



Notitie Vakevaluaties



Casper Albers
de Personeelsfractie
September 2016

Bij de meeste vakken aan de RUG wordt na afloop via antwoorden van studenten op vragenlijsten een vakevaluatie opgesteld. Uit meerdere wetenschappelijke onderzoeken is gebleken dat deze manier van evalueren vaak niet meet wat het dient te meten. In deze memo legt De Personeelsfractie kort uit wat het probleem is, om vervolgens het College van Bestuur een aantal suggesties te bieden om in de toekomst vakken beter te kunnen evalueren¹.

Inleiding

Bij verreweg de meeste vakken aan de RUG wordt na afloop een vakevaluatie gehouden. Hierbij geven de studenten die het vak gevolgd hebben aan hoe zij het vak ervaren hebben. De Personeelsfractie erkent het belang van de meningen van studenten over (de kwaliteit van) een vak: deze notitie is geenszins een pleidooi tot afschaffing van de evaluaties.

De precieze afname van deze vakevaluaties verschilt per opleiding (bv. evaluatie op paper direct na afloop van het tentamen, of online op een later moment), maar de essentie is overal hetzelfde: via een aantal vragen ('de colleges sloten goed aan bij de literatuur', 'het tentamen was een representatieve afspiegeling van de stof', etc.) kunnen de studenten (anoniem) aangeven welke aspecten van het vak goed danwel slecht ervaren werden. Deze evaluaties worden door docenten, opleidingscommissies en anderen gebruikt om de studieprogramma's verder te verbeteren.

Tevens worden deze vakevaluaties in sommige situaties bij de RUG gebruikt bij de beoordeling van de betrokken docent. Zo is bijvoorbeeld "*Uitvoering van het onderwijs is goed blijkend uit vakevaluaties*" een van de eisen bij Faculteit GMW voor Tenure Track-medewerkers. Het niet voldoen aan Tenure Track-eisen kan een reden zijn om het arbeidscontract niet voort te zetten. Vakevaluaties kunnen bij de RUG dus letterlijk carrières breken.

Samengevat, vakevaluaties hebben direct invloed op de kwaliteit van het onderwijs alsmede op wie er bij de universiteit werkt. Ze raken dus de kern van de universiteit. Als men vakevaluaties wil gebruiken om de kwaliteit van het onderwijs en de docent te meten, dan moeten deze evaluaties daar wel een geschikt middel voor zijn.

Probleem

En dat is het probleem: zij is geen geschikt middel. Een goede docent is iemand die goede leeruitkomsten bij zijn/haar studenten realiseert. Het realiseren van leeruitkomsten is *het* doel van de opleiding en daarmee belangrijker dan satisfactie van de studenten – terwijl vakevaluaties vaak alleen dit laatste meten.

De afgelopen decennia is er veelvuldig onderzoek gedaan naar de effectiviteit en kwaliteit van vakevaluaties. Dit onderzoek is recentelijk (2016) samengevat in meerdere reviewartikels^{2, 3}, waarin ook enkele grootschalige en invloedrijke studies zijn opgenomen.

Uttl, White & Gonzalez³ hebben gekeken naar meta-analyses rond vakevaluaties en concluderen dat een aantal oude studies incorrect zijn uitgevoerd ("*these meta-analyses* suffer from multiple critical methodological flaws that render their conclusions unwarranted"). Door de meta-analyse correct uit te voeren, berekenen zij de correlatie tussen de evaluatie van een vak en de mate waarin de studenten hun leerdoelen bij dat vak behaald hebben: *maximaal 0.01*. Dit is praktisch nul: het verschil in vakevaluatie tussen docenten die wel en docenten die niet voor elkaar krijgen dat hun studenten de gewenste leerdoelen bereiken is verwaarloosbaar klein.

¹ De Personeelsfractie bedankt dr. Ellen Jansen (Faculteit GMW, afdeling Lerarenopleiding) voor haar nuttige adviezen bij het opstellen van deze notitie.

² Kornell, N. & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in Psychology – Educational Psychology*, doi:10.3389/fpsyg.2016.00570

³ Uttl, B., White, C.A. & Gonzalez, D.W. (2016). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, doi: 10.1016/j.stueduc.2016.08.007

Uit hun onderzoek concluderen Kornell & Hausman² niet alleen dat vakevaluaties geen goede maat voor docentkwaliteit zijn, maar zelfs dat de *beste docenten* gemiddeld genomen de *laagste waardering* krijgen bij evaluaties. Braga et al.⁴ verklaren dat paradoxaal resultaat als volgt:

“[...] that good teachers are those who require their students to exert effort; students dislike it, especially the least able ones, and their evaluations reflect the utility they enjoyed from the course. Overall, our results cast serious doubts on the validity of students’ evaluations of professors as measures of teaching quality or effort.”

Vakevaluaties zijn een degelijk middel om studenttevredenheid te meten, maar niet om onderwijskwaliteit te meten. Los van het feit dat vakevaluaties niet, en mogelijk soms zelfs het omgekeerde, de kwaliteit van het onderwijs meten, zijn er meer problemen. Zo krijgen mannen stelselmatig hogere waarderingcijfers dan vrouwen⁵ (ook na correctie voor aantal jaren onderrichtserving, etc.) en worden vakevaluaties positiever ingevuld als de zon schijnt⁶ tijdens het evalueren dan wanneer het regent.

