

Opinie: studentenevaluaties in het hoger onderwijs

*Pieter Spooren**

Deze opiniebijdrage gaat over de evaluaties door studenten uit het hoger onderwijs van de opleidingsonderdelen (vakken) die ze hebben gevolgd. Aan zowat alle instellingen voor hoger onderwijs zijn deze evaluaties een belangrijke parameter van de kwaliteit van het verstrekte onderwijs en worden ze gebruikt voor zowel formatieve als summatieve doeleinden. Oordelen van studenten kunnen een belangrijke rol spelen in de optimalisering van de onderwijspraktijk, maar dienen met de grootste zorg te worden verzameld en geïnterpreteerd. Aan de hand van inzichten uit de recente onderzoeksliteratuur over dit thema worden enkele suggesties geformuleerd voor verder onderzoek en voor de praktijk.

Introductie

Iedereen die een beetje vertrouwd is met het hoger onderwijs van vandaag, weet dat docenten aan hogescholen en universiteiten voortdurend worden beoordeeld (Larsen, 2005). Voor de Vlaamse universiteitsprofessoren, bijvoorbeeld, is decretaal bepaald dat hun drie kerntaken, onderzoek, onderwijs en dienstverlening, vijfjaarlijks worden geëvalueerd. Een dergelijke evaluatie gebeurt op basis van een uitgebreid personeelsdossier, waarin parameters van elk van de kerntaken worden opgesomd. Zo wordt op het vlak van onderzoek onder andere gekeken naar het aantal wetenschappelijke publicaties en de verworven financiering voor onderzoeksprojecten. Onder dienstverlening vallen de bijdragen aan het maatschappelijke debat, lidmaatschappen van werkgroepen en commissies, en bestuursfuncties binnen de instelling. Voor wat onderwijs betreft wordt onder andere gekeken naar het aantal opleidingsonderdelen dat men doceert en hoeveel studenten daarmee worden bediend. Een belangrijke parameter voor de kwaliteit van het verstrekte onderwijs zijn bovendien de studentenevaluaties van opleidingsonderdelen.

Studenten worden immers op regelmatige basis bevraagd over de opleidingsonderdelen die zij het voorafgaande semester of academiejaar hebben gevolgd. Daarbij wordt onder andere gepeild naar hun visie op de pedagogische kwaliteiten van de lesgever, de kwaliteit van het studiemateriaal en de opbouw van het opleidingsonderdeel. Dit gebeurt doorgaans door middel van een meer of minder uitgebreide vragenlijst, die wordt afgenomen tijdens een college of eindtoets, of via een elektronische enquête. De didactische vaardigheden van de docenten worden dus beoordeeld op basis van studentenevaluaties, die de dominante (en in een

* Dr. Pieter Spooren (pieter.spooren@ua.ac.be) is werkzaam bij de Faculteit Politieke en Sociale Wetenschappen (PSW) van de Universiteit Antwerpen

aantal gevallen slechts de enige) indicator vormen voor de kwaliteit van het onderwijs. De resultaten van deze evaluaties worden besproken met de betrokken docent ter verdere verbetering van de onderwijspraktijk. Ook beleidsverantwoordelijken (onderwijs- of evaluatiecommissies) krijgen inzage in de resultaten, die in vele gevallen worden gebruikt bij beslissingen over bijvoorbeeld een vaste benoeming of promotie.

De organisatie van onderwijsevaluaties op regelmatige tijdstippen, met uitgewerkte procedures en gestandaardiseerde instrumenten, is bovendien een belangrijk aandachtspunt bij de instellingstoets kwaliteitszorg van de NVAO. Standaard 3 van het beoordelingskader stelt immers dat 'de instelling zicht heeft op de mate waarin haar visie op de kwaliteit van haar onderwijs wordt gerealiseerd, en regelmatig de kwaliteit van haar opleidingen bij studenten, medewerkers, alumni en vertegenwoordigers van het beroepenveld meet en evalueert' (NVAO, 2011, p. 7). Er moet met andere woorden sprake zijn van een 'kwaliteitscultuur', waarin studenten een prominente rol krijgen toebedeeld. Een instelling die over een uitgebreid systeem van onderwijsevaluaties door studenten beschikt, kan bijgevolg aan een auditcommissie laten zien dat ze 'in control' is wat de interne kwaliteitszorg betreft.

