# Toetsen op afstand vereist een doordacht onderwijsontwerp

De coronapandemie heeft in het voorjaar van 2020 duidelijk gemaakt dat de traditionele manier van toetsen niet passend is op het moment dat studenten niet meer bij elkaar in een ruimte gebracht kunnen worden. Daarnaast is duidelijk geworden dat het wegvallen van toetsen aan het einde van een onderwijsperiode docenten te weinig inzicht geeft in de studievoortgang van hun studenten. Als er echter gesproken wordt over toetsen, wordt veