andersom. Een interessant fenomeen gerelateerd aan fouten maken - dat voor mij nieuw was - betreft het hypercorrectie-effect. Studenten die het onjuiste antwoord geven, maar overtuigd zijn dat het antwoord wel juist is, onthouden na feedback sneller het juiste antwoord (ze 'schrikken' dat het antwoord toch niet goed was) dan studenten die het onjuiste antwoord geven en hier ook niet overtuigd van zijn. 'High-confidence' fouten worden hypergecorrigeerd. Overigens stelt Wiliam dat de antwoorden van zelfverzekerde studenten een slechte indicatie zijn van wat de rest van de groep denkt. Ook hier geldt weer: probeer van alle studenten informatie te verkrijgen op belangrijke momenten in het leerproces. Als het gaat om versterken van de zelfregulatievaardigheden, vraagt dit besef dat de focus van formatief handelen ligt op het verbeteren van de *student*, niet het werk. Daarmee bedoelt Wiliam dat het niet gaat om het alleen maar kunnen uitvoeren van bijvoorbeeld een project of opdracht, maar om langetermijnveranderingen in de kennis en vaardigheden die nodig zijn om bepaalde taken te kunnen uitvoeren. Kennis en vaardigheden die specifieke taken ontstijgen. Hij geeft dan ook aan geen problemen te hebben met bijvoorbeeld projectmatig werken, zolang de focus maar ligt op leren. Zelfregulatievaardigheden zijn nodig te kunnen blijven leren ook na het behaalde diploma. Het werk van David Boud over [sustainable assessment](#) sluit hierbij aan: hij stelt dat studenten toerusten met zelfregulatievaardigheden één van de belangrijkste doelen is van het onderwijs.

6. Feedback is niet altijd zinvol

Formatief handelen wordt vaak gelijk gesteld aan feedback. Of anders gezegd: formatief handelen zonder feedback bestaat niet. Feedback benoemt Wiliam als één van de strategieën in het territorium van formatief handelen. Feedback geven is echter zeer arbeidsintensief en heeft lang niet altijd het gewenste effect (vaak tot frustratie van docenten). Wiliam stelt dat bij feedback acht dingen kunnen gebeuren: twee zijn goed, zes zijn slecht (zie afbeelding).

Response type	Feedback indicates performance...	
	falls short of goal	exceeds goal
Change behavior	Increase effort	Exert less effort
Change goal	Reduce aspiration	Increase aspiration
Abandon goal	Decide goal is too hard	Decide goal is too easy
Reject feedback	Feedback is ignored	Feedback is ignored

Afbeelding 3. Bij feedback kunnen acht dingen gebeure, twee zijn slechts goed

Als het gaat om het geven van feedback bij het nog niet behalen van het leerdoel, kan feedback bij de student leiden tot a) een een afname in de wil iets te bereiken, b) het besluit dat het doel niet te behalen is of c) het negeren van de feedback. Als het gaat om het geven van feedback bij het behalen van het leerdoel, kan feedback bij de student leiden tot a) minder moeite doen, b) het besluit dat het doel dat behaald dient te worden te makkelijk is of c) het negeren van feedback. Bij de student die nog het doel *niet* heeft behaald werkt feedback alleen als het leidt tot een toename in moeite die de student wil nemen het doel te bereiken (*increase effort*); bij de student die het doel *wel* behaalt dient feedback voor het versterken van de wil een nieuw doel te bereiken (*increase aspiration*). Kernboodschap is dus dat feedback de student in beweging brengt en houdt. Dit betekent ook dat de student in staat moet zijn de feedback te begrijpen om er vervolgacties aan te kunnen koppelen. Feedback heeft een duurzaam effect als een student niet alleen het betreffende werk kan verbeteren, maar ook tot nieuw beter werk komt. Hier ligt dan ook het uiteindelijke doel van feedback: het heeft studenten geholpen om meer onafhankelijk te worden. Effectieve (zelfregulerende) studenten zijn in staat hun eigen feedback te organiseren in relatie tot succescriteria en met deze feedback te werken.

David Didau schreeft enkele interessante blogs over feedback. Hij benoemt drie redenen voor het het geven van feedback:

- helderheid over doel geven (*understanding*);
- studenten stimuleren meer moeite te doen (*effort*);
- studenten stimuleren tot meer uitdaging (*aspiration*).