In deze bijdrage wordt, na een korte problematisering, een beknopte stand van zaken geschetst van het recente, (hoofdzakelijk) internationale onderzoek naar de validiteit van studentenevaluaties. Op basis van enkele belangrijke vaststellingen worden vervolgens suggesties geformuleerd, enerzijds voor verder onderzoek en anderzijds voor de praktijk.

Het delicate karakter van studentenevaluaties

Onderwijsevaluaties door studenten gebeuren zowel vanuit een verbeter- als vanuit een beoordelingsperspectief. Ze leveren enerzijds feedback ter verbetering, en anderzijds beslissingsgrond voor beleids- en personeelsdoelinden. Het is precies deze spanning die studentenevaluaties zo delicaat maakt. Docenten zijn weliswaar zeer geïnteresseerd in de ervaringen van de studenten (Balam & Shannon, 2010), maar houden ook steeds meer rekening met dat andere doel van de evaluaties. Velen van hen beschouwen het systematisch verzamelen en gebruiken van studentenopinions bovendien als symptomatisch voor het 'nieuwe' hoger onderwijs, dat verantwoordingsplicht, transparantie en (vooral) klanttevredenheid hoog in het vaandel draagt (Blackmore, 2009; Molesworth, Nixon & Scullion, 2009; Titus, 2008) en zich onderwerpt aan een 'permanent kwaliteitstribunaal' (Masschelein & Simons, 2003, p. 48). Bovendien lijkt het soms alsof de opinie van studenten over de kwaliteit van onderwijs wordt gelijkgesteld aan kennis over de kwaliteit van dat onderwijs, temeer omdat aan studentenevaluaties een objectief karakter wordt toegekend: ze zijn anoniem, zijn gebaseerd op grote groepen studenten, en ze worden veelal kwantitatief verwerkt en gepresenteerd.

Het gevolg van deze ontwikkeling is dat heel wat docenten zich vragen stellen bij de betrouwbaarheid en de geldigheid van studentenevaluaties. Deze vragen gaan er bijvoorbeeld over of studenten wel in staat zijn om het onderwijs te evalueren

en of ze geen andere visie op goed onderwijs hebben dan de docent zelf. Ook stellen docenten zich de vraag wiens mening belangrijker is, of studentenevaluaties niet veeleer populariteitspolls zijn, of studenten niet beïnvloed worden door allerlei factoren die niets te maken hebben met de kwaliteit van het onderwijs en of aan de studenten geen andere vragenlijsten kunnen worden voorgelegd. De vraag of het niet beter is om de studenten op een meer persoonlijke manier te betrekken, komt eveneens aan de oppervlakte.

Dit leidt ertoe dat veel docenten wat ongemakkelijk uitkijken naar hun eerstvolgende evaluatierapport. Enkelen van hen vrezen de tirannie van het evaluatieformulier en gaan actief op zoek naar mogelijkheden om de resultaten van hun onderwijsbeoordelingen op te krikken zonder daarvoor iets aan hun onderwijs te doen (Simpson & Siguaw, 2000): ze geven hoge examenscores aan de studenten, verwachten weinig studiewerk of verlenen gemakkelijk vrijstellingen. Een andere strategie is het koudweg negeren van de resultaten van studentenevaluaties. Door deze praktijken worden heel wat interessante en waardevolle opmerkingen en suggesties van de studenten niet benut.

Daarenboven is het gros van de docenten niet op de hoogte van de talloze studies die de voorbije honderd jaar werden verricht naar de betrouwbaarheid en validiteit van studentenevaluaties. Heel wat van hun verzuchtingen vormden immers reeds het voorwerp van onderzoek. Uit dit onderzoek blijkt enerzijds dat het verzamelen en interpreteren van gegevens die voortkomen uit studentenevaluaties inderdaad niet zo eenvoudig is als dat op het eerste gezicht lijkt. Anderzijds wordt uit de literatuur ook duidelijk dat heel wat *urban myths* kunnen worden genuanceerd.

Uit het voorgaande is duidelijk geworden dat belangrijke en terechte vragen worden gesteld over het gebruik en de validiteit van studentenoordelen in het kader van onderwijsbeoordelingen. Deze vragen worden niet het minst door de docenten zelf gesteld. Dit is zeker het geval wanneer de oordelen van de studenten een invloed kunnen uitoefenen op de loopbaan van de betrokkenen. We nemen deze vaststelling als uitgangspunt voor een bespreking van recente onderzoeksresultaten die betrekking hebben op het gebruik en de validiteit van studentenevaluaties.

