

Naar formatieve evaluatie van onderwijs voor het bevorderen van *teacher agency*

Max Kusters en Mario de Jonge

Samenvatting Het efficiëntie denken binnen universiteiten legt een sterke nadruk op meetbare prestaties, zoals publicaties en studenttevredenheid. Hierdoor wordt de professionele ontwikkeling van docenten naar ons oordeel belemmerd. In dit opiniërend artikel verkennen we hoe de huidige evaluatiemethoden bijdragen aan *teacher agency*, i.e., het vermogen van docenten om weloverwogen en doelgericht hun onderwijs te verbeteren. Veel universiteiten gebruiken de Student Evaluation of Teaching (SET) als maatstaf voor docentprestaties, maar deze methode is problematisch. Onderzoek toont aan dat SETs onbetrouwbaar zijn, gevoelig voor gender- en raciale bias, en kunnen leiden tot cijferinflatie. Bovendien moedigen SETs docenten aan om tegemoet te komen aan de wensen van studenten in plaats van uitdagende leermogelijkheden aan te bieden. Als alternatief pleiten wij voor formatieve evaluatie, waarin constructieve feedback en professionele ontwikkeling centraal staan. Dit sluit beter aan bij initiatieven zoals *Erkennen & Waarderen*, die een meer holistische benadering van academische waardering bevorderen. Door evaluatiesystemen te herzien en *teacher agency* te versterken, kunnen docenten hun onderwijs continu verbeteren, afgestemd op zowel de leerbehoeften van studenten als de bredere academische en maatschappelijke doelen.

Trefwoorden teacher agency, student evaluaties van onderwijs, formatieve evaluaties van onderwijs

Artikelgeschiedenis

Ontvangen: 24 februari 2025
Geaccepteerd: 6 september 2025
Online: 22 september 2025

Contactpersoon

Max Kusters, m.c.j.kusters@rug.nl

Over de auteur(s)

Max Kusters is Universitair Docent Onderwijs-wetenschappen, GION onderwijs/onderzoek, Rijksuniversiteit Groningen; Mario de Jonge is Onderwijskundig adviseur & Onderzoeker, ICLON Graduate School of Teaching, Universiteit Leiden.

Copyright

© Author(s); licensed under Creative Commons Attribution 4.0. This allows for unrestricted use, as long as the author(s) and source are credited.

Introductie

Binnen universiteiten in Nederland ligt de nadruk sterk op efficiëntie en meetbaarheid (Visser et al., 2024). Deze focus komt onder andere tot uiting in de manier waarop onderwijskwaliteit wordt beoordeeld, met name via de *Student Evaluation of Teaching* (SET). Deze evaluaties – kortgezegd evaluaties van het onderwijs doorgaans in de vorm van vragenlijsten die studenten aan het eind van een cursus invullen – passen binnen een breder, instrumenteel discours waarin onderwijspraktijken worden teruggebracht tot eenvoudig te kwantificeren indicatoren (Olssen & Peters, 2005). Voorstanders van de SET benadrukken dat dit meetinstrument snel en eenvoudig betrouwbare en valide data van onderwijskwaliteit oplevert. Het gebruik van SETs brengt echter ook het risico met zich

mee dat complexe onderwijskwaliteit wordt gereduceerd tot oppervlakkige tevredenheidsscores (Ali et al., 2021). Deze instrumentele benadering roept daarom fundamentele vragen op over de waarde van de SET als beoordelingsinstrument van onderwijskwaliteit (cf. Stroebe, 2020). Bovendien is het de vraag wat het docenten daadwerkelijk oplevert – afgezien van een cijfermatig oordeel dat nauwelijks ruimte laat voor professionele reflectie of inhoudelijke diepgang. Daarbij komt dat dergelijke evaluatiepraktijken weinig ruimte laten voor *teacher agency* – het vermogen van docenten om weloverwogen keuzes te maken en doelgericht te handelen in hun onderwijspraktijk. Immers, wanneer onderwijskwaliteit primair wordt afgemeten aan studenttevredenheidscijfers, kan dit de mogelijkheid ondermijnen om docentprofessionalisering, inhoudelijke vernieuwing en de kwaliteit van onderwijs te bevorderen (cf. Imants & Van der Wal, 2020).

In dit opiniërend artikel staat de vraag centraal in hoeverre de traditionele SET nog van toegevoegde waarde is voor het hoger onderwijs. De meeste SETs bevatten in de regel, naast een overwegend kwantitatieve beoordelingscomponent, ook een paar afsluitende open vragen om kwalitatieve data te verzamelen. Vooropgesteld, het zou te verstrekkend zijn om te stellen dat de SET in de huidige vorm helemaal geen waarde heeft. De uiteindelijke vraag zou wat ons betreft daarom moeten zijn of de traditionele SET *voldoende* toegevoegde waarde heeft om de brede en voortdurende inzet te rechtvaardigen. Om alvast op de zaken vooruit te lopen, dit laatste zullen wij in ons betoog nadrukkelijk in twijfel trekken. Universiteiten spenderen kostbare tijd en middelen voor het afnemen, verwerken en interpreteren van SETs, en daarnaast vragen zij hun studenten hiervoor een inspanning te leveren. Op basis van de literatuur zullen wij beargumenteren dat deze tijd en middelen veel beter zouden kunnen worden besteed dan nu het geval is. Zeker in het licht van recente ontwikkelingen denken wij dat de tijd nu rijp is voor een grondige herziening van de evaluatie van hoger onderwijs.