Ik verwijs naar [Didau's blogs](#) om zijn visie te aanzien van feedback te lezen. Ook de blogs van [Paul Kirschner](#), [René Kneyber](#) en [Tom Sherrington](#) zijn het lezen waard.

7. Wees kritisch over het gebruik van rubrics

In het onderwijs zien we dat er steeds vaker rubrics worden ingezet als beoordelingsinstrument voor bepaalde leeruitkomsten. Een rubric beschrijft enerzijds de criteria waaraan een bepaalde uitkomst dient te voldoen, anderzijds de niveaus van beheersing, bijvoorbeeld onvoldoende-voldoende-goed-excellent. Hierdoor ontstaat een cellenstructuur, waarbij elke cel een kwalitatieve beschrijving geeft van het betreffende criterium en niveau. Rubrics worden in de onderwijspraktijk gebruikt voor zowel formatieve als summatieve doeleinden. In het seminar kraakte Wiliam vier kritische noten over het gebruik van rubrics. Op de eerste plaats zijn er in een rubric nooit lege cellen, in elke cel staat wel een

aantal aspecten beschreven ten aanzien van het criterium en niveau. Dit zorgt ervoor dat bij het ontwerpen ervan wellicht gezocht moet worden naar invulling, in plaats van dat men nadenkt of vulling wel zo zinvol is. Tweede probleem is dat een rubric maar beperkte informatie geeft hoe te verbeteren. Er staat wel wat een bepaald criterium op een bepaald niveau betekent, maar dat zegt nog niks over HOE bijvoorbeeld naar een 'goed' toe te werken. Juist in de feedback over het HOE zit de kwaliteit en informatie om verder te komen, niet in de criteria zelf. Rubrics bevatten bovendien vaak vage criteria (wat is bijvoorbeeld het verschil tussen onheldere en heldere structuur?). Een derde probleem is dat studenten de inhoud van een rubric niet hetzelfde zullen interpreteren als docenten (de experts). Het goed kunnen begrijpen van criteria en de inhoud van een rubric wordt bepaald door de voor- en domeinkennis die studenten hebben. Als hier geen informatie over is, bestaat het gevaar dat studenten criteria misinterpreteren en niet tot de juiste vervolgstappen komen. Een vierde probleem van rubrics is dat wordt aangegeven wanneer het werk onvoldoende is. Wiliam stelt dat het beter werkt studenten te laten werken met *goede* voorbeelden, in plaats van slechte voorbeelden. Dit maakt het mogelijk goede voorbeelden te 'deconstrueren' en expertaanpakken te modelleren.

Een mogelijk beter alternatief voor uitgebreide rubrics is de [single-point rubric](#). Bij deze rubric wordt alleen het voldoende niveau per criterium beschreven. Dit biedt de kans om met studenten in gesprek te gaan waarom bepaald werk nog niet voldoet (en dit goed te beschrijven), of wanneer er sprake is van meer dan voldoende beheersing (en wat dat betekent voor de volgende stap). Ook geeft een single point rubric mogelijkheden voor zelf- en peer-beoordeling. Zover ik dat kan overzien, is er nog weinig empirisch onderzoek naar de werking van single point rubrics. Het is interessant een single-point rubric op te stellen met studenten op basis van het bekijken van uitgewerkte voorbeelden van gewenste prestaties.

8. Stimuleer als leider dat onze goede docenten zich kunnen verbeteren

Wiliam benadrukt dat docenten voldoende kans dienen te krijgen formatief handelen te oefenen in de dagelijkse onderwijspraktijk. De opdracht voor leiders is ervoor te zorgen dat docenten zich kunnen verbeteren in datgene wat voor studenten het beste is. Als het gaat om toetsbekwaamheid zijn de volgende vragen van belang: Begrijp je wat dit toetsresultaat betekent? Heeft het waarde als het gaat om het beoogde gebruik ervan? Welke boodschap straalt de toetsing uit naar relevante betrokkenen? Wat zal het effect van deze toets zijn op studenten? Investeren in toetsbekwaamheid is overigens een van de belangrijkste adviezen uit het recent verschenen advies [Toets wijzer](#) van de Onderwijsraad. Om formatief te leren denken en handelen stelt Wiliam het inrichten van een lerend docentgezelschap in dat werkt volgens een vast patroon (zie afbeelding).

