



Plaats- en tijdonafhankelijk toetsen

Gedifferentieerd onderwijs

Differentiatie binnen het hoger onderwijs staat al een decennium hoog op de politieke agenda, maar dit heeft in de beginperiode voornamelijk betrekking gehad op verbreding van bacheloropleidingen en profilering van het master aanbod (OCW, 2011). In de strategische agenda van 2015-2025 (OCW, 2015) pleit men voor meer flexibele leerwegen om het onderwijs meer toe te spitsen op de individuele behoeften. In reactie bereiden universiteiten zich hier al enige tijd op voor, waarbij men zich onder andere richt op blended learning, integratie van massive open on-line courses (MOOC) in het curriculum en streeft naar plaats- en tijdonafhankelijke toetsing.

In het meerjarenplan 2015-2018 van SURF (2014) werd in het programma "Onderwijs op maat" de ambitie voor gepersonaliseerd onderwijs ook expliciet geformuleerd. SURF erkende dat one-size-fits-all niet toereikend is voor de diversiteit aan studenten (p. 34). De reden dat juist SURF dit op de agenda zette was niet verwonderlijk. Als ICT-samenwerkingsorganisatie van het hoger onderwijs in Nederland ziet SURF de onderwijstechnologische ontwikkelingen die de gewenste individualisering in alle redelijkheid mogelijk zouden kunnen maken. Ook de huidige versnellingsagenda (VSNU, VH en SURF 2018) noemt de flexibilisering van het onderwijs in haar ambities, waarbij wederom het plaats- en tijdonafhankelijk toetsen als streven op de agenda staat. Tot slot plaatst ook het meest recente "SURF 2Jarenplan 2021-2022 NL" (SURF, 2020) flexibilisering van het onderwijs op de prioriteitenlijst.

Hoewel het digitaal toetsen de eerste twee jaar van het versnellingsplan niet geheel prioriteit heeft gekregen, noodzaakt de huidige coronacrisis de vraagstukken hieromtrent met spoed op te pakken. Met name de plaatsafhankelijke toetsing staat in de schijnwerpers, terwijl voor vele instellingen de transitie naar digitaal toetsen ook al een behoorlijke uitdaging is.

Digitaal toetsen

Het summatief digitaal toetsen heeft de afgelopen tien jaar een prominente positie gekregen in het hoger onderwijs. Tentamenzalen zijn omgebouwd tot digitale tentamenzalen, waarbij de toename in het gebruik op sommige instellingen de maximale capaciteit al overstijgt. Landelijk was reeds een trend gaande naar meer mobiele oplossingen voor digitaal toetsen. Zo worden op een aantal instellingen chromebooks gebruikt of is gekozen voor bring your own device (BOYD) oplossingen. Ook zien we grote verschillen in de procesinrichtingen rond digitaal toetsen. Waar op sommige plekken docenten enkel verantwoordelijk zijn voor het maken van toetsvragen, moeten anderen de planning van de afname zelfstandig uitvoeren. Er zijn tevens initiatieven waarbij voor sommige domeinen de toetsing landelijk wordt aangeboden en de docent zich helemaal niet hoeft bezig te houden met de toetsing.

Deze ontwikkelingen doen vermoeden dat het digitaal summatief toetsen al lang en breed gemeengoed is. De transitie is echter met name voor de meer decentraal georganiseerde instellingen een grotendeels afdelingsspecifieke keuze. Het is dus zeker niet zo dat het digitaal toetsen overal al werd toegepast. De coronacrisis heeft iedereen hiertoe gedwongen, waarbij het plaatsafhankelijke toetsen er nog eens bovenop kwam.