Recente ontwikkelingen

Tegenover de dominante kwantitatieve beoordelingscultuur staan initiatieven die pleiten voor een bredere waardering van academici. Het landelijke programma *Erkennen & Waarderen* (VSNU et al., 2019) beoogt bijvoorbeeld een holistische benadering waarin de verscheidenheid aan taken van academici worden erkend. Hiermee wordt, naast leiderschap, maatschappelijke impact en samenwerking, ook expliciet het belang van de onderwijskundige bijdrage van docenten benadrukt. Met betrekking tot het universitair onderwijs betekent dit dat onderwijs geëvalueerd zou moeten worden op een manier die beter recht doet aan de professionele ontwikkeling en behoeften van de docent (cf. Graham, 2018).

In navolging van *Erkennen & Waarderen* herzien veel Nederlandse universiteiten hun HR-beleid. Zo introduceerde de Universiteit Leiden begin dit jaar een vernieuwde opzet van jaargesprekken onder de naam GROW (*Gesprekken over Resultaat, Ontwikkeling en Wel-*

*zijn*¹⁾. GROW markeert een verschuiving van een summatieve benadering met jaarlijkse eendoordelen naar een meer formatieve aanpak. De traditionele beoordelingsgesprekken maken plaats voor een doorlopende dialoog, waarin het niet alleen draait om *wat* iemand bereikt, maar juist ook om *hoe* en *waarom*. GROW nodigt dus uit tot bredere, rijkere vormen van reflectie en gesprek (cf. Academia in Motion, z.d.).

Ondanks deze ontwikkelingen blijven SETs en vergelijkbare evaluatie-instrumenten een dominante rol spelen in de beoordeling van onderwijskwaliteit. Hoewel SETs inzicht kunnen bieden in de studentbeleving, zijn ze oorspronkelijk bedoeld als feedbackinstrument, niet als beoordelingsmaatstaf. In de praktijk worden ze echter vaak wél op die manier gebruikt (Hammonds et al., 2017). Dit zet druk op docenten om te voldoen aan studentverwachtingen – zoals populariteit, toegankelijkheid en tevredenheid – in plaats van zich te richten op het stimuleren van kritisch denken en intellectuele verdieping (Stroebe, 2020). Daarmee dreigen SETs de kern van academisch onderwijs te ondermijnen. Bovendien leveren ze doorgaans weinig bruikbare inzichten op voor inhoudelijke of didactische verbetering. In het licht van landelijke en institutionele bewegingen richting een bredere waardering van onderwijs is het van cruciaal belang om feedback op onderwijs meer te richten op professionele groei en inhoudelijke ontwikkeling van docenten, en minder op oppervlakkige tevredenheidsscores.

Er is veel geschreven over de mogelijke rol van SETs in het verbeteren van de hoger onderwijspraktijk. Wij zullen eerst kort beschrijven hoe SETs doorgaans worden ingezet in het HO, en daarna bespreken wij de problemen met de SET

Het gebruik van studentevaluaties van onderwijs

Aan de meeste hoger onderwijsinstellingen wordt het onderwijs geëvalueerd via SETs. De SET is een vragenlijst voor een bepaald vak of een aantal vakken/onderdelen waarin studenten via een reeks Likertschaalvragen het genoten onderwijs beoordelen. Stroebe (2020) beschrijft in zijn overzichtsartikel hoe deze vorm van evalueren in de jaren twintig van de vorige eeuw werd ontwikkeld door Remmers & Brandenburg (1927) en Guthrie (1953). Het oorspronkelijke doel van de SET was om docenten van feedback te voorzien om hun onderwijs mee te kunnen verbeteren. Saillant detail hierbij is dat het eigenlijk niet de bedoeling was dat de evaluatieresultaten ook zouden dienen als input voor andere belanghebbenden dan de docent.

Het gebruik van SETs nam in de jaren '70 een vlucht en werd tussen de jaren '80 en '90 gemeengoed bij de meeste instellingen van hoger onderwijs wereldwijd (Rojstaczer & Healy, 2012). Later zouden de evaluatieresultaten ook gebruikt worden voor het beoordelen van onderwijsprogramma's als geheel, of voor het beoordelen van en beslissen over de promotie van individuele docenten (Chalmers & Hunt, 2016). Met name dit laatste punt stuitte al regelmatig op felle kritiek (e.g., Close, 2024; Heffernan, 2022; Uttl, 2021). Heffernan stelde bijvoorbeeld dat universiteiten die SETs blijven inzetten geen

veilige inclusieve werkomgeving zouden creëren voor docenten. Hij riep universiteiten op tot een onmiddellijke verandering van het beleid op dit gebied. Uttl (2021) deed er nog een schepje bovenop en stelde zelfs dat de formele beoordeling van docenten met SETs gezien kan worden als een schending van de mensenrechten. In het verleden werden er al verschillende concrete aanbevelingen gedaan voor verbetering op dit gebied (e.g., Algozzine et al., 2004) en blijkens initiatieven zoals het eerdergenoemde GROW wordt hier nu ook gehoor aan gegeven door sommige universiteiten in Nederland.

In dit betoog zullen wij ons niet zozeer buigen over de ethiek van studentevaluaties met betrekking tot het nemen van personele beslissingen en beoordeling. In plaats daarvan richten wij ons specifiek op het doel waar de SET oorspronkelijk voor werd ontworpen, namelijk het verkrijgen van feedback voor de docent om het onderwijs te verbeteren. Er zijn sterke aanwijzingen dat ook dit doel niet voldoende wordt gediend door de SET.

De problemen met de SET

De problematiek betreft: (1) De onbetrouwbaarheid van SETs in relatie tot onbedachtzaamheid en vooroordelen van studenten, (2) Het twijfelachtige verband tussen SETs en studentprestaties, (3) De beperkte bijdrage van SETs aan onderwijskwaliteit, en (4) De beperkte bijdrage van SETs voor docentprofessionaliseringdoeleinden en *teacher agency*.

De onbetrouwbaarheid van SETs

SETs zijn hoofdzakelijk kwantitatieve meetinstrumenten. De kwaliteit van ieder meetinstrument is grotendeels afhankelijk van de mate van betrouwbaarheid waarmee metingen kunnen worden verricht. Van SETs is al langer bekend dat de betrouwbaarheid van de meting soms van twijfelachtig niveau is. Hiervoor zijn twee redenen: De onbedachtzaamheid tijdens het invullen, en vooroordelen die hun intrede doen bij het invullen van de vragenlijst.

Onbedachtzaamheid

Voorstanders van de SET zullen wellicht beargumenteren dat de betrouwbaarheid van SETs deels geborgd is als de respons groot genoeg is. Dit zou ook een reden kunnen zijn voor sommige instellingen om het invullen van de SET verplicht te maken voor hun studenten. Aan de andere kant, wordt de lage respons voor facultatieve (online) SETs an sich ook wel als een teken aan de wand gezien dat studenten over het algemeen weinig betrokken zijn bij de evaluatie (Anderson et al., 2006). Daarenboven zijn er verschillende studies die laten zien dat de resultaten van SETs vaak met een flinke korrel zout moeten worden genomen (e.g., Reynolds, 1977; Uijtdehaage en O'Neal, 2015). Bijvoorbeeld in de studie van Uijtdehaage en O'Neal (2015) werden er, in de reguliere studentevaluaties van

twee vakken, naast de echte docenten, met opzet ook enkele zogenaamde spookdocenten (met fictieve namen als “Pat Turner” of “Chris Miller”) en fictieve college titels (e.g., “Longziekte, een introductie”) opgevoerd. De resultaten lieten zien dat een meerderheid van studenten er doodleuk voor koos om de spookdocenten te evalueren terwijl zij ook de mogelijkheid hadden dit niet te doen. Toen er naast fictieve namen van de spookdocenten ook nog portretfoto’s werden toegevoegd, daalde het aantal onbedachtzame evaluaties weliswaar, maar nog steeds werden de spookdocenten door nagenoeg de helft van studenten ten onrechte geëvalueerd.

De resultaten Uijtdehaage en O’Neal (2015) bevestigen die uit eerder onderzoek (Reynolds, 1977), en zijn ook in lijn met de bevinding dat SETs ondermaats presteren op de zogenaamde test-hertest betrouwbaarheid (e.g., Vanacore en Pellegrino, 2019). Niet gespeend van enige ironie concludeerde Reynolds in 1977 al dat het wellicht helemaal niet nodig is om tot het eind van een vak te wachten met evalueren. Studenten zouden een SET ook best direct aan het begin van een vak alvast kunnen invullen wanneer zij nog fris en alert zijn.

Vooroordelen

Naast de onbetrouwbaarheid als gevolg van onbedachtzaamheid, is er nog een andere reden om de resultaten van studentevaluaties te wantrouwen. Volgens Michael Birnbaum is er namelijk een fundamenteel probleem met de opzet van vragenlijsten zoals de SET. Respondenten zouden geneigd zijn om, bij gebrek aan een concreet referentiekader, zich een eigen schaal voor de geest te halen. Birnbaum (1999) liet zien dat dit uiteindelijk kan resulteren in absurde en ontegenzeggelijk incorrecte conclusies (e.g., het getal “221” dat als hoger wordt ingeschat dan het getal “9” op een beoordelings-schaal). Het mechanisme waarbij studenten zichzelf een schaal voor de geest halen zou ook de deur openzetten voor mogelijk vooroordelen die hun intrede kunnen doen in SETs.

Dat de deur wagenwijd open staat blijkt wel uit tal van overzichtsartikelen die een hele reeks vooroordelen bespreken die de resultaten van SETs zouden kunnen beïnvloeden (bv., Heffernan, 2022; Neath, 1996; Uttl, 2021). Het valt buiten het bereik van dit artikel om alle in de literatuur gedocumenteerde vooroordelen de revue te laten passeren. Ter illustratie beperken wij ons hier tot wat misschien wel het meestgenoemde en meest controverste voorbeeld is, namelijk dat van seksistische vooroordelen. Neath (1996) stelde op basis van zijn literatuurreview dat studenten misschien (onbewust) meer ondersteuning verwachten van vrouwelijke dan van mannelijke docenten. Wanneer vrouwelijke docenten echter niet aan dit verwachtingspatroon voldoen, kunnen studenten geneigd zijn hun beoordeling vervolgens naar beneden bij te stellen (Bennett, 1982; Unger, 1997). Recentelijk werd het effect van seksistische vooroordelen nog onderzocht in een *online setting* waarbij studenten de betrokken docent nooit in het echt kregen te zien (MacNell et al., 2015). Deze setting biedt de unieke mogelijkheid om het geslacht van de docent te manipuleren door studenten hierover simpelweg verkeerd te informeren. MacNell et al.

vonden dat docenten een significant hogere beoordeling kregen wanneer de studenten dachten dat het een man betrof. Om Birnbaum (1999) in herinnering te roepen, er werd duidelijk met twee verschillende maten gemeten.

Voorstanders van de SET zouden kunnen beargumenteren dat de soep heus niet zo heet gegeten wordt als dat zij wordt opgediend. Een recent overzichtsartikel van Kreitzer & Cushman (2021) stelt bijvoorbeeld dat de mate waarin seksistische vooroordelen een rol spelen in de SET ook afhankelijk is van andere (context)variabelen. Er zijn ook situaties waarin juist vrouwen een hogere beoordeling krijgen op de SET dan mannen. Het is echter belangrijk om te benadrukken dat het probleem wat ons betreft niet zozeer zit in *hoe* vooroordelen de SET beïnvloeden. Het probleem is de vatbaarheid voor vooroordelen in het algemeen. De exacte vooroordelen die op een bepaald moment of in een bepaalde context wel of niet een rol spelen is buiten de kwestie. Het doet wat ons betreft niets af aan het grotere fundamentele probleem met de SET dat het een onbetrouwbaar meetinstrument maakt. Er zijn weliswaar initiatieven en aanbevelingen geweest om de mate van bevooroordeeldheid in SETs zoveel mogelijk te bepreken bijvoorbeeld door de vraagstelling gericht te maken (e.g., Cook, 2024). Het lijkt echter praktisch onmogelijk om effectief te controleren of te corrigeren voor alle mogelijke vooroordelen in een SET (Uttl, 2021).

Het twijfelachtige verband tussen SETs en studentprestaties

Vooropgesteld, de resultaten van enkele vroege meta-analyses laten een kleine tot middelgrote correlatie zien tussen studentevaluaties van onderwijs en studentprestaties (e.g., Cohen, 1980, 1981; Feldman, 1989). Dit verband gaf voorstanders van de SET in de afgelopen 40 jaar wellicht voldoende aanleiding om te veronderstellen dat de SET een valide maat zou zijn voor het meten van onderwijskwaliteit (Uttl, 2021). Echter, een recente heranalyse van de resultaten van eerdere meta-analyses werpt een ander licht op de kwestie. Uttl et al. (2017) beargumenteerden dat er in de eerder uitgevoerde meta-analyses onvoldoende rekening is gehouden met de steekproefgroottes van de geïncludeerde studies, en mogelijke publicatie bias. Op basis van een meer geavanceerde uitgebreide meta-analyse vonden Uttl et al. dat er praktisch geen correlatie lijkt te zijn tussen SETs en studentprestaties. Zij concludeerden dat beoordelingen met studentvragenlijsten zoals de SET waarschijnlijk geen goede indicatie zijn als het gaat om het beoordelen van de effectiviteit van individuele docenten.

De beperkte bijdrage van SETs aan onderwijskwaliteit

In zowel de vroegere meta-analyses (e.g., Cohen, 1980, 1981; Feldman, 1989) als in die van Uttl (2017) werd er gekeken naar het verband tussen SET-beoordelingen en de studentprestaties aan het eind van een vak. Echter, er kan ook worden beargumenteerd dat de impact van eerdere genoten onderwijs pas na verloop van tijd zou moeten worden bekeken (i.e., na een langere termijn). Er zijn relatief weinig studies die dit hebben gedaan. Kornell en Hausman (2016) vatten de resultaten van twee van zulke studies (Braga et al.,

2014; Carell & West, 2010) samen in een kort overzichtsartikel. Wanneer men keek naar de bijdrage van docenten aan studentprestaties tijdens een later gegeven vervolgvak, dan leken juist docenten met een lagere beoordeling een grotere bijdrage te hebben geleverd.

Een mogelijke verklaring voor deze enigszins contra-intuïtieve bevinding is dat docenten die het makkelijker maken voor hun studenten misschien ook hogere beoordelingen krijgen (Kornell & Hausman, 2016; Stroebe, 2020). Stroebe (2020) beschrijft een model van het werkzame proces dat hieraan ten grondslag zou liggen. In het kort stelt Stroebe dat studenten makkelijker vakken en soepele beoordelingen van docenten belonen met een hogere beoordeling op de SET. Dit gegeven zorgt voor een perverse prikkel voor docenten om hun vakken makkelijker te maken en de lat steeds wat minder hoog te leggen. Het gevolg is een glijdende schaal waarbij het onderwijs makkelijker zou worden, en de beoordelingen die studenten van hun docenten krijgen steeds hoger. Het model van Stroebe is overigens geheel in lijn met het fenomeen dat bekend staat als *cijferinflatie*: De trend dat het gewogen gemiddelde cijfer van studenten in de Verenigde Staten (*National Trends in Grade Inflation, American Colleges and Universities*, n.d.; Rojstaczer, & Healy, 2012) en in Europa (François & De Witte, 2025) over een langere periode van tijd geleidelijk aan lijkt te stijgen. Studies waarin beleidsmatige interventies werden onderzocht om cijferinflatie tegen te gaan, lijken een verband tussen beoordelingen en SETs ook te bevestigen. Zo vonden Butcher en collega's (2014) dat maatregelen om buitensporig hoge beoordelingen in verschillende studieprogramma's in te perken vervolgens resulteerden in een lagere SET-beoordeling van de betreffende programma's.

Samenvattend, lijkt het er sterk op dat het inzetten van SETs in het hoger onderwijs op zichzelf een negatief effect kan hebben op datgene wat het beoogt te meten. Uit vragenlijstonderzoek blijkt althans dat docenten zich bewust zijn van de perverse prikkel die uitgaat van de SET. Lakeman et al. (2022) vonden bijvoorbeeld dat docenten aangaven vaak moeite te hebben met het vinden van een goede balans tussen onderwijskwaliteit aan de ene kant en studenttevredenheid aan de andere kant.

De beperkte bijdrage van SETs voor docentprofessionaliseringdoeleinden en teacher agency

Om onderwijs- en docentontwikkeling optimaal te stimuleren, zijn drie voorwaarden essentieel: 1) een actieve stimulans vanuit beleidsmakers, 2) betrokken docenten die de waarde en noodzaak van onderwijskwaliteit onderschrijven, en 3) de mogelijkheid voor docenten om *teacher agency* te tonen en te ontwikkelen (cf. Adams et al., 2023). Om aan deze voorwaarden te voldoen, hebben docenten – naast voldoende kennis van hun vakgebied en didactiek – ook toegang nodig tot betekenisvolle feedback (Kusters et al., 2023).

Teacher agency verwijst naar het vermogen van docenten om weloverwogen en doelgerichte acties te ondernemen ten behoeve van onderwijsontwikkeling (Priestley et al., 2015). *Teacher agency* vormt daarmee een cruciale pijler voor professionele groei en onderwijsontwikkeling, omdat het docenten in staat stelt om proactief en reflectief hun onderwijspraktijken te verbeteren (Leijen et al., 2020; Vähäsantanen, 2015).

Echter, blijvende professionele ontwikkeling wordt op dit moment aan universiteiten onvoldoende gestimuleerd. Kusters et al. (2023) constateerden dat de formele professionele ontwikkeling van academici, via cursussen en kwalificaties bijvoorbeeld in het kader van de Basiskwalificatie Onderwijs, vaak al ophoudt na de eerste twee jaar van een *tenured* positie. Daarna leunen docenten wellicht voor een groot deel op de feedback verkregen via studentevaluaties van onderwijs. Het risico bestaat dus dat docenten hun *teacher agency* richten op, vanuit onderwijskundig perspectief, beperkte of zelfs misleidende indicatoren van onderwijsontwikkeling. Namelijk op SET-scores, die voornamelijk studenttevredenheid meten, maar weinig zeggen over didactische vaardigheden of inhoudelijke ontwikkeling (Stroebe, 2020). Hierdoor ontstaat het risico dat de *teacher agency* juist leidt tot het versterken van conformisme aan laagdrempelig onderwijs.

Conclusie

Wij stelden ons de vraag of de traditionele SET voldoende toegevoegde waarde heeft om de brede en voortdurende inzet te rechtvaardigen. Op basis van de besproken ontwikkelingen en problemen lijkt de waarde van de traditionele SET hiervoor te beperkt. Hoewel SETs inzicht kunnen bieden in de studentbeleving, bieden zij onvoldoende basis voor betekenisvolle verbetering van onderwijskwaliteit. Het is daarom wenselijk dat universiteiten hun HR-beleid herzien en daarbij meer ruimte creëren voor *teacher agency* en bredere vormen van feedback en reflectie.

Aanbevelingen voor HR-beleid

Het HR-beleid dat *teacher agency* ondersteunt, schenkt aandacht aan de feedbackbronnen, maakt dialogische feedback mogelijk, en let op de kwaliteit en wijze waarop feedback wordt aangeboden.

Welke bronnen dragen wel bij aan ondersteuning van teacher agency?

Een sleutelfactor in het ontwikkelen van *teacher agency* is toegang tot waardevolle feedbackbronnen, die docenten ondersteunen in hun reflectie, professionele groei en strategische besluitvorming (Leijen et al., 2020). Om *teacher agency* te bereiken, hebben docenten toegang nodig tot feedback die hen inzicht geeft in hun onderwijspraktijk en hun impact op het leren van studenten. Feedback moet verder gaan dan alleen een oordeel; het moet docenten de mogelijkheid bieden om kritisch te reflecteren, verbanden te leggen met hun professionele doelen, en effectieve beslissingen te nemen over hun onderwijsstrategieën. Het aanreiken van inzichtelijke feedback stelt docenten in staat om zich bewust te worden van hun sterktes, zwaktes en de bredere context waarin zij werken, wat essentieel is voor het maken van weloverwogen en geïnformeerde beslissingen (Ali

et al., 2021). Hier ligt overigens ook een dankbare taak voor onderwijsonderzoekers. Hoewel de meeste studies het belang lijken te onderkennen van studentenfeedback om de onderwijskwaliteit te verbeteren, is het vooralsnog onduidelijk hoe studentenfeedback in de praktijk gebruikt wordt door docenten om verbeteringen door te voeren in het onderwijs (Algozzine et al., 2006; Harrison et al., 2022).