Het recente onderzoek naar de validiteit van studentenevaluaties

De onderzoeksliteratuur over de validiteit van studentenopinions in relatie tot het gebruikte instrumentarium is sterk Angelsaksisch gekleurd, omdat de Angelsaksische onderwijswereld een lange traditie kent van het systematisch uitvoeren van studentenevaluaties. Europese studies zijn schaars, omdat onderwijsbeoordelingen door studenten daar een recenter fenomeen zijn (Spiel, Schober & Reimann, 2006). In Vlaanderen, bijvoorbeeld, wordt deze vorm van interne kwaliteitszorg pas sinds de jaren negentig op grote schaal toegepast. Vooral de invoering van de onderwijsbeoordelingen door de NVAO en de decretale evaluaties van het onderwijzend personeel hebben ertoe bijgedragen dat zowat elke instelling vandaag haar

studenten bevreemdt over de opleidingsonderdelen die zij hebben gevolgd, en dit zowel vanuit verbeter- als beoordelingsperspectief.

Het recente onderzoek naar de validiteit van studentenevaluaties kan worden opgesplitst in vier grote thema's: studies over de psychometrische waarde van het gebruikte instrumentarium; de zogenaamde 'bias'-studies; onderzoek naar de wijze waarop studentenevaluaties worden verzameld; en studies over de relatie tussen studentenevaluaties en de verbetering van het onderwijs. Algemeen kan worden gesteld dat, ook na honderd jaar onderzoek, het thema studentenevaluaties nog steeds leidt tot verhitte discussies. Bovendien blijven heel wat stakeholders (onderzoekers, docenten, studenten) zich vragen stellen bij de praktijk en/of de zinvolheid van deze studentenevaluaties. Onderzoekers slagen er immers niet in om duidelijke en eenduidige antwoorden te formuleren op verschillende van deze vragen. Kulik (2001) noteert dat 'one might suppose that the research studies on [student] ratings [of teaching] are similar to many other studies in education: conflicting, confusing and inconclusive' (p. 10).

Het instrumentarium

Een eerste belangrijke vaststelling is dat nogal wat instrumenten voor studentenevaluaties niet gelinkt zijn aan de onderwijskundige theorie en/of nauwelijks werden getest op hun bruikbaarheid (Spooren, Mortelmans & Denekens, 2007). Nochtans zijn er verschillende instrumenten beschikbaar die uitvoerig werden gevalideerd (zie onder andere Ginns, Prosser & Barrie, 2007; Kember & Leung, 2008). Deze instrumenten hebben, naast het feit dat zij op theorie en empirie zijn gebaseerd, twee belangrijke voordelen ten opzichte van vele andere bestaande instrumenten voor onderwijsevaluatie door studenten. Ten eerste wordt in deze instrumenten gebruik gemaakt van *schaaltechnieken*, waarbij elk aspect van onderwijs wordt gemeten aan de hand van meerdere items. Schaaltechnieken zijn, in vergelijking met de veelvoorkomende *single-item*-benadering, veel beter geschikt voor het in kaart brengen van percepties over didactische vaardigheden, die moeten worden beschouwd als latente constructen en dus niet onmiddellijk meetbaar zijn (McIver & Carmines, 1981). Schalen zijn daarenboven veel minder gevoelig voor sociale wenselijkheid, ambigue interpretaties en toevallige fluctuaties in de antwoorden van respondenten. Bovendien kan de interne consistentie van een schaal bij elke afname worden gemeten aan de hand van bijvoorbeeld een Cronbachs alpha-statistiek. De score voor elke dimensie in elk evaluatierapport wordt dan vergezeld van een indicator die iets vertelt over de betrouwbaarheid van die score. Scores die niet betrouwbaar zijn, kunnen uit het evaluatierapport worden geweerd en worden niet geïnterpreteerd.

Ten tweede worden aan de hand van deze instrumenten verscheidene onderwijsdimensies in kaart gebracht. In de literatuur bestaat consensus over de *multidimensionaliteit* van onderwijs, namelijk dat onderwijs bestaat uit verschillende aspecten, zoals het studiemateriaal, de studiebegeleiding, de opbouw van de leerinhouden, enzovoort. Aanbevolen wordt om in het kader van onderwijsevaluatie aandacht te besteden aan zoveel mogelijk dimensies om gedetailleerde feedback met betrekking tot specifieke aspecten van een opleidingsonderdeel mogelijk te

maken (Marsh, 1991). Tegelijkertijd wordt echter ook gepleit voor een unidimensionele benadering van de evaluatiescores, die leidt tot één globale score en dus snel zicht biedt op 'de' onderwijskwaliteit binnen een opleidingsonderdeel (Abrami & d'Apollonia, 1991). Deze werkwijze biedt immers voordelen voor het summatief gebruik van evaluatieresultaten (McKone, 1999). Apodaca en Grad (2005) toonden aan dat het ene het andere niet in de weg hoeft te staan, aangezien de multidimensionele structuur van hun data compatibel bleek met een sterke onderliggende factor, die als indicator optreedt voor een *general instructional skill*. Confirmatorische factoranalyse op het aan de Universiteit Antwerpen ontwikkelde SET37-instrument (Mortelmans & Spooen, 2009) toont aan dat ook in dit instrument een onderliggende factor kan worden geïdentificeerd, die veel variantie in zeven schalen uit het instrument verklaart. Dit betekent dat het mogelijk is om zowel gedetailleerde feedback (per gemeten dimensie van onderwijs) als een meer algemene evaluatiescore uit een instrument af te leiden. Andere auteurs (Shevlin, Banyard, Davies & Griffiths, 2000) spreken op hun beurt van een 'halo'-effect, waarbij één kenmerk van een docent, dat niet altijd rechtstreeks betrekking heeft op diens didactische vaardigheden, het evaluatiegedrag van een student beïnvloedt.

Heel wat recent onderzoek heeft overigens betrekking op technische aspecten bij de constructie van het instrumentarium. Achterliggende theoretische en methodologische overwegingen en keuzes kunnen immers een invloed uitoefenen op de evaluatiescores die een instrument, dat doorgaans vooral gesloten vragen bevat, oplevert. Zo tonen Onwuegbuzie en Weems (2004) aan dat het gebruik van een neutrale antwoordcategorie de interne consistentie van een schaalscore kan verzwakken. Verschillende auteurs stellen bovendien vast dat respondenten erg worden beïnvloed door het aantal antwoordcategorieën aan de hand waarvan een gesloten vraag kan worden beantwoord (Sedlmeier, 2006). Studenten maken bijvoorbeeld veel beter gebruik van een vijfpunten-schaal dan van een tienpunten-schaal bij de evaluatie van een opleidingsonderdeel (Landrum & Braitman, 2008). Een ander veelvoorkomend fenomeen in survey-onderzoek is dat van 'janknikkerij' (*acquiescence*), waarbij respondenten bevestigend antwoorden op een vraag, ongeacht of die vraag positief of negatief werd geformuleerd. Dit kan te wijten zijn aan motivationele (sociale wenselijkheid) of cognitieve factoren (een vraag niet goed begrijpen, of nog weinig moeite willen doen om erover na te denken omdat het de zoveelste bevraging betreft) (Knowles & Gordon, 1999). In het onderzoek naar studentenevaluaties zijn de vaststellingen over het voorkomen van *acquiescence* verdeeld. Richardson (2012) stelde in zijn onderzoek aan een Britse instelling voor afstandsonderwijs janknikkerij vast, terwijl de studenten aan de Antwerpse universiteit minder gevoelig blijken voor dit fenomeen (Spooen, Mortelmans & Thijssen, 2012).

Een laatste vaststelling met betrekking tot het gebruikte instrumentarium is dat studenten, docenten en ontwerpers van instrumenten niet altijd dezelfde mening delen over 'goed' onderwijs (Onwuegbuzie et al., 2007) en de betekenis van (op het eerste gezicht) gangbare concepten inzake goed onderwijs (Billings-Gagliardi, Barrett & Mazor, 2004). Een instrument dat niet bij de verschillende stakeholders

werd afgetoetst, kan dus verwarrende en moeilijk te interpreteren resultaten opleveren.

