



Gastcolumn

■ Martijn Leenknecht

Dr. M.J.M. Leenknecht is onderwijskundige bij HZ University of Applied Sciences (Hogeschool Zeeland), extern lid examencommissie Academie Paramedische Studies HAN, en oprichter en mede-coördinator van het landelijke Platform Leren van toetsen. Hij schrijft deze column op persoonlijke titel.
E-mail: m.j.m.leenknecht@hz.nl

Zelftest

Van “zelftesten zijn niet betrouwbaar” tot “vanaf nu hoeft je alleen nog een zelftest uit te voeren bij corona-gerelateerde klachten”. Dat is zo’n beetje hoe het ging tijdens de coronapandemie. Bij de introductie van de testkits voor thuisgebruik (antigeentest) was de scepsis groot. Kunnen mensen dat wel zelf? Stoppen ze het stokje wel diep genoeg? De betrouwbaarheid van de tests werd in twijfel getrokken, omdat de test niet onder gecontroleerde omstandigheden wordt afgenomen. Bovendien is de thuistest minder fijnmazig en spoort deze ook bij correct gebruik minder nauwkeurig coronadeeltjes op dan de laboratoriumtests (PCR-test). Die scepsis herken ik vanuit het onderwijs wanneer we het hebben over peer- en self-assessment. Er is alleen één groot verschil. Bij de coronatests gingen we uiteindelijk massaal aan de zelftest, in het onderwijs durven we dat (nog?) niet aan.

Tien jaar geleden schreef ik samen met Harm Tillema een artikel voor Examens over peer assessment (Leenknecht & Tillema, 2012). We gingen specifiek in op de vrees van docenten én studenten dat de kwaliteit van beoordelingen op het spel staat als studenten elkaar gaan beoordelen. Die vrees bestaat nog altijd. In het artikel spreken we bijvoorbeeld over de mogelijkheid dat de studentbeoordeling de docentbeoordeling vervangt, maar ik zie ook tien jaar later weinig opleidingen of scholen die dat aandurven.

Studenten die elkaar beoordelen of studenten die een zelfbeoordeling maken, zien we wel terug als onderdeel van formatief handelen. Peer- en self-assessment worden dan met name ingezet als leerinstrument, niet als beoordelingsinstrument dat deel uitmaakt van het programma van toetsen. Ook worden peer- en self-assessment toegepast als hulpmiddel bij docentbeoordelingen, bijvoorbeeld wanneer een student een zelfbeoordeling inlevert als oplegger van een portfolio. Op die manier vormt de zelfbeoordeling een leeswijzer voor de docent bij het doorneemen van het portfolio. Een dergelijke vorm van zelfbeoordeling kan ook worden ingezet voor het stimuleren van feedbackvragen en -gebruik, denk daarbij aan voorbeelden als de interactieve cover sheet (Arts et al., 2021).

Toch is er nog koudwatervrees voor het inzetten van studentbeoordelingen in het (formeel) beoordelingsproces. Een vrees die volgens mij veel gelijkenissen vertoont met de scepsis voor zelftesten tijdens de coronapandemie. En de kern van die vrees is betrouwbaarheid. Je kunt bij een zelftest of een studentbeoordeling nu eenmaal niet met zekerheid stellen dat beoordeling een volgende keer tot eenzelfde uitkomst leidt. Je weet niet of de beoordeling helemaal klopt. Dat weet je bij een PCR-test of docentbeoordeling ook niet met zekerheid, maar de zekerheidsmarge daarvan is groter. Al is dit laatste in de vergelijking tussen docent- en studentbeoordelingen trouwens eerder een ervaren gevoel dan een in onderzoek bevestigd gegeven (zie Falchikov & Goldfinch, 2000; Li et al., 2016).

Als het gaat om de coronatesten, hebben we op een bepaald moment gezamenlijk geaccepteerd dat de betrouwbaarheid van de zelftest minder is. De voordelen van de eenvoudig toepasbare zelftest – breder zicht op de verspreiding van het virus en snel kunnen afzonderen van (mogelijk) besmette personen – waren belangrijker dan de zekerheid waarmee een besmetting kon worden vastgesteld. We namen eventuele foute diagnoses dus op de koop toe, omdat we met meer zelftesten het virus effectiever konden indammen. Betrouwbaarheid werd daarmee niet minder belangrijk, maar het besef dat ook andere factoren van belang zijn, groeide. We brachten balans tussen de behoefte aan betrouwbare resultaten en de behoefte aan snel en breed zicht op de verspreiding.