Dialogische feedback

Onderzoek toont aan dat dialogische feedback, waarbij docenten actief betrokken worden bij het interpreteren van data en het formuleren van actieplannen, de potentie heeft om *teacher agency* te bevorderen. Dialogische feedback creëert een co-constructieve leeromgeving, waarin docenten samen met collega's en studenten reflecteren op hun onderwijsmethoden en manieren zoeken om deze te verbeteren. In plaats van een top-down feedbackmodel, waarin docenten slechts passieve ontvangers zijn, stimuleert dialogische feedback een actieve rol in het leerproces (Charteris & Smardon, 2015). Dit proces moedigt zelfreflectie aan, versterkt de autonomie, en biedt een kader waarin docenten bewuste keuzes kunnen maken en verantwoordelijkheid nemen voor hun professionele ontwikkeling. Daarnaast onderzochten Verberg et al. (2016) hoe formatieve feedback een belangrijke rol speelt in het bevorderen van *teacher agency*. In hun onderzoek tonen zij aan dat wanneer docenten feedback niet slechts passief ontvangen, maar actief interpreteren en inbedden in hun professionele doelen, zij een verhoogd gevoel van controle en initiatief ontwikkelen. Dit sluit aan bij de theorie van Emirbayer en Mische (1998), die *agency* definiëren als een proces dat voortkomt uit de interactie tussen verleden, heden en toekomst, waarbij eerdere ervaringen, huidige contexten en toekomstige doelen allemaal bijdragen aan de besluitvorming van de docent.

Kwaliteit en manier van feedback

Naast de inhoud van de feedback is de kwaliteit en de manier waarop feedback wordt aangeboden cruciaal. Carless (2013) stelt dat vertrouwen een fundamenteel element is van effectieve feedbackprocessen. Wanneer docenten zich veilig voelen binnen een feedbackcultuur, zijn ze eerder geneigd om zich open te stellen voor kritische reflectie, en om daadwerkelijk actie te ondernemen op basis van ontvangen feedback. Dit is vooral van belang in universiteiten waar feedback niet zozeer als controlemechanisme wordt ingezet, maar als een middel om samen leren te bevorderen (Groundwater-Smith & Mockler, 2009).

Pleidooi voor formatieve onderwijsbeoordelingen

In dit opiniërend artikel stellen wij dat kwantitatieve onderwijsbeoordelingen hun langste tijd gehad hebben. We zullen de evaluatie van onderwijs moeten herzien op een manier die recht doet aan de complexiteit en diversiteit van onderwijs en -ontwikkeling. Wij betogen

dat formatieve evaluatiemethoden – waarbij de nadruk ligt op constructieve feedback en continue verbetering – beter aansluiten bij de doelstelling om de ontwikkeling van *teacher agency* te ondersteunen. Hierbij is overigens ook de *timing* van de evaluatie het (her)overwegen waard (Hounsel, 2006). Zo kan het voor formatieve doeleinden aan te raden zijn om een vak tussentijds te evalueren in plaats van aan het einde van een vak (Algozzine et al., 2004). Dit kan als bijkomend voordeel hebben dat de betrokkenheid bij de evaluatie onder de studenten wordt verhoogd, en er minder sprake is van onbedachtzaamheid bij het geven van feedback.

Door feedback op een systematische manier te integreren in professionele ontwikkeling, kunnen universiteiten een cultuur creëren waarin docenten zich gesteund en gemootiveerd voelen om proactief hun onderwijspraktijken te verbeteren. In een tijd waarin het onderwijs steeds meer onder druk staat van externe verantwoordingsmodellen, biedt een *agency*-gecentreerde benadering docenten de autonomie en het vertrouwen om hun onderwijs vorm te geven op basis van hun professionele expertise en leerlinggerichte inzichten. Kortom, om *teacher agency* te bereiken, is feedback meer dan een hulpmiddel; het is een katalysator voor professionele groei en weloverwogen besluitvorming. Inzichtelijke feedback stelt docenten in staat om hun onderwijspraktijk niet alleen te verbeteren, maar ook om deze bewust en strategisch vorm te geven binnen de specifieke context waarin zij werken.

Onderwijs is een multidimensionale activiteit die niet adequaat kan worden beoordeeld met overwegend kwantitatieve data uit een gestandaardiseerde vragenlijst. In lijn met andere recente publicaties (e.g., Cook et al., 2024; Uttl, B., 2024) en initiatieven (de Jonge en Veenstra, 2024) pleiten wij voor meer formatieve evaluaties waarin kwalitatieve feedback centraal staat. Evaluaties die gericht zijn op verbetering in plaats van beoordeling. Om dit te realiseren moeten bestaande evaluatiesystemen fundamenteel worden herzien. We zullen moeten erkennen dat studentenevaluaties van onderwijs te veel zijn gebaseerd op globale, cijfermatige oordelen die weinig zeggen over de werkelijke onderwijskwaliteit (Algozzine et al., 2004). Naar onze mening is de overwegend kwantitatieve component in reguliere studentevaluaties een wassen neus die geen potentieel heeft om een bijdrage te leveren aan kwaliteit van onderwijs. Initiatieven om de SET te verbeteren of te repareren zijn ongetwijfeld goed bedoeld, maar plaveien, om de woorden van Uttl (2024) aan te halen “de snelweg naar de hel”. Niet langer zou cijfermatige beoordeling de norm moeten zijn. Er is een evaluatiecultuur nodig die gericht is op professionele groei, waarbij kwalitatieve feedback fungeert als katalysator voor reflectie en ontwikkeling. Binnen zo’n cultuur kunnen docenten *teacher agency* ontwikkelen, en zich positioneren als actieve en verantwoordelijke professionals die in staat zijn om effectief in te spelen op de sociale, structurele en culturele context van hun onderwijs.