Kortom: (i) vakevaluaties meten niet of beperkt wat ze moeten meten, (ii) vakevaluaties discrimineren, (iii) vakevaluaties worden beïnvloed door factoren die niets met de kwaliteit van het vak en de docent te maken zijn. Er zijn nog veel meer punten van kritiek – de geïnteresseerde lezer kan terecht bij het artikel van Kornell & Hausman² en de daar gebruikte referenties.

Een kanttekening die hierbij wel gemaakt moet worden is het volgende. De in de wetenschappelijke literatuur beschreven problemen zijn grotendeels gebaseerd op onderzoek aan Amerikaanse onderwijsinstellingen. Het is niet bekend deze problemen zich op dezelfde schaal voordoen in Groningen. Het is belangrijk dat hier meer over bekend wordt.

Los van bovenstaande is er nog een flink aantal praktische problemen met de huidige manier van vakken evalueren. De focus van deze notitie ligt op de genoemde fundamentele problemen rond de huidige manier van vakevaluaties afnemen. Daarom zal in deze notitie geen uitgebreide aandacht besteed worden aan de praktische problemen.

Kern van ons probleem met evaluaties is dat de resultaten uit (vak)evaluaties veelal een zwaarder gewicht krijgen dan ze op basis van de validiteit van het middel verdienen. Om met Kornell & Hausman² te spreken:

“Despite their limitations, we do not suggest abandoning student ratings, but do recommend that student evaluation scores should not be the sole basis for evaluating college teaching and they should be recognized for what they are.”

Suggesties

Op basis van het hierboven geschetste probleem, komt de Personeelsfractie met de volgende twee suggesties:

1 Het College van Bestuur heeft recentelijk aangegeven het Tenure-Trackbeleid centraal aan te willen sturen. Met betrekking tot de rol van vakevaluaties in het personeelsbeleid is ons advies helder. Stop met *onmiddellijke ingang en bij alle faculteiten* met het hanteren van vakevaluaties als criterium in het Tenure-Track- en ander personeelsbeleid. Docentkwaliteit is op andere manieren beter te meten. Denk hierbij aan inhoudelijke beoordeling van de cursusopzet, college-observaties van peers, etc. Het is voor de Personeelsfractie onacceptabel dat medewerkers (mede) op basis van een bewezen ineffectief instrument hun baan kunnen verliezen.

⁴ Braga, M., Paccagnella, M., and Pellizzari, M. (2014). Evaluating students’ evaluations of professors. *Economics of Education Review* 41, 71–88. doi: 10.1016/j.econedurev.2014.04.002

⁵ MacNell, L., Driscoll, A., and Hunt, A. N. (2014). What’s in a name: exposing gender bias in student ratings of teaching. *Innovations in Higher Education* 40, 1–13. doi: 10.1007/s10755-014-9313-4

⁶ Carrell, S. E., and West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy* 118, 409–432. doi: 10.1086/653808

2 Vakevaluaties gaan over onderwijs aan individuele opleidingen. De Personeelsfractie erkent dat dit in eerste instantie een zaak van de opleiding zelf (docent, studenten, opleidingscommissie, examencommissie, onderwijsdirecteur) is en dat het College van Bestuur hier geen directe rol in heeft. Het College kan echter wel een adviserende rol spelen. De Personeelsfractie pleit niet voor afschaffing van de vakevaluaties – net zoals bijvoorbeeld de Nationale Studenten Enquête (NSE) en universitaire rankings hebben deze een functie om problemen te signaleren, alsmede om studenttevredenheid te meten. De Personeelsfractie pleit echter wel voor (i) bewustwording van de beperkte ‘macht’ van vakevaluaties en (ii) verbetering van de manier van evalueren.

Derhalve adviseren wij het College om een onderzoek te starten naar de manieren binnen de RUG waarop momenteel vakevaluaties plaatsvinden en plaats zouden kunnen vinden⁷. Naast de klassieke schriftelijke vragenlijst (al dan niet online), gebeurt dit nu ook al via bijvoorbeeld klassieke evaluaties (biologie) en via een jaarvertegenwoordiging (psychologie). In het onderzoeksrapport dient te worden opgesomd wat de verschillende mogelijkheden zijn, en wat per mogelijkheid de bijbehorende voor- en nadelen zijn. Deze opsomming dient *evidence-based* te zijn, dus onderbouwd met wetenschappelijke referenties. Tevens dient onderzocht te worden of de evaluaties voortaan gekoppeld kunnen worden aan de studieresultaten van individuele studenten, zonder dat hun anonimiteit in het geding komt: alleen met deze koppeling is meetbaar of de evaluaties daadwerkelijk gebruikt kunnen worden als maat voor onderwijskwaliteit.

Het onderzoeksrapport kan vervolgens verspreid worden over de opleidingen zodat zij een geïnformeerde beslissing kunnen nemen over de vorm van evalueren. De Personeelsfractie wil, vanzelfsprekend, te zijner tijd het verslag met het College bespreken.

Namens de Personeelsfractie,
Casper Albers

⁷ De ‘Stuurgroep Blue’ lijkt ons een geschikte gesprekspartner in dit onderzoek.