

# Open commentaar bij gesloten vragenlijsten

Willem van Os

**Het is gebruikelijk dat onderwijs met hulp van een gesloten vragenlijst geëvalueerd wordt onder studenten. Vaak kunnen studenten ook aanvullende opmerkingen maken. Het blijkt dat studenten die commentaar geven niet wezenlijk verschillen van studenten die dit niet doen, en dat het in zijn algemeenheid aanbeveling verdient het commentaar van studenten als representatief te beschouwen en ter harte te nemen.**

## Inleiding

In het Nederlandse hoger onderwijs is het sinds jaar en dag gebruik om het onderwijs te evalueren met behulp van vragenlijsten die bij studenten worden afgenomen. Meestal zijn die lijsten in meerdere of mindere mate gestandaardiseerd met 'gesloten' antwoordmogelijkheden, waarbij daarnaast ruimte is voor studenten om aanvullend commentaar te geven op vak, docent, tentamen of wat dan ook. Aan de Vrije Universiteit Amsterdam is dat in elk geval wel zo, en die opmerkingen fungeren in de praktijk als een nadere toelichting op of verantwoording van de cijfermatige studentoordelen op de gesloten vragen. Soms wordt in dit verband zelfs het onderscheid 'kwalitatieve' versus 'kwantitatieve' evaluatie gebezigd, iets wat hier wegens de veronderstelde tegenstelling overigens niet wordt toegejuicht. Zowel docenten als opleidingscommissies hechten vaak veel waarde aan die opmerkingen. Juist daarom is het van belang

*Studenten die open commentaar geven verschillen niet wezenlijk van studenten die dit niet doen*

hoe representatief dat commentaar eigenlijk is voor alle studenten, en met die vraag wordt dan ook begonnen. Daarna komen punten aan de orde als: waarover worden de meeste opmerkingen gemaakt, verschillen studenten van de diverse faculteiten in hun schrijflust, kun je studenten stimuleren om commentaar te geven (of anderzijds de omvang beperken), komt het vaak voor dat studenten 'over de schreef gaan', en meer in het algemeen: heb je er echt wat aan?

## Representativiteit

Over de betrouwbaarheid, validiteit en bruikbaarheid van evaluatievragenlijsten is buitengewoon veel onderzoek gedaan, vooral in de Verenigde Staten. Overzichten daarvan zijn onder meer te vinden in Marsh (1987) en Marsh en Dunkin (1992), en voor wat betreft Nederland kan bijvoorbeeld worden verwezen naar de dissertaties van Gijselaers (1988) en Van Os (1999). Een onderwerp dat van meet af aan sterk in de belangstelling heeft gestaan is bijvoorbeeld de relatie tussen de beoordeling van docenten enerzijds en het (verwachte) studieresultaat anderzijds (meer in het bijzonder het zogenaamde 'leniency effect'). Zie hiervoor onder anderen Baird, 1987; Cohen, 1981; Greenwald & Gill-

more, 1997a, 1997b; Marsh & Roche, 2000; Palmer, Carliner & Romer, 1978 en Pounder, 2007. Andere aandachtspunten waren en zijn het mogelijke verband tussen die beoordelingen en docentkenmerken als presentatiewijze (de 'dr. Fox studies'; Abrami et al., 1982; Marsh & Ware, 1982), leeftijd, ervaring, sekse (Basow & Silberg, 1987; Bennett, 1982), studentkenmerken als de studiefase, de groeps grootte, cursuskenmerken (hoofdvak, bijvak), eigenlijk alles wat ook maar enigszins van invloed zou kunnen zijn. Zo rapporteert Zabaleta (2007) zelfs een significant verband tussen het tijdstip waarop onderwijs wordt gegeven en de waardering voor de docent (tussen 10 en 12 uur is in zijn studie het beste moment).

In al dit onderzoek gaat het wel steeds om gesloten vragenlijsten en niet om analyses van het bijgevoegde commentaar: hiervan zijn ons geen voorbeelden bekend. Omdat evaluatievragenlijsten bijna altijd anoniem worden ingevuld zijn er ook weinig mogelijkheden om 'schrijvers' (commentaargevers) te vergelijken met 'zwijgers'. Wat resteert is de manier waarop zij de vragen hebben beantwoord: geven schrijvers gemiddeld veel lagere, of misschien juist wel hogere, waarderingen dan zwijgers? In het eerste geval zouden docenten zich van negatief commentaar bijvoorbeeld kunnen afmaken met de opmerking dat het hier gaat om een klein groepje 'dat altijd wat te klagen heeft', maar dat de overgrote meerderheid 'gewoon tevreden is'.

Om deze vraag te beantwoorden zijn enkele jaren geleden bij de Faculteit der Sociale Wetenschappen aan de Vrije Universiteit bij 28 studieonderdelen de formulieren met commentaar apart verwerkt. Daarbij hebben we ons beperkt tot die onderdelen waarbij minstens 10 studenten commentaar hadden gegeven. In de praktijk komt dit er op neer dat de vergelijking alleen plaats heeft gevonden bij studieonderdelen met veel studenten, want doorgaans geeft namelijk slechts een minderheid commentaar. Dit betekent weer dat we het hier vooral hebben over onderwijs in de vorm van (hoor)colleges.

Bij 28 cursussen konden dus afzonderlijke scores worden berekend voor beide groepen, en in de onderstaande tabel wordt het gemiddelde vermeld op de overallvragen *totaaloordeel over de docent*, *over de cursusinhoud* en *over de eindtoets*.

Hoewel studenten die commentaar geven op deze overallvragen iets kritischer zijn dan studenten die er het zwijgen toe doen, gaat het hier om minieme en niet-significante verschillen. Zonder tegenbewijs nemen we daarom aan dat schrijvers in hun commentaar over deze zaken ook de mening van hun collega's vertegenwoordigen.

Tabel 1 Algemene mening van schrijvers en zwijgers

	schrijvers	zwijgers	verschil	p-verschil
ocontent	3,65	3,69	-0,04	0,33
oorsusinhoud	3,51	3,57	-0,06	0,22
oindtoets	3,38	3,40	-0,02	0,54

### Belangrijke kwesties

Op bijna alle afzonderlijke vragen is het oordeel van de studenten die extra commentaar geven wat negatiever, net als dat bij de bovenstaande overall-vragen het geval was. Dat ligt misschien ook wel voor de hand, gegeven het feit dat zij zich de moeite getroosten om nog een extra kanttekening aan hun formulier toe te voegen, maar er zijn een paar uitzonderingen die juist daarom de aandacht verdienen.

Tabel 2 Verschillen tussen schrijvers en zwijgers

	schrijvers	zwijgers	verschil	p-verschil
Studieonderdeel interessant	3,56	3,46	+0,10	0,23
Regelmatige collegevoorbereiding	3,16	3,09	+0,07	0,36
Collegebezoek einde cursus	3,98	3,71	+0,26	0,00
Collegebezoek 67% of meer	72,6	62,9	+9,7	0,00
Geschatte studielast in uren	44,5	41,9	+2,6	0,17

De verschillen zijn natuurlijk niet erg groot en doorgaans evenmin significant, afgezien van het collegebezoek, maar ze wijzen wel consistent in dezelfde richting: schrijvers zijn in het algemeen de wat actievere en meer betrokken studenten, en hoewel dat achteraf misschien wel vanzelf spreekt, is het toch prettig dat dit uit de gegevens daadwerkelijk naar voren komt. Significante verschillen zijn er wel op de vragen in Tabel 3. Kritischer zijn ze ook over het overige studiemateriaal, maar dat verschil was niet significant.

Het ligt in de rede dat hiermee ook de onderwerpen genoemd zijn waarop hun opmerkingen vooral gericht zijn, en voor een deel is dat ook zo. Toch heeft door de bank genomen het meeste commentaar betrekking op onderwijsaspecten die nu juist *niet* door de desbetreffende vragenlijst worden bestreken. Dat geldt bijvoorbeeld voor eventuele opdrachten, en meer in het bijzonder de tijd die daarmee gepaard gaat – dit vooral wanneer bij parallel lopende studieonderdelen óók opdrachten moeten worden gemaakt. Bij een bepaalde cursus leverden 62 studenten, bijna 46% van het totaal, commentaar. Dit is uitzonderlijk veel, maar 56 van die 62 studenten klaagden onder meer over de blijkbaar evident ontoereikende tijd die voor één van de opdrachten beschikbaar was.

Tabel 3 Enkele significante verschillen

	schrijvers	zwijgers	verschil	p-verschil
Adequaat gebruik av-media	3,67	3,74	-0,07	0,04
Kwaliteit van het collegedictaat	3,49	3,58	-0,09	0,03
Gebruikte onderwijsmethode geschikt	3,33	3,49	-0,16	0,01

Hoewel studenten in het algemeen nogal kritisch zijn ten opzichte van de kwaliteit van het tentamen, niet zonder reden trouwens, wordt over het tentamen zelf relatief niet eens zoveel commentaar geleverd. Wellicht komt dit omdat over dat tentamen bij de vragenlijst waar het hier om gaat al zes vragen worden gesteld. Anders ligt dat bij de omstandigheden waaronder het tentamen wordt afgenomen, iets wat niet door de vragenlijst wordt gedekt. Het gaat bijvoorbeeld om hinderlijke (met elkaar pratende of rondlopende) surveillanten, rumoer in het algemeen, ongerieflijk meubilair (te lage stoelen, te hoge tafels), een geldend eetverbod, het afgeven van tassen, enzovoort. Hieruit kan men overigens opmaken dat aan de Vrije Universiteit evaluatievragenlijsten overwegend op het tentamen worden uitgedeeld en ingevuld. Responsproblemen zoals gemeld door Kamphorst en Oostindiër in het vorige nummer van *Onderzoek van Onderwijs* spelen mede daarom niet of nauwelijks een rol: gemiddeld nadert die respons de 70%. Het bovenstaande komt er op neer dat bij de constructie van vragenlijsten als hier bedoeld niet kan, en ook niet hoeft te worden gestreefd naar volledigheid: wanneer studenten echt iets dwars zit, maken ze dat via commentaar heus wel duidelijk. Die gelegenheid moet men ze dan wel bieden. Dat kan eenvoudig door bij bepaalde vragen (bijvoorbeeld over de moeilijkheidsgraad van de studiestof of het tentamen, de tijdbesteding, het studiemateriaal) expliciet te verzoeken om een nadere toelichting op de achterkant van het formulier. Onze ervaring is dat studenten aan zo'n simpele aansporing genoeg hebben om die toelichting ook daadwerkelijk te geven.

*Het commentaar is doorgaans constructief en bruikbaar voor onderwijsverbetering*

### Verschillen tussen faculteiten

Voor deze bijdrage is niet gekwantificeerd of bij verschillende faculteiten ook verschillende hoeveelheden commentaar wordt geleverd. Die verschillen zijn er naar onze stellige indruk wel. Niet voor niets zijn de eerdere gegevens ontleend aan de Faculteit der Sociale Wetenschappen. Deze faculteit wordt kennelijk bevolkt door graag en veel schrijvende studenten: per studieonderdeel bedraagt het aantal schrijvers gemiddeld bijna 19%. Soms leek de docent nadrukkelijk gevraagd te hebben om bij de cursevaluatie met suggesties over zijn onderwijs te komen. Zo'n opwekking mist zijn uitwerking niet: studenten leiden er vermoedelijk uit af dat serieus wordt omgegaan met wat ze te berde brengen, wat daarnaast in zijn algemeenheid ook nog responsverhogend werkt. Studenten uit de Bètafaculteiten en Letteren schrijven minder vaak,

en als ze het doen zijn ze aanzienlijk beknopter. Eerder werd aangegeven dat de opmerkingen vaak betrekking hebben op zaken waarover geen enquêtevragen worden gesteld. Een implicatie daarvan is dat het aantal opmerkingen toeneemt naarmate de vragenlijst korter en globaler is: studenten willen toch hun ei kwijt. Wie denkt tijd te besparen door korte vragenlijsten uit te delen komt in dit opzicht dan ook bedrogen uit. Onzerzijds is wel getracht de omvang van dat commentaar enigszins in te dammen (de verwerking ervan is namelijk nogal arbeidsintensief) door op het formulier van zo'n korte vragenlijst een kader aan te brengen, met de aanwijzing dat geschreven teksten binnen dat kader moesten blijven. Het was opmerkelijk hoe goed studenten zich daar aan hielden, overigens met als neveneffect dat dikwijls niet zozeer het aantal woorden als wel de lettergrootte afnam. Men ging als het ware priegelen, en een tijdsbesparing leverde het uiteindelijk niet zozeer op.

### Conclusie

Uit het voorafgaande kan misschien de indruk ontstaan dat studenten de hun geboden ruimte vooral gebruiken om te klagen. Dat is zeker niet zo, in elk geval lang niet altijd. Niet zozeer de ongemakken wekken hun ergernis op, als wel de *vermijdbare* ongemakken, zoals eerdere tentamens die na zes weken nog niet zijn nagekeken. Opvallend

is verder juist hoe constructief studenten zijn in hun commentaar, en hoe vaak ze de docent nog eens extra lof toezwaaien voor zijn of haar inspanningen. Zelden of nooit worden ze beledigend. Het is natuurlijk niet leuk om te lezen dat je doceercapaciteiten waardeloos zijn, of dat je erg saai "les geeft"; zeker wanneer je dat als docent helemaal niet vindt van jezelf, maar dat is toch echt iets anders dan gescheld. Vermoedelijk speelt hierbij ook de situatie waarin de vragenlijsten worden ingevuld - op de tentamenzitting - een rol.

Natuurlijk is niet elke opmerking van elke student inhoudelijk juist en van even groot belang. Toch is de slotconclusie dat hun commentaar vaak een waardevolle aanvulling is op de meer getalsmatige beoordeling van de kwaliteit van het gevolgde onderwijs, en beslist behulpzaam kan zijn bij datgene waar het in dit verband vooral om gaat: onderwijsverbetering.

### Personalia

Willem van Os is verbonden aan de Taakgroep Onderwijsevaluatie / Kwaliteitszorg van het Onderwijscentrum van de Vrije Universiteit in Amsterdam.

### Literatuur

- Abrami, P. C., Leventhal, L. & Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52, 107-118.
- Baird, J. S. (1987). Perceived learning in relation to student evaluation of university instruction. *Journal of Educational Psychology*, 79, 90-91.
- Basow, S. A. & Silberg, N. T. (1987). Student evaluations of college professors: are female and male instructors rated differently? *Journal of Educational Psychology*, 79, 308-314.
- 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, 170-179.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: a meta-analysis of multi-section validity studies. *Review of Educational Research*, 51, 281-309.
- Greenwald, A. G. & Gillmore, G. M. (1997a). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89, 743-751.
- Greenwald, A. G. & Gillmore, G. M. (1997b). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.
- Gijselaers, W. H. (1988). *Kwaliteit van het onderwijs gemeten*. Academisch proefschrift. Maastricht: Rijksuniversiteit Limburg.
- Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H. W. & Dunkin, M. J. (1992). Students' evaluation of university teaching: a multidimensional perspective. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research - Volume VII*, pp. 1430-233. New York: Agathon Press.
- Marsh, H. W. & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: popular myth, bias, validity or innocent bystanders? *Journal of Educational Psychology*, 92, 202-228.
- Marsh, H. W. & Ware, J. E. (1982). Effects of expressiveness, content coverage and incentive on multidimensional student rating scales. *Journal of Educational Psychology*, 74, 126-134.
- Os, W. van (1999). *Bruikbaarheid en effectiviteit van studentenoordelen over het onderwijs*. Academisch proefschrift. Amsterdam: Vrije Universiteit Amsterdam.
- Palmer, J., Carliner, G. & Romer, T. (1978). Leniency, learning and evaluations. *Journal of Educational Psychology*, 70, 855-863.
- Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? *Quality Assurance in Education*, 15, 178-191.
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12, 55-76.

Het geven van commentaar kan worden gestimuleerd door aan sommige vragen toe te voegen: "licht uw antwoord s.v.p. toe op de achterzijde van het formulier"

