



Opsteller: Sanne Jongeleen
Telefoon: (030) 253 3613
Email: s.jongeleen@uu.nl

GW/BBO/SJ/018/058

Aankruisen wat van toepassing is

ter bespreking

ter advisering

ter instemming

Aan de leden van de faculteitsraad

Onderwerp: Evaluatie pilot anoniem tentamineren

Evaluatie anoniem tentamineren

Op initiatief van twee docenten die verbonden zijn aan het departement Filosofie en Religiewetenschap, Ingrid Robeyns en Sander Werkhoven, is in de faculteit Geesteswetenschappen in blok 3 en 4 van collegejaar 2017-2018 een proef gestart met anoniem tentamineren: de docent ziet daarbij uitsluitend het studentnummer en niet de naam van de student wiens tentamen nagekeken wordt. Daarmee worden mogelijke (bewuste en onbewuste) *biases*, die zowel positief als negatief kunnen uitwerken, voorkomen. De argumentatie voor anoniem tentamineren hebben Robeyns en Werkhoven uitgewerkt in een artikel dat in maart 2018 is verschenen in DUB (zie hieronder).

Om anoniem tentamineren mogelijk te maken is speciaal tentamenpapier gedrukt, waarop studenten alleen hun studentnummer en niet hun naam kunnen invullen. Elektronisch toetsen binnen Remindo is eenvoudig te anonimiseren.

Bij circa 15 bachelor- en mastercursussen hebben docenten met anoniem tentamineren geëxperimenteerd. Docenten van deze cursussen is per mail gevraagd om enkele vragen te beantwoorden over hun ervaringen hiermee. Hieronder volgt een samenvatting van de reacties van zeven docenten.

(1) Reacties studenten op anoniem tentamineren

Studenten waren weinig verrast met de nieuwe manier van tentamineren, en vonden het niet meer dan logisch, of zelfs een goede zaak. Dit zou kunnen komen doordat docenten 'gewoon zijn gestart' met het anoniem tentamineren en de meesten het niet als verandering of iets nieuws hebben aangekondigd.

(2) Uitkomsten tentamens anders?

De meeste docenten geven aan dat dit moeilijk te zeggen is, aangezien het bij gebrek aan vergelijkingsmateriaal niet te meten is. Toch geven de meesten aan dat ze *denken* dat de uitkomsten wel degelijk anders kunnen zijn. Zo vinden zij het aannemelijk dat zij bij anoniem tentamineren eerder geneigd zijn om iemand die zij kennen als topstudent of iemand die goed

meedoet tijdens werkcolleges, beter te beoordelen dan wanneer objectief (anoniem) wordt getentamineerd.

(3) Best practices/tips

Docenten geven aan dat zij moesten wennen aan het werken met nummers in plaats van met namen. Tips heeft men vooral over de administratie. Zo geeft een docent aan dat het nu eerder kan gebeuren dat er iets fout gaat: je voert eerder een verkeerd nummer in dan een verkeerde naam. Eén docent verwijst naar een gebruik op een Engelse universiteit waar de secretaresses de cijfers in de systemen zetten aan de hand van een Excellijst.

(4) Voortzetten?

Docenten zijn erg positief over het anoniem tentamineren en raden aan dit voort te zetten. Zij zijn ervan overtuigd dat deze manier van toetsing en beoordelen objectiever is. Dat wil zeggen bij tentamens: docenten geven aan dat deze manier van toetsen niet, of minder goed, werkt bij essays of papers. Als kanttekening moet opgemerkt worden dat studenten die met speciale voorzieningen een tentamen maken geen gebruik kunnen maken van de anonieme variant: zij dienen op hun tentamenpapier aan te geven dat zij bijvoorbeeld dyslexie hebben. Bij kleine studentenaantallen is dit te herleiden tot de student.

Vervolg

Aangezien de eerste ervaringen positief zijn, zal de pilot het volgende vervolg hebben:

- De mogelijkheid om anoniem te tentamineren wordt behouden. Docenten die van de mogelijkheid gebruik willen (blijven) maken, kunnen hier dan voor kiezen, zonder dat dit een verplichtend karakter heeft.
- Docenten worden breed geïnformeerd over de mogelijkheid en voordelen van anoniem tentamineren en daarbij worden de resultaten van de pilot verspreid.
- Er wordt verder uitgezocht hoe anoniem tentamineren vergemakkelijkt kan worden en hoe eventuele knelpunten op te lossen zijn. Uit de pilot komen in ieder geval al de volgende punten naar voren:
 - De invoer van cijfers in Osiris minder foutgevoelig en docentvriendelijker maken (wat in elk geval al mogelijk is, is dat de docent bij het invullen van de cijferlijst in Osiris studenten kan sorteren op studentnummer in plaats van op achternaam).
 - Onderzoeken of er een koppeling tussen systemen (Remindo, Blackboard en Osiris) gemaakt kan worden om cijferinvoer te vergemakkelijken.
 - Inventariseren welke grofstoffelijke basisvoorzieningen op orde moeten zijn en deze leveren (b.v. drukken tentamenpapier, instellingen in Remindo).
 - Bekijken in hoeverre anoniem tentamineren de kans op fraude (iemand geeft zich uit voor een bepaalde student) vergroot.

Bijlage

<https://www.dub.uu.nl/nl/achtergrond/anoniem-tentamineren-verdient-de-voorkeur>

Anoniem tentamineren verdient de voorkeur

ACHTERGROND

Body:

Is een student vriendelijk, knap, of heel gemakzuchtig en lui? Docenten hebben te kampen met bewuste en onbewust veronderstellingen die mede bepalen of ze een student bij een subjectieve vorm van tentamen een hoger dan wel lager cijfer geeft. Sander Werkhoven en Ingrid Robeyns beargumenteren dat anoniem tentamineren dit kan voorkomen.

Door [Ingrid Robeyns](#), [Sander Werkhoven](#)

12/03/2018

Het is algemeen bekend dat elk mens, bewust of onbewust, bepaalde *biases* heeft: veronderstellingen waarmee we de wereld benaderen en begrijpen, en de handelingen van anderen proberen te voorspellen, anticiperen, en beoordelen. Ook al worden we in toenemende mate bewust van onze *biases*, niemand is er volledig vrij van. Bovendien is het al geruime tijd bekend dat bewustwording en zelfs de meest oprechte pogingen om *biases* te neutraliseren nooit geheel effectief zijn.

Bij open vragen of essays is er echter een onherroepelijke subjectieve component aanwezig in het beoordelingsproces

Ook docenten hebben *biases*, die op allerlei manier hun onderwijsactiviteiten beïnvloeden. Dat hoeft uiteraard niet altijd een probleem te zijn. Er *kan* een problematisch invloed ontstaan bij het beoordelen van toetsen, tentamens en essays. Wanneer een toets volledig 'objectief' is (zoals een meerkeuzevragen, of simpelweg goed of fout vragen), dan is het risico op *biases* bij de puntentoekenning vrijwel nihil. Bij open vragen of essays is er echter een onherroepelijke subjectieve component aanwezig in het beoordelingsproces, en bestaat het risico dat *biases* een invloed hebben op de beoordeling, en dus op het cijfer dat de student krijgt.

Gunstig als student een leuke of veelvoorkomende naam heeft

Er is onderzoek gedaan in hoeverre veronderstellingen inderdaad een rol spelen bij de beoordeling. In deze literatuur wordt verder gekeken dan naar de bekende *biases* op basis van het geslacht en de huidskleur van de persoon die beoordeeld wordt. De literatuur suggereert dat er ook een *bias* kan bestaan als een student

- (1) een leuke of veelvoorkomende naam heeft;
- (2) een hoog of juist laag cijfer heeft behaald op eerdere toetsmomenten – dit wordt het 'halo-effect' genoemd;
- (3) een aantrekkelijk of onaantrekkelijk uiterlijk heeft;
- (4) hard en bewust studeert of juist lui en gemakzuchtig is;
- (5) al dan niet interesse toont in het onderwerp van een cursus;
- (6) gedreven (of wanhopig) is om een hoog cijfer te halen;
- (7) bij een hoog cijfer belangrijke doeleinden kan bereiken;

- (8) intelligent spreekt en handelt of juist niet;
- (9) een connectie heeft met de docent.

Deze factoren kunnen ook een omgekeerd effect hebben. Als een docent probeert het effect te vermijden en daarin te ver doorschiet – dit wordt het *reverse* effect genoemd.

Een meta-analyse uit 2016 (waarbij 13 experimenten uit de VS en 7 elders waarvan 1 uit Nederland) blijkt dat er inderdaad een significant effect is bij de beoordelingen. In negatieve zin werd er effect gevonden bij

- (a) studenten uit specifieke ethnische groepen,
- (b) studenten die eerder slecht hebben gepresteerd;
- (c) minder aantrekkelijke studenten;
- (d) studenten met negatieve educatieve labels (bijvoorbeeld dyslectie)

Deze uitkomsten komen overeen met resultaten van ander onderzoek naar de rol van *biases*. De heterogeniteit van de resultaten suggereert dat ook *reverse biases* een rol spelen. Het maakt hierbij niet uit of een beoordelaar meer of minder ervaring heeft met beoordelen. Wel suggereert het bestaande onderzoek dat het gebruik van een beoordelingsrubriek (waarbij de docenten volgens een vaste reeks van beoordelingscriteria de toets beoordelen) het effect van *biases* vermindert (maar niet wegneemt). Positieve effecten worden niet genoemd of gespecificeerd in deze analyse.