Teacher learning communities

86

- Plan that the TLC will run for two years.
- Identify 10 to 12 interested colleagues:
 - Composition:
 - Similar assignments (e.g., early years, math/science)
 - Mixed subject/mixed phase
 - Hybrid
- Secure institutional support for:
 - Monthly meetings (75–120 minutes each, inside or outside school time)
 - Time between meetings (two hours per month in school time):
 - Collaborative planning
 - Peer observation
 - Any necessary waivers from school policies

Learning Sciences
DylanWiliamCenter

Afbeelding 4. Een aanpak voor een lerend docentgezelschap

Voor het slagen van een dergelijke aanpak is een goede moderator om de bijeenkomsten goed te leiden essentieel. Ook is het zinvol een onderzoeker aan het gezelschap te verbinden. Door flankerend onderzoek te doen naar de werking en effecten van deze aanpak, wordt kennis ontwikkeld over formatief werken op teamniveau. Ook is het zinvol opgedane ervaringen uit te wisselen met opleidingen die met een gelijke aanpak experimenteren. De huidige [leernetwerken formatief evalueren](#) zijn een mooi voorbeeld hoe scholen van elkaar kunnen leren.

9. Zeven principes voor curriculumontwerp

Naast veel kennis en ervaringen over formatief handelen, deelde Wiliam in de tweede dag ook een aantal richtlijnen voor curriculumontwerp. Formatief handelen kan immers niet los worden gezien van de doelen van het onderwijs en het leerplan dat wordt ingericht om die doelen te bereiken. Tot nu was mij alleen het beoogde (voorgeschreven), uitgevoerde (geïnterpreteerde) en bereikte (ervaren) curriculum bekend, waarbij er sprake is van een expliciet curriculum (de programma's, cursussen, vakinhouden) en een impliciet/verborgen curriculum (waarden en verwachtingen, deze zijn er, maar wordt niet expliciet over gesproken). Wat ik niet wist is dat er ook een 0-curriculum is. Het 0-curriculum verwijst naar wat opleidingen *niet* expliciet onderwijzen. Dit kan betekenen dat studenten bepaalde keuzes niet zullen worden aangeboden, met bepaalde perspectieven niet worden geconfronteerd en een aantal concepten en vaardigheden geen onderdeel zullen gaan uitmaken van hun intellectuele repertoire. Wiliam presenteerde vervolgens zeven principes voor curriculumontwerp, waarover [hier](#) meer achtergronden te lezen zijn: 1) gebalanceerd: er is sprake van een evenwichtige verdeling van inhouden, 2) stringent: er worden krachtige manieren van denken ontwikkeld door voortdurend met een bepaald onderwerp of discipline aan de slag te zijn, 3) samenhangend: studenten leren te zien hoe verschillende inhouden/disciplines zich tot elkaar verhouden, 4) verticaal geïntegreerd: de nadruk ligt op vooruitgang in leren waarbij inhouden herhaaldelijk aan bod komen (met verwijzing naar het [spiraalcurriculum van Bruner](#)), 5) geschikt: past het curriculum bij het niveau van de student en welke implicaties heeft dit voor het organiseren

van het onderwijs (op leerfase of jaargroep), 6) gefocused: geef vooral aandacht aan de *big ideas* binnen een discipline en durf te kiezen, en 7) belangrijk: hoe kun je studenten geïnteresseerd krijgen voor zaken waarvan ze niet wisten dat ze er interesse voor zouden hebben? Voor alle principes stelt Wiliam dat er geen eenduidige antwoorden zijn hoe het curriculum te ontwerpen en dat de principes onderling op spanning kunnen staan. Ook hier wordt weer duidelijk hoe ingewikkeld het ontwerpen van een curriculum is.