Is dat niet ook de crux in het onderwijs? Zou meer en tijdige informatie over het ontwikkelingsniveau van onze leerlingen en studenten niet even belangrijk moeten zijn als betrouwbaarheid van de beoordeling? Ik schreef hier in 2017 samen met Simone Kooij een betoog over voor HO Management (zie Leenknecht & Kooij, 2017). En ik herhaal hier graag, in iets andere bewoordingen, de kernboodschap van dat betoog: laten we peer- en self-assessment omarmen zoals we ook de zelftest hebben omarmt! We hoeven als docenten niet iedere beoordeling die door studenten is uitgevoerd over te doen, we kunnen soms, for the time being, varen op de studentbeoordeling en gaandeweg werken aan de betrouwbaarheid van de beoordeling. Door het toevoegen van meerdere perspectieven, bijvoorbeeld van meerdere (mede)studenten, maar ook vanuit stage-instellingen of het bijbaantje van de student.

Daar komt bij dat als we studentbeoordelingen een meer prominente plaats geven in ons programma van toetsen, deze beoordelingen ook winnen aan betrouwbaarheid. Bij antiegeen-zelftesten staat de betrouwbaarheid vast, maar bij studenten is groei en ontwikkeling mogelijk in de betrouwbaarheid van hun beoordelingen. Als studenten meer ervaring opdoen met het beoordelen én ze leren goede prestaties te herkennen doordat ze meer inzicht krijgen in criteria en standaarden, dan neemt de betrouwbaarheid van hun beoordelingen alleen maar toe. Die beoordelingservaring draagt bovendien niet alleen bij aan de betrouwbaarheid van de beoordeling, maar stimuleert ook de kritische houding en zelfregulerende vaardigheden van studenten. Win-win dus.

Ik denk overigens niet dat we toe moeten naar “vanaf nu hoeft je alleen nog maar studentbeoordelingen te verzamelen om je diploma te krijgen”, zoals we wel toegingen naar alleen nog zelftesten bij corona-gerelateerde klachten, maar ik houd graag een pleidooi voor het meer serieus nemen van studentbeoordelingen in het onderwijs. Niet alleen als leerinstrument, maar ook als (onderdeel van de) formele beoordeling. Zowel studentbeoordelingen van medestudenten als

beoordelingen van eigen werk zijn kansen om breder en sneller zicht te krijgen op de ontwikkeling van de student, zowel voor de docent als de student zelf, die we niet kunnen laten liggen.

Referenties

- Arts, J. G., Jaspers, M., & Joosten-ten Brinke, D. (2021). Enhancing written feedback: The use of a cover sheet influences feedback quality. *Cogent Education*, 8(1), 1901641. <https://doi.org/10.1080/2331186X.2021.1901641>
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. <https://doi.org/10.3102/00346543070003287>
- Leenknecht, M. J. M., & Kooij, S. (2017). Leren van toetsen. Ontwikkeling als doel. *Hoger Onderwijs Management*, 2, 5-7.
- Leenknecht, M. J. M., & Tillema, H. H. (2012). Wie let op de kwaliteit van peer assessment? De rolverdeling tussen docent en student. *Examens, tijdschrift voor de toetspraktijk*, 9(1), 5-8.
- Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264. <https://doi.org/10.1080/02602938.2014.999746> ■

Gesignaleerd

Ook examiner heeft invloed op slagingspercentage rijscholen

Gepubliceerd op 05-08-2022, rijkschoolpro.nl

Uit een onderzoek van RTL Nieuws blijkt dat het veel uitmaakt bij welke examiner een leerling in de auto stapt. De onderzoeksredactie heeft door middel van cijfers van het CBR aangetoond dat tien procent van de examinatoren lagere of hogere scores heeft dan je zou verwachten.

Opvallend is dat er bij een aantal examinatoren uitzonderlijke verschillen zijn tussen de slagingspercentages van mannelijke en vrouwelijke kandidaten. “Om uit te sluiten dat dit verschil door toeval is ontstaan, heeft onze onderzoeksredactie drie hoogleraren gevraagd om naar de data te kijken. Zij stellen vast dat er zeker 38 examinatoren zijn bij wie aanwijzingen bestaan dat ze leerlingen benadelen op basis van hun geslacht,” vermeldt RTL Nieuws.

“Een uitschieter is een examiner bij wie 67 procent van alle mannelijke kandidaten in één keer het rijbewijs haalt. Kom je als vrouw bij deze examiner terecht dan is de kans op slaan aanzienlijk lager, namelijk 45 procent. Maar andersom gebeurt dit ook: bij een andere examiner slaagde 52 procent van de vrouwelijke kandidaten in één keer, maar slechts 33 procent van de mannen haalde in één keer het rijbewijs.”