Cijfers tonen niet aan dat iedere docent bevooroordeeld is, maar er bestaat wel een risico

Wat moeten we hiermee? Deze cijfers tonen natuurlijk niet aan dat iedere docent bevooroordeeld is en niet in staat zou zijn objectief te beoordelen. Het geeft wel reden te denken dat er een risico bestaat dat, in een organisatie zo groot als een universiteit, *biases* een rol spelen en onterecht studenten bevoordelen of benadelen.

Naast het gebruik van rubrieken, is een eenvoudige manier om *biases* te omzeilen de tentamens te anonimiseren. Alle bovengenoemde persoonsgebonden factoren zijn dan onbekend voor de beoordelaar, en dan kunnen ze dus ook geen rol van betekenis meer spelen. Deze strategie wordt sterk aanbevolen door verschillende onderzoekers, waaronder ook door de Nobelprijswinnaar Daniel Kahneman op basis van zijn onderzoek naar het nemen van beslissingen.

Eerder is er een appèl gedaan voor anoniem beoordelen door de studenten van Middlesex University, door Collegeboard.com, en vorig jaar door een groep hoogleraren uit Leuven. De reden is simpel: het is een effectief en goedkoop middel om de *mogelijke* effecten van naam, geslacht, kleur, talent, eerdere resultaten, en uiterlijk te voorkomen – en we hebben gegronde reden te denken dat deze effecten, in meer of mindere mate, bestaan. In het uitzonderlijke geval dat een docent toch redenen heeft om persoonsgebonden factoren mee te laten wegen, dan is dit nog altijd mogelijk (door het studentnummer op te zoeken) – maar dan doet een docent dit in ieder geval bewust.

We moeten de inzichten die voor anoniem tentamineren pleiten, serieus nemen

In alle andere gevallen kan een docent een onbewuste invloed van *biases*, in positieve en negatieve zin, voorkomen door middel van anoniem nakijken. Om die reden pleiten wij er dan ook voor om anoniem nakijken van schriftelijke tentamens de standaard te maken binnen de universiteit.

Wellicht zijn er collega's die praktische bezwaren zullen aandragen. Wij kunnen ze geruststellen. Wij hebben, overtuigd door de argumenten uit de literatuur, zelf al een aantal jaar onze tentamens anoniem beoordeeld. Natuurlijk zijn er toetsen waarbij het lastig is (bijvoorbeeld een paper over een onderwerp dat de student zelf kiest, en waarvan de docent met de student de commentaar op een eerdere versie doorsprak). Maar onze ervaring met anoniem nakijken hebben ons gesterkt in onze overtuiging dat we de inzichten uit de wetenschappelijke literatuur die voor anoniem tentamineren pleiten, serieus moeten nemen. We zijn meermaals verrast door de cijfers die sommige studenten kregen bij anoniem tentamineren - zowel lagere cijfers dan we zouden verwachten bij studenten die zeer actief waren tijdens het college, maar ook hogere cijfers van introverte of onzekere studenten die echter zeer goede toetsen afleggen. Op basis van de wetenschappelijke literatuur over *biases* mogen we ervan uitgaan dat die lagere cijfers wellicht (onterecht) iets hoger geweest waren bij niet-anoniem tentamineren, en dat de hogere cijfers van de onzekere studenten wellicht (onterecht) iets lager geweest waren. Anoniem nakijken is dus eerlijker, en dat zijn we studenten verschuldigd.

Literatuur

- Archer, J., and B. McCarthy. (1988). 'Personal biases in student assessment,' *Educational Research* 30:142–45.
- Baron, R. A., and D. Byrne. (2004). *Social psychology*, 10th ed. Boston: Pearson.
- Brennan, D. J. (2008). 'University student anonymity in the summative assessment of written work,' *Higher Education Research & Development*, 27: 43–54.
- Archer, J., and B. McCarthy. (1988). 'Personal biases in student assessment,' *Educational Research*, 30:142–45.
- Collegeboard.com. (2006). *SAT scoring guide*. http://www.collegeboard.com/student/testing/sat/about/sat/essay_scoring.html (accessed November 2017)
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Malouff, J. M. (2008). 'Bias in grading,' *College Teaching*, 56, 191–192.
- Malouff, J. M., Emmerton, A. J., & Schutte, N. S. (2013). 'The risk of a halo bias as a reason to keep students anonymous during grading,' *Teaching of Psychology*, 40, 233–237.
- Malouff, J. M. & Thorsteinsson, E.B. (2016). 'Bias in grading: A meta-analysis of experimental research findings,' *Australian Journal of Education* 60(3): 245–256.
- Nisbett, R., and L. Ross. 1980. *Human inference*. Upper Saddle River, NJ: Prentice Hall.

Warren Piper, D., McNulty, D. D., & O'Grady, G. (1996). *Examination practices and procedures in Australian universities*. Canberra, Australia: Australian Government Publishing Service.