Noten

- 1 <https://www.medewerkers.universiteitleid.nl/po/ontwikkeling/grow?cf=interfacultaire-instituten&cd=iclon>

Literatuur

- Academia in Motion* (z.d.). Universiteit Leiden. <https://www.universiteitleid.nl/dossiers/academia-in-motion>
- Adams, B., Thomas, L., Moens, M., & Valcke, M. (2023). University teachers as versatile masters: Evaluating the effectiveness of a professional development programme on student-centred teaching competencies. *Studies in Educational Evaluation*, 77, 101260. <https://doi.org/10.1016/j.stueduc.2023.101260>
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., Mohanty, G., & Spooner, F. (2004). Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching*, 52(4), 134–141. <http://www.jstor.org/stable/27559201>
- Ali, A., Crawford, J., Cejnar, L., Harman, K., & Sim, K. (2021). What student evaluations are not: Scholarship of Teaching and Learning using student evaluations. *Journal of University Teaching & Learning Practice*, 18(8). <https://doi.org/10.53761/1.18.8.1>
- Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. *Innovate: Journal of Online Education*, 2(6), 1–7.
- Bennett, S.K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74(2), 170–179. <https://doi.org/10.1037/0022-0663.74.2.170>
- Birnbaum, M.H. (1999). How to Show That $g > 221$: Collect Judgments in a Between-Subjects Design. *Psychological Methods*, 4(3), 243–249. <https://doi.org/10.1037/1082-989X.4.3.243>
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88. <https://doi.org/10.1016/j.econedurev.2014.04.002>
- Butcher, K.F., McEwan, P.J., & Weerapana, A. (2014). The Effects of an Anti-Grade Inflation Policy at Wellesley College. *The Journal of Economic Perspectives*, 28(3), 189–204. <https://doi.org/10.1257/jep.28.3.189>
- Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In *Routledge eBooks* (pp. 100–113). <https://doi.org/10.4324/9780203074336-11>
- Chalmers, D., & Hunt, L. (2016). Evaluation of teaching. *HERDSA Review of Higher Education*, 3(27), 25–55.
- Charteris, J., & Smardon, D. (2015). Teacher agency and dialogic feedback: Using classroom data for practitioner inquiry. *Teaching and Teacher Education*, 50, 114–123. <https://doi.org/10.1016/j.tate.2015.05.006>
- Close, D. (2024). Why Student Ratings of Faculty Are Unethical. *Journal of Academic Ethics*. <https://doi.org/10.1007/s10805-024-09562-3>

- Cohen, P.A. (1980). *A meta-analysis of the relationship between student ratings of instruction and student achievement*. University of Michigan.
- Cohen, P.A. (1981). Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies. *Review of Educational Research*, 51(3), 281–309. <https://doi.org/10.3102/00346543051003281>
- Cong-Lem, N. (2021). Teacher agency: A systematic review of international literature. *Issues in Educational Research*, 31(3), 718–738. <https://search.informit.org/doi/10.3316/informit.190851857034060>
- Cook, S., Watson, D., & Webb, R. (2024). Performance evaluation in teaching: Dissecting student evaluations in higher education. *Studies in Educational Evaluation*, 81, Article 101342. <https://doi.org/10.1016/j.stueduc.2024.101342>
- De Jonge, M. & Veenstra E. (2024, 10 oktober). *AI-supported formative evaluation of teaching*. Teaching and Teacher Learning. <https://researchblog.iclon.nl/ai-supported-formative-evaluation-of-teaching/>
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, 103(4), 962–1023. <https://doi.org/10.1086/231294>
- Feldman, K.A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30(6), 583–645. <https://doi.org/10.1007/bf00992392>
- François, M. & De Witte, K. (2025). A decade of grade inflation boosted by the COVID-19 pandemic – An empirical analysis of a top European university. *British Educational Research Journal*, 00, 1–69. <https://doi.org/10.1002/berj.4172>
- Graham, R. (2018). *Career framework for university teaching*. Retrieved from <https://www.rhgraham.org/resources/Career-Framework-University-Guidance-Note.pdf>
- Groundwater-Smith, S., & Mockler, N. (2009). *Teacher Professional Learning in an Age of Compliance: Mind the Gap*. Springer. 10.1007/978-1-4020-9417-0
- Guthrie, E.R. (1953). The Evaluation of Teaching. *The American Journal of Nursing*, 53(2), 220–221. <https://doi.org/10.1097/00000446-195302000-00053>
- Hammonds, F., Mariano, G.J., Ammons, G., & Chambers, S. (2017). Student evaluations of teaching: Improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education*, 21(1), 26–33. <https://doi.org/10.1080/13603108.2016.1227388>
- Heffernan, T. (2022). Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment and Evaluation in Higher Education*, 47(1), 144–154. <https://doi.org/10.1080/02602938.2021.1888075>
- Kusters, M., Van der Rijst, R., De Vetten, A., & Admiraal, W. (2023). University lecturers as change agents: How do they perceive their professional agency? *Teaching and Teacher Education*, 127, 104097. <https://doi.org/10.1016/j.tate.2023.104097>
- Lakeman, R., Coutts, R., Hutchinson, M., Massey, D., Nasrawi, D., Fielden, J., & Lee, M. (2022). Playing the SET game: how teachers view the impact of student evaluation on the experience of teaching and learning. *Assessment & Evaluation in Higher Education*, 48(6), 749–759. <https://doi.org/10.1080/02602938.2022.2126430>

- Leijen, Ä., Pedaste, M., & Lepp, L. (2020). Teacher Agency following the Ecological Model: How it is achieved and how it could be strengthened by different types of reflection. *British Journal of Educational Studies*, 68(3), 295–310. <https://doi.org/10.1080/00071005.2019.1672855>
- MacNell, L., Driscoll, A., & Hunt, A.N. (2015). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, 40(4), 291–303. <https://doi.org/10.1007/s10755-014-9313-4>
- National trends in grade inflation, American Colleges and universities.* (n.d.). <https://www.gradeinflation.com/>
- Neath, I. (1996). How to Improve Your Teaching Evaluations without Improving Your Teaching. *Psychological Reports*, 78(3_suppl), 1363–1372. <https://doi.org/10.2466/pro.1996.78.3c.1363>
- Olssen, M., & Peters, M.A. (2005). Neoliberalism, higher education and the knowledge economy: from the free market to knowledge capitalism. *Journal of Education Policy*, 20(3), 313–345. <https://doi.org/10.1080/02680930500108718>
- Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher Agency: An Ecological Approach*. Bloomsbury Academic.
- Remmers, H.H., & Brandenburg, G.C. (1927). Experimental data on the Purdue rating scale for instructors. *Educational Administration and Supervision*, 13(6), 399–406.
- Reynolds, D.V. (1977). Students Who Haven't Seen a Film on Sexuality and Communication Prefer it to a Lecture on The History of Psychology They Haven't Heard: Some Implications for the University. *Teaching of Psychology*, 4(8), 82–83.
- Rojstaczer, S., & Healy, C. (2012). Where a is Ordinary: The Evolution of American College and University Grading, 1940–2009. *Teachers College Record the Voice of Scholarship in Education*, 114(7), 1–23. <https://doi.org/10.1177/016146811211400707>
- Stroebe, W. (2020). Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation: A Theoretical and Empirical Analysis. *Basic and Applied Social Psychology*, 42(4), 276–294. <https://doi.org/10.1080/01973533.2020.1756817>
- Uijtdehaage, S., & O'Neal, C. (2015). A curious case of the phantom professor: mindless teaching evaluations by medical students. *Medical Education*, 49(9), 928–932. <https://doi.org/10.1111/medu.12647>
- Uttl, B. (2021). Lessons Learned from Research on Student Evaluation of Teaching in Higher Education. In: Rollett, W., Bijlsma, H., Röhl, S. (eds) *Student Feedback on Teaching in Schools*. Springer, Cham. https://doi.org/10.1007/978-3-030-75150-0_15
- Uttl, B. (2024). Student Evaluation of Teaching (SET): Why the Emperor has no clothes and what we should do about it. *Human Arenas*, 7, 403–437. <https://doi.org/10.1007/s42087-023-00361-7>
- Uttl, B., White, C.A., & Gonzalez, D.W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Vähäsantanen, K. (2015). Professional agency in the stream of change: Understanding educational change and teachers' professional identities. *Teaching and Teacher Education*, 47, 1–12. <https://doi.org/10.1016/j.tate.2014.11.006>

- Vanacore, A., & Pellegrino, M.S. (2019). How Reliable are Students' Evaluations of Teaching (SETs)? A Study to Test Student's Reproducibility and Repeatability. *Social Indicators Research*, 146(1–2), 77–89. <https://doi.org/10.1007/s11205-018-02055-y>
- Verberg, C.P.M., Tigelaar, D.E.H., Van Veen, K., & Verloop, N. (2016). Teacher agency within the context of formative teacher assessment: an in-depth analysis. *Educational Studies*, 42(5), 534–552. <https://doi.org/10.1080/03055698.2016.1231060>
- Visser, M., Stokes, P., Ashta, A., & Andersson, L.M. (2024). The 'performative' university: Theoretical and personal reflections. *Journal of Education Policy*, 39(6), 1030–1048. <https://doi.org/10.1080/02680939.2024.2403431>
- VSNU, NFU, KNAW, NWO, & ZonMw. (2019). *Room for Everyone's Talent: Towards a New Balance in Recognising and Rewarding Academics*. <https://recognitionrewards.nl/wp-content/uploads/2020/12/position-paper-room-for-everyones-talent.pdf>

Towards formative evaluation of teaching for fostering teacher agency

Abstract Efficiency thinking within universities places a strong emphasis on measurable performance, such as publications and student satisfaction. As a result, the professional development of teachers is, in our opinion, hindered. This opinion article explores how evaluation methods can contribute to teacher agency, i.e., the capacity of teachers to make deliberate and goal-oriented improvements to their teaching. Many universities use the Student Evaluation of Teaching (SET) as a measure of teaching performance, but this method is problematic. Research shows that SETs are unreliable, susceptible to gender and racial bias, and can contribute to grade inflation. Moreover, SETs encourage teachers to cater to students' wishes instead of offering challenging learning opportunities. As an alternative, the authors advocate for formative evaluation, which emphasizes constructive feedback and professional development. This approach aligns more closely with initiatives such as *Recognition & Rewards*, which promote a more holistic approach to academic recognition. By revising evaluation systems and strengthening teacher agency, teachers can continuously enhance their teaching, ensuring it aligns with both students' learning needs and broader academic and societal goals.

Keywords teacher agency, student evaluations of teaching, formative evaluations of teaching