Bias-studies

In de literatuur wordt veel aandacht besteed aan de zogenaamde *bias*-studies, waarin wordt gepeild in hoeverre bepaalde factoren, die met de kwaliteit van onderwijs weinig tot niets te maken hebben, een invloed uitoefenen op de door de studenten uitgebrachte evaluaties (Centra & Gaubatz, 2000). Voorbeelden zijn het geslacht of de seksuele geaardheid van de docent, of het tijdstip waarop het college plaatsvindt. Een populaire variabele die ook vaak in *bias*-studies opduikt, is de examenscore die een student kreeg of verwacht voor het opleidingsonderdeel dat hij of zij moet evalueren. De positieve correlatie tussen examenscore en de door de student uitgebrachte onderwijsevaluatie is statistisch significant maar zwak, want ze ligt ongeveer tussen .10 en .30 (Feldman, 1997). Hoewel voor dit verschijnsel verschillende mogelijke verklaringen zijn (Brockx, Spooren & Mortelmans, 2011), wordt nogal eens gezegd dat een docent goede evaluaties kan 'kopen' door hoge examencijfers te geven. In dit opzicht moet de examenscore van een student worden beschouwd als een *bias*-variabele, daar ze weinig te maken heeft met de kwaliteit van het onderwijs binnen een opleidingsonderdeel. Recent onderzoek toont echter aan dat de verklaringskracht van dit soort *bias*-variabelen eerder laag is (Beran & Violato, 2005; Spooren, 2010) en dat studentenoordelen vooral afhangen van de onderwijsprestaties van docenten (Greimel-Fuhrmann & Geyer, 2003).

Dataverzameling

Studenten blijken over het algemeen bereid zijn om deel te nemen aan evaluaties, al denken ze veelal dat zowel de instelling als de betrokken docenten maar weinig gebruik maken van de resultaten (Spencer & Schmelkin, 2002). Docenten beschouwen studentenevaluaties als een vrij betrouwbare indicator voor zowel het beoordelen van individuele onderwijsprestaties als het aantonen van de onderwijskwaliteit van een instelling (Beran & Rokosh, 2009). Toch maken zij zelf weinig gebruik van de resultaten van studentenevaluaties voor de verbetering van hun onderwijs (Nasser & Fresko, 2002). Docenten blijken bovendien moeilijker om te kunnen gaan met deze resultaten dan men zou verwachten (Kogan, Schoenfeld-Tacher & Helleyer, 2010). Het is daarom van groot belang dat omzichtig wordt omgesprongen met studentenevaluaties en dat docenten kunnen rekenen op *peers* en onderwijsexperts voor de interpretatie en het gebruik van hun persoonlijke resultaten. Beleidsmedewerkers verantwoordelijk voor het verzamelen en interpreteren van studentenevaluaties moeten over de nodige (onder andere statistische) competenties beschikken én voldoende thuis zijn in de thematiek. Pas dan kunnen zij een afdoend antwoord bieden op de bekommernissen van individuele docenten.

Tegenwoordig worden vooral elektronische enquêtes gebruikt voor de afname van studentenevaluaties (Nulty, 2008). Onderzoek toont aan dat de resultaten van deze evaluaties niet erg verschillen van de resultaten afkomstig van de meer klas-

sieke *paper-and-pencil*-methoden (Leung & Kember, 2005) en dat studenten bovendien meer tijd nemen om de evaluatie grondig in te vullen wanneer deze wordt afgenomen in een elektronische omgeving (Venette, Sellnow & McIntyre, 2010). Een nadeel van deze methode blijkt evenwel de lagere responsgraad in vergelijking met klassikale afname van SET (Dommeyer, Baum, Hanna & Chapman, 2004).

Studentenevaluaties en de verbetering van onderwijs

Verschillende studies tonen aan dat de bijdrage van studentenevaluaties aan de verbetering van het onderwijs niet bijster groot is (Marsh, 2007; Lang & Kersting, 2007). Studentenevaluaties *an sich* leiden niet tot beter onderwijs. Dit betekent ten eerste dat studentenevaluaties moeten worden geïntegreerd in een meer holistische benadering van onderwijsevaluatie, waarin docenten worden uitgedaagd om actief na te denken over de verbetering van hun onderwijs. Ten tweede impliceert dit dat studentenevaluaties niet de enige manier kunnen en mogen zijn om de (verbetering van) onderwijsprestaties binnen een bepaald opleidingsonderdeel in kaart te brengen.

Beperkingen en suggesties voor verder onderzoek

Het onderzoek naar de geldigheid van studentenevaluaties kent twee belangrijke beperkingen. Ten eerste is het bestaande instrumentarium gebaseerd op een aantal principes betreffende goed onderwijs uit de literatuur en/of op de onderwijsvisie van een instelling. Echter, tot op heden beschikt men niet over een conceptueel kader voor 'goed onderwijs', dat wordt gedragen door alle stakeholders (studenten, docenten, beleidsmakers, onderwijskundigen) en waarop instrumenten voor onderwijsevaluatie kunnen worden geënt (Penny, 2003). Dit heeft cruciale gevolgen voor de inhoudsvaliditeit en de constructvaliditeit van het instrumentarium (Onwuegbuzie, Daniel & Collins, 2009).