10. Toetsing dient het curriculum te dienen, niet te sturen

Een interessant thema dat een rode draad is door alle eerder beschreven lessen, is de aansluiting tussen de doelen van het onderwijs, het curriculum (het leerplan) en de toetsing. De laatste jaren is er veel discussie hoe die aansluiting er zou moeten uitzien, bijvoorbeeld in het [nadenken over een andere aanpak van de eindexamens](#). Wiliam is duidelijk dat de wijze van toetsing het curriculum moet dienen, niet beheersen. Zo ziet de praktijk er echter niet uit, voor studenten zijn de toetsen leidend en bepalen deze waar ze energie in steken, niet het curriculum ('Moeten we dit weten voor de toets?'). Wiliam is scherp in het duiden van de tekortkomingen in de wijze van toetsen gericht op een summatieve conclusie. Ten eerste waarschuwt hij voor Campbell's Law: als toetsresultaten het doel worden van onderwijzen, verliezen deze resultaten hun waarde en zullen zij ongewenste effecten hebben op het leerproces. Toetsresultaten worden uiteindelijk betekenisloos. Een tweede risico is het gebrek aan validiteit door *construct underrepresentation* (niet alle te beoordelen aspecten komen aan de orde in de toets, bijvoorbeeld het spreken wordt bij Engels niet beoordeeld) en *construct irrelevant variance* (er worden aspecten beoordeeld die niet er toe zouden mogen doen, bijvoorbeeld begrijpend lezen in een rekentoets). Dit is vooral aan de orde in de eindexamens, die lang niet altijd een goede afspiegeling zijn van wat in het curriculum aan bod komt. Een mooi voorbeeld hiervan beschrijft Martin Ringenaldus in zijn [blog](#) over het examen Duits. Een derde probleem betreft de accuraatheid van een score/cijfer. Wiliam illustreert op basis van onderzoek naar betrouwbaarheid dat bijvoorbeeld een 7.1 op moment 1 net zo goed een 6.1 of 8.1 had kunnen zijn (hoe preciezer de score, des te lager de accuraatheid). Wiliam citeerde hierbij de definitie van een cijfer van Dressel (geciteerd in Chickering, 1983): *"A grade is an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student had attained an undefined level of mastery of an unknown proportion of an indefinite material."* Het feit dat cijfers niet betrouwbaar zijn is een belangrijke constatering, aangezien we bij bijvoorbeeld het eindexamen wel zeer zwaarwegende beslissingen nemen op basis van dit cijfer. Om te voorkomen dat toetsing zijn doel voorbijschiet, adviseert Wiliam een toetsprogramma te ontwerpen dat voldoet aan de volgende kenmerken: 1) bewijs van leren wordt niet pas aan het einde verzameld, maar op een groot aantal momenten, 2) er is voldoende ruimte kennis en vaardigheden op te bouwen, 3) er is voldoende kans alle aspecten zowel in de diepte als breedte te leren, 4) het is organiseerbaar zodat kosten in verhouding zijn met opbrengsten, en 5) alle betrokkenen hebben vertrouwen in de uitkomsten. Mijn persoonlijke vertaling is dat in het onderwijs vooral sprake is van activiteiten waaraan formatieve conclusies worden verbonden (formatief handelen gericht op langetermijnleren), en dat toetsing met summatieve doeleinden uiteindelijk een ceremonieel karakter heeft (met dank aan Ben Wilbrink voor het woord ceremonieel). Dit zou dan een andere invulling geven van 'stakes' (Wat staat er op het spel?): high-stake summatieve toetsen worden low-stake (het eindexamen leidt niet meer tot verrassingen), low-stake formatieve toetsen worden high-stake. Het *leren* gericht op een langetermijneffect staat immers centraal.

Korte beschouwing

Het seminar van Dylan Wiliam heeft mij geholpen in het bijstellen van een aantal beelden over formatief toetsen. Het wordt steeds duidelijker wat formatief handelen wel en ook niet is. Ook zet het