Plaatsonafhankelijk

Het plaatsonafhankelijk toetsen werd pre-corona mondjesmaat toegepast voor speciale gevallen, zoals toetsing voor topsporters en decentrale selectie voor internationaal georiënteerde opleidingen. Het gebruik van online proctoring software is vanuit deze optiek al uitvoerig bestudeerd in het *Online proctoring for remote examination Erasmus+ project* (OP4RE, 2016). De schaal waarmee online proctoring gedurende de coronacrisis wordt ingezet is echter van ongekenne omvang. Het voornaamste doel van de inzet van online proctoring richt zich op fraudepreventie. Waarbij studentidentificatie en het naleven van regelgeving centraal staat. Hoewel online proctoring niets meer betekent dan online surveilleren zijn er echter verschillende toepassingsmogelijkheden. Deze lopen uiteen van het live toezicht op papieren tentamens via videostreamingdiensten tot gespecialiseerde software voor opnames en analyse. Met name deze laatste categorie doet veel stof opwaaien. Waarbij op het hoogtepunt op 25 november 2020 een kamermotie voor het afschalen van online proctoring is aangenomen. Ondanks duidelijke afspraken over bewaartermijnen, on campus alternatieven en verwerkersovereenkomsten is het sentiment over de inbreuk op privacy groot. Schaal- en kennisniveau vormen echter nog wel de voornaamste argumenten voor het ontbreken van alternatieven. Kennisreproductie toetsen met een groot aantal studenten blijft slecht te doen met essayvragen. Daarbij komt dat bij identificatie in een online setting het eenmalig controleren van de studentenkaart niet volstaat. Als we de identiteit van de student niet kunnen vaststellen dan gaat het plaatsonafhankelijk toetsen behoorlijk lastig worden.

Tijdonafhankelijk

Een toets maken op het moment dat het jou uitkomt klinkt heel mooi en wordt zelfs al toegepast. Het probleem is echter dat, in combinatie met plaatsonafhankelijk toetsen, na de eerste afname de tentamenvragen op straat liggen. Random item selectie zou hier natuurlijk uitkomst kunnen bieden, maar daar kleven flink wat haken en ogen aan. Instellingen die hun kwaliteitsborgingsproces rondom toetsing goed op orde hebben maken voornamelijk gebruik van klassieke testtheorie om de betrouwbaarheid van hun toetsen vast te stellen. Hoewel er technieken zijn om de betrouwbaarheid van toetsen met random getrokken items te bepalen (Draaijer & Klinkenberg, 2015), worden deze niet ondersteund in beschikbare toetsapplicaties. Een psychometrisch beter alternatief is moderne test theorie (IRT) waardoor met behulp van gekalibreerde itembanken computer adaptief (CAT) getoetst kan worden. Het toepassen van CAT is echter niet triviaal. Het vereist een grote bank met toetsvragen die niet op straat mogen komen te liggen, een toetsapplicatie die dit ondersteunt, een begrip van moderne testtheorie en deze kunnen uitleggen aan studenten. Hoe verklaar je bijvoorbeeld dat de ene student met 10 beantwoorde vragen het vak wel heeft gehaald versus een ander die er met 20 vragen niet doorheen komt. De toelichting dat de bovengrens van het betrouwbaarheidsinterval van theta onder de cesuur lag volstaat waarschijnlijk niet. Zonder inzet van moderne test theorie vormt het gebruik van random getrokken toetsvragen een grote bedreiging voor de kwaliteit van het toetsproces. Door gebrek aan grip op de moeilijkheid van de afgenomen vragen introduceren we willekeurige ongelijkheid. Waarbij de ene student eenvoudigere vragen voorgelegd kan krijgen dan een ander. Daarnaast hebben we geen enkel besef van de betrouwbaarheid van het cijfer. Hiermee introduceren we nog meer onzekerheid over de beoordeling van studenten. Vooral de onterechte negatieve beoordelingen kunnen hier veel schade aanrichten. Tijdonafhankelijke toetsing lijkt momenteel dan ook nog een brug te ver voor toepassing in het hoger onderwijs. Zolang de psychometrie niet op orde is, is dit zeer af te raden.



Toekomstperspectief

Voorzichtige schattingen laten zien dat het aantal formele digitale toetsen richting verdubbeling gaat. Dit zijn toetsen die worden afgenomen met de formele toets applicaties binnen de instellingen. Dezelfde applicaties waarvoor pre-corona een digitaal tentamen in een zaal geboekt zou worden. Als we ervanuit gaan dat docenten die hierop zijn overgestapt niet zullen terugkeren naar papieren tentaminering, dan doemt het beeld op dat we na de coronacrisis de fysieke capaciteit niet zullen hebben om al deze toetsen op de campus onder te brengen.

Het plaatsafhankelijk toetsen zou hier een uitkomst kunnen bieden. Hierbij zijn scenario's voor te stellen waarbij studenten van tevoren kunnen aangeven of zij op de campus of thuis willen toetsen. Daarbij moeten we echter wel gemeenschappelijke waarden formuleren waar studenteninstellingen en politiek het over eens zijn. Tevens moet dan worden vastgesteld dat scores voor de twee verschillende afnames niet structureel verschillen.

Waarom niet alleen formatief

Schildkamp (2020) toont in haar essay hoe formatieve toetsing inzicht geeft in het leerproces van studenten en hoe digitale varianten ingezet kunnen worden in een exclusief online scenario. Als formatieve toetsing zo informatief is, waarom tentamineren we dan nog? Als de psychometrie ons één ding geleerd heeft over toetsing, dan is het wel dat zelfs onder zeer gecontroleerde tentamenomstandigheden het lastig is om betrouwbare en valide toetsen te maken. Vergelijking van formatieve toetsresultaten met summatieve resultaten laat slechts een matig verband zien tussen de twee (Harlen, 2006; Klinkenberg, 2013). Daarbij geeft Harlen (2006) aan dat de informatie die voortvloeit uit formatieve toetsing wel goed te gebruiken is voor procesbegeleiding maar dat dit niet goed bruikbaar is voor cijferbepaling. Dat is ook niet zo vreemd gezien studenten die midden in het leerproces zitten de criteria van de leerdoelen nog niet behaald hebben. Het louter bepalen van het vaardigheidsniveau op basis van oefengegevens zou een verkeerd beeld geven van de werkelijke vaardigheid van studenten. De summatieve toets valt dus nog niet weg te denken uit ons onderwijssysteem.

Conclusie

De droom van flexibel onderwijs lijkt op het gebied van toetsing nog ver weg. Hoewel op sommige plekken al gepretendeerd wordt dat dit ideaal al gehaald is, moeten we ons realiseren dat de eenvoudige route zal leiden tot structurele kwaliteitsschending met ernstige gevolgen voor de eerlijke beoordeling van studenten. Plaats- en tijdonafhankelijk toetsen vereist een grondige herziening van de toegepaste psychometrie en een sluitend antwoord op online surveillance.

De coronacrisis heeft ons het digitale tijdperk in geschoten en er lijkt geen weg meer terug. Om hier ongeschonden uit te komen moeten we het toetsen op afstand nu goed regelen. In de komende jaren kunnen we ons richten op het veranderen van de psychometrische kwaliteitszorg en moeten we het formatief toetsen de plaats geven die het verdient, namelijk midden in het onderwijs.



Dit essay werd geschreven ter gelegenheid van het *NRO Symposium Hoger Onderwijs: Nieuwe richtingen na de pandemie?* dat op 15 januari 2021 heeft plaatsgevonden. Dit essay valt binnen het thema *Toetsen op afstand en monitoren van de studievoortgang*.

Contactgegevens auteur

Sharon Klinkenberg

Docent binnen de faculteit Psychologie Department of Psychology

Universiteit van Amsterdam

s.klinkenberg@uva.nl

Referenties

- Commissie Toekomstbestendig Hoger Onderwijs Stelsel. (2010). *Differentiëren in drievoud - advies van de Commissie Toekomstbestendig hoger onderwijs stelsel*
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record; Teach. Coll. Rec.*, 115(3)
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning. A meta-analysis and review of online learning studies*; U.S. Department of Education.
- OCW. (2011). *Kwaliteit in verscheidenheid - strategische agenda hoger onderwijs, onderzoek en wetenschap*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap
- OCW. (2015). *De waarde(n) van weten - strategische agenda hoger onderwijs en onderzoek 2015-2025*. Retrieved from <https://www.cdho.nl/assets/uploads/2016/08/Strategische-agenda-hoger-onderwijs-2015.pdf>. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap
- SURF. (2014). SURF meerjarenplan 2015 - 2018.
- SURF. (2020). *SURF 2Jarenplan 2021-2022*. Retrieved from <https://www.surf.nl/files/2020-12/surf-2jarenplan-2021-2022.pdf>
- VSNU, VH, & SURF. (2018). *Versnellingsplan onderwijsinnovatie met ICT*. Nederland: Retrieved from <https://versnellingsplan.nl/wp-content/uploads/2019/11/Versnellingsplan-2018.pdf>
- Draijer, S., & Klinkenberg, S. (2015, June). A practical procedure for the construction and reliability analysis of fixed-length tests with randomly drawn test items. In *International Computer Assisted Assessment Conference* (pp. 47-60). Springer, Cham.
- Op4re. (2016). Retrieved from <https://www.onlineproctoring.eu/nl/home-3/>
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. *Assessment and learning*, 2, 95-110.
- Schildkamp, K. (2021). Online formatief toetsen: Valkuilen en kansen. Paper presented at the *NRO Symposium Hoger Onderwijs: Nieuwe Richtingen Na De Pandemie?*