Ten tweede moeten de resultaten van heel wat studies (met inbegrip van de studies die worden vermeld in deze bijdrage) voorzichtig worden geïnterpreteerd, omdat zij afhangen van contextgebonden factoren. Als studenten in een instelling bijvoorbeeld niet worden verplicht om deel te nemen aan een onderwijsevaluatie en vrij kunnen beslissen of hun achtergrondgegevens mogen worden gebruikt voor verdere analyse, treedt er onvermijdelijk een selectie-effect op. Dit beperkt mogelijk de generaliseerbaarheid van de resultaten en is problematisch in heel wat survey-onderzoek. Daarom moet worden nagegaan welke studenten wel/niet deelnemen aan onderwijsevaluaties en in hoeverre deze studenten een al dan niet een verschillend profiel hebben (Spooen & Van Loon, 2012). Met deze gegevens kan dan rekening worden gehouden bij de interpretatie van de resultaten van een onderwijsevaluatie. Ook is de kans op *response sets* ('jankikkerij', extreem antwoordgedrag, willekeurig antwoorden) en dubieuze resultaten veel groter wanneer studenten worden gedwongen om deel te nemen aan evaluaties.

Suggesties voor de praktijk

Uit het voorgaande is duidelijk geworden dat studentenoordelen een belangrijke rol kunnen spelen in de optimalisering van de onderwijspraktijk, maar dat ze met de grootste zorg moeten worden verzameld en geïnterpreteerd. Enkele suggesties:

- Studentenevaluaties worden nog te vaak gebaseerd op *ad-hoc-instrumenten* die niet of nauwelijks op hun waarde werden getest. Dit kan ernstige gevolgen hebben voor zowel de verbetering van de onderwijspraktijk als de loopbaan van de betrokken docenten. Het komt er dus op aan om werk te maken van degelijke instrumenten, die worden gebaseerd op de theorie én de empirie.
- Het *verzamelen en interpreteren* van studentenevaluaties dient met de grootste zorg te gebeuren. Op elk moment in de procedure kunnen fouten worden gemaakt die leiden tot een vertekening of een verkeerde interpretatie van de resultaten. Daarom is het aangewezen om stelselmatig te onderzoeken of er bijvoorbeeld geen antwoordpatronen of andere storende factoren opduiken die nefast zijn voor de kwaliteit van de resultaten. Beleidsmedewerkers die evaluaties organiseren en interpreteren, spelen een belangrijke rol in het onderwijsevaluatiegebeuren. Het is van groot belang dat zij goed bekend zijn met onderwijstheorie en het onderwijsevaluatieonderzoek. Zij moeten beschikken over de nodige vaardigheden om het hele proces in goede banen te leiden. De kwaliteit van de resultaten en de interpretatie ervan ligt in hun handen.
- Wat betreft het *gebruik* van studentenevaluaties lijken heel wat onderwijsinstellingen (bewust of onbewust) te vergeten dat de opinie van studenten slechts één indicator is van de kwaliteit van het onderwijs. Studentenevaluaties zijn erg populair. Hier zijn enerzijds economische redenen voor, met name de mogelijkheid om de vragenlijsten af te nemen van een grote groep stakeholders en ze vervolgens snel te verwerken. Anderzijds hebben studentenevaluaties ook de schijn objectief te zijn. Ze passen bovendien perfect in het hedendaagse denken over hogeronderwijsinstellingen als spelers op de onderwijsmarkt die rekening willen houden met de noden en verzuchtingen van hun 'klanten'. Maar te weinig aandacht wordt besteed aan de mening van collega-docenten, onderwijsexperts en de docent in kwestie zelf. Het is belangrijk dat ook zij, op basis van hun specifieke inhoudelijke of onderwijsexpertise, de kans krijgen en nemen om hun spreekwoordelijke duit in het zakje te doen. Dit kan de kwaliteit van een onderwijsevaluatie alleen maar verbeteren. Verbetering van onderwijs moet immers niet alleen worden gebaseerd op de macht van het getal en op korte en anonieme antwoorden op gestandaardiseerde instrumenten. Verbetering van onderwijs gebeurt het beste in een cultuur van betrokkenheid en samen-denken, en dit zowel op formele als informele wijze.

Wanneer, ten slotte, studentenevaluaties worden gebruikt voor zowel summatieve als formatieve doeleinden, is het codewoord 'vertrouwen'. Docenten moeten goed en precies geïnformeerd worden over het gebruik en het belang van dit soort

evaluaties. Zij moeten kunnen rekenen op collega's en beleidsmedewerkers bij het interpreteren van en het omgaan met deze resultaten. Eén of twee minder goede studentenevaluaties mogen niet beslissend zijn wanneer de onderwijsprestaties van een docent worden beoordeeld. Integendeel, docenten moeten worden aangemoedigd om te experimenteren met cursusinhouden en onderwijsmethoden (in plaats van te handelen naar de vereisten opgelegd door het instrument of hoge examenscores toe te kennen in ruil voor betere evaluatieresultaten). Ongeacht wat de gevolgen daarvan zijn voor het resultaat van een onderwijsevaluatie. Alleen dan kunnen en zullen onderwijsevaluaties door studenten ten volle bijdragen aan de verbetering van onderwijs.

Referenties

- Abrami, P.C. & d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness-generalizability of 'N=1' research. Comment on Marsh (1991). *Journal of Educational Psychology*, 83, 411-415. doi: 10.1037/0022-0663.83.3.411
- Apodaca, P. & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni- and multidimensional models. *Studies in Higher Education*, 30, 723-748. doi:10.1080/03075070500340101
- Balam, E. & Shannon, D. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment and Evaluation in Higher Education*, 35, 209-221. doi:10.1080/02602930902795901
- Beran, T.N. & Rokosh, J.L. (2009). Instructor's perspectives on the utility of student ratings of instruction. *Instructional Science*, 37, 171-184. doi:10.1007/s11251-007-9045-2
- Beran, T. & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education*, 30, 593-601. doi:10.1080/02602930500260688.
- Billings-Gagliardi, S., Barrett, S.V. & Mazor, K.M. (2004). Interpreting course evaluation results: Insights from thinkaloud interviews with medical students. *Medical Education*, 38, 1061-1070. doi:10.1111/j.1365-2929.2004.01953.x
- Blackmore, J. (2009). Academic pedagogies, quality logics and performative universities: Evaluating teaching and what students want. *Studies in Higher Education*, 34, 857-872. doi: 10.1080/03075070902898664
- Brockx, B., Spooen, P. & Mortelmans, D. (2011). Taking the 'grading leniency' story to the edge: The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, 23, 289-306. doi:10.1007/s11092-011-9126-2
- Centra, J.A. & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71, 17-33. Retrieved from www.jstor.org/stable/2649280
- Dommeyer, C. J., Baum, P., Hanna, R. W. & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29, 611-623. doi: 10.1080/02602930410001689171
- Feldman, K.A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. Perry & J. Smart (Eds.), *Effective teaching in higher education: Research and Practice* (pp. 368-395). New York: Agathon.

- Ginns, P., Prosser, M. & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education*, 32, 603-615. doi:10.1080/03075070701573773
- Greimel-Fuhrmann, B. & Geyer, A. (2003). Students' evaluation of teachers and instructional quality: Analysis of relevant factors based on empirical research. *Assessment & Evaluation in Higher Education*, 28, 229-238. doi:10.1080/0260293032000059595
- Kember, D. & Leung, D. (2008). Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education*, 33, 341-353. doi:10.1080/02602930701563070
- Knowles, E.S. & Gordon, C.A. (1999) Why people say 'yes': A dual process theory of acquiescence. *Journal of Personality and Social Psychology*, 77, 379-386. doi:10.1037/0022-3514.77.2.379
- Kogan, L., Schoenfeld-Tacher, R. & Helleyer, P. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, 15, 623-636. doi:10.1080/13562517.2010.491911
- Kulik, J.A. (2001). Student ratings: Validity, utility and controversy. *New Directions for Institutional Research*, 27, 9-25. doi:10.1002/ir.1
- Landrum, R.E. & Braitman, K.A. (2008). The effect of decreasing response options on students' evaluation of instruction. *College Teaching*, 56, 215-217. doi:10.3200/CTCH.56.4.215-218
- Lang, J.W.B. & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, 35, 187-205. doi:10.1007/s11251-006-9006-1
- Larsen, M.A. (2005). A critical analysis of teacher evaluation policy trends. *Australian Journal of Education*, 49, 292-305.
- Leung, D.Y.P. & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education*, 46, 571-591. doi:10.1007/s11162-005-3365-3
- Masschelein, J. & Simons, M. (2003). *Globale immuniteit: een kleine cartografie van de Europese ruimte voor onderwijs*. Leuven: Acco.
- Marsh, H.W. (1991). A multidimensional perspective on students' evaluations of teaching effectiveness: Reply to Abrami & d'Apollonia (1991). *Journal of Educational Psychology*, 83, 416-421. doi:10.1037/0022-0663.83.3.416
- Marsh, H.W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluation of teaching over 13 years. *Journal of Educational Psychology*, 99, 775-790. doi:10.1037/0022-0663.99.4.775
- McIver, J.P. & Carmines, E.G. (1981). Unidimensional scaling. *Sage University Paper series on Quantitative Applications in the Social Sciences*, 24-001. Sage: Newbury Park, CA.
- McKone, K.E. (1999). Analysis of student feedback improves instructor effectiveness. *Journal of Management Education*, 23, 396-415. doi:10.1177/105256299902300406
- Molesworth, M., Nixon, E. & Scullion, R. (2009). Having, being and higher education: The marketisation of the university and the transformation of the student into consumer. *Teaching in Higher Education*, 14, 277-287. doi:10.1080/13562510902898841
- Mortelmans, D. & Spooren, P. (2009). A revalidation of the SET37-questionnaire for student evaluations of teaching. *Educational Studies*, 35, 547-552. doi:10.1080/03055690902880299
- Nasser, F. & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27, 187-198. doi:10.1080/02602930220128751