tot nadenken over de vraag hoe formatieve en summatieve doeleinden beter kunnen worden gebalanceerd. Hier zit mijns inziens nog wel een grote uitdaging: In hoeverre is het mogelijk om als school of opleiding beide functies van toetsing waar te maken zonder aan kwaliteit van één van beiden in te boeten? Zou het onderwijs er niet alleen moeten zijn om te (leren) leren, in plaats van een plek van voortdurend presteren? Wat verder opvalt is dat formatief handelen een steeds sterkere theoretische inbedding krijgt. Wiliam lijkt de laatste jaren vooral het cognitief psychologisch perspectief te hebben ingenomen, blijkend uit vele verwijzingen naar bijvoorbeeld de cognitieve belastingstheorie en het werk van Ausubel, Kirschner en van Merriënboer. Het is een goede ontwikkeling dat formatief handelen een sterkere binding krijgt met andere perspectieven op toetsen, zoals het [programmatisch toetsen](#), [effectieve leerstrategieën](#) en het werken aan kansengelijkheid (zie bijvoorbeeld werk van Ben Wilbrink over [fair schooling](#)). Formatief handelen gaat over leren en minder over toetsen. De vraag is dan ook of we überhaupt nog over formatief toetsen moeten willen spreken. Ik hoop echter dat we in de verdere kennisontwikkeling aandacht blijven houden voor de relatie tussen formatief handelen en de andere doeldomeinen van het onderwijs van Biesta, zoals socialisatie en persoonsvorming. Juist de aandacht voor de student, de interactie tussen studenten en de interactie tussen docenten en studenten is iets wat mij altijd in formatief handelen zo heeft aangesproken. Ik hoop dat door de sterke cognitieve focus op formatief handelen, de pedagogische en meer persoonsvormende functie niet op de achtergrond raakt. Onderwijs is er om iets te leren - laat daar geen twijfel over zijn - maar het is ook de plek waar je in verbinding met anderen je identiteit met bijbehorende waarden en normen vormgeeft. Dat is iets dat we niet meetbaar moeten willen maken met toetsen met summatieve doeleinden, maar kan wel worden verrijkt met interacties, ontmoetingen, frusterende ervaringen waarbij je als student weet dat daarin falen en opnieuw proberen mag. Voor mij is ook dat nog steeds formatief handelen, al is het dan niet met een cognitieve insteek, maar gericht op het leren jezelf te blijven ontwikkelen als mens.

Tot slot

Formatief handelen gaat over goed onderwijs. Mijn hoop is dat we formatief niet te snel voor lief nemen, maar steeds bewust blijven waar het bij formatief handelen om gaat en waarom dat zo is vanuit de wetenschap over leren. Een seminar van Dylan Wiliam helpt om scherp te blijven, maar ook kritisch te zijn en vragen te blijven stellen. Hoewel onderzoek ons helpt goed geïnformeerd te handelen, zal de specifieke context bepalen wat kan en lukt. Wiliam zegt hierover: *Alles werkt wel een keer ergens*. Het blijft dus van belang ook in onderzoek te blijven rapporteren in wat ook niet lukt en feedback te organiseren welke redenen hiervoor te bedenken zijn.

Dank René Kneyber, voor het organiseren van deze ontmoeting met - naar mijn idee - toch wel dé expert op het terrein van formatief handelen. Ik denk dat met name de eerste dag en de wijze waarop Wiliam deze vormgaf, velen heeft geïnspireerd. Het terrein van formatief handelen en dat zelf in praktijk brengen is toch echt zijn ding. Alhoewel in de onderwijspraktijk nog vele vragen blijven bestaan over implementatie, scholing, facilitering en het spanningsveld tussen leren en examineren, is de wijze waarop docenten zelf steeds meer regie nemen in het beantwoorden van deze vragen (bijvoorbeeld in Facebookgroepen als [deze](#) en [deze](#)) inspirerend en aanstekelijk enthousiasmerend. Nu gaat het om de durf de focus weer te leggen op het verzorgen van goed, formatief onderwijs, en wellicht wat minder op de instrumentele benadering van de summatieve functie van toetsing. Ik kijk uit naar een nieuwe ontmoeting met Wiliam in 2028.

Dylan Wiliam is dé expert op het terrain van formative assessment. Zijn boek 'Embedded formative assessment' biedt veel concrete handvatten formatief handelen in de dagelijkse onderwijspraktijk vorm te geven. Ook zijn boek 'Leadership for teacher learning: Creating a culture where all teachers

improve so that all students succeed' kan ik van harte aanbevelen. Meer weten over Wiliam? Op zijn websites (zie [hier](#) en [hier](#)) zijn veel webinars en interviews te vinden.