- NVAO (2011). Beoordelingskaders accreditatiestelsel hoger onderwijs Instellingstoets kwaliteit zorg. Op 1 februari 2013 ontleend aan http://www.nvao.net/page/downloads/DEFINITIEVE_KADERS_INSTELLINGSTOETS_22_november_2011.pdf
- Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, 33, 301-314. doi: 10.1080/02602930701293231
- Onwuegbuzie, A. J. & Weems, G.H. (2004). Response categories on rating scales: Characteristics of item respondents who frequently utilize midpoint. *Research in the Schools*, 9, 73-90.
- Onwuegbuzie, A.J., Witcher, A.E., Collins, K.M.T., Filer, J.D., Wiedmaier, C.D. & Moore, C.W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal*, 44, 113-160. doi:10.3102/0002831206298169
- Onwuegbuzie, A.J., Daniel, L.G. & Collins, K.M.T. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity*, 43, 197-209. doi:10.1007/s11135-007-9112-4
- Penny, A.R. (2003). Changing the agenda for research into student's views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8, 399-411. doi: 10.1080/13562510309396
- Richardson, J.T.E. (2011). The role of response biases in the relationship between students' perceptions of their courses and their approaches to studying in higher education. *British Educational Research Journal*. Advance online publication.
- Shevlin, M., Barnyard, P., Davies, M.N.O. & Griffiths, M.D. (2000). The validity of student evaluations in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25, 397-405. doi:10.1080/713611436
- Simpson, P. & Siguaw, J. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22, 199-213. doi: 10.1177/0273475300223004
- Spencer, K.J. & Schmelkin, L.P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, 27, 397-409. doi: 10.1080/026029302200009285
- Spiel, C., Schober, B. & Reimann, R. (2006). Evaluation of curricula in higher education: Challenges for evaluators. *Evaluation Review*, 30, 430-450.
- Spooen, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on student evaluations of teaching. *Studies in Educational Evaluation*, 36, 121-131. doi: 10.1016/j.stueduc.2011.02.001
- Spooen, P., Mortelmans, D. & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert scales. *Assessment and Evaluation in Higher Education*, 32, 667-679. doi: 10.1080/02602930601117191
- Spooen, P., Mortelmans, D. & Thijssen, P. (2012). Content vs. style: Acquiescence in student evaluations of teaching? *British Educational Research Journal*, 38, 3-21. doi: 10.1080/01411926.2010.523453
- Spooen, P. & Van Loon, F. (2012). Who participates (not)? A non-response analysis on students' evaluations of teaching. *Procedia of Social and Behavioral Sciences*, 69, 990-996. doi: 10.1016/j.sbspro.2012.12.025
- Titus, J. (2008). Student ratings in a consumerist academy: Leveraging pedagogical control and authority. *Sociological Perspectives*, 51, 397-422. doi:10.1525/sop.2008.51.2.397

Venette, S., Sellnow, D. & McIntyre, K. (2010). Charting new territory : Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, 35, 101-115. doi: 10.1080/02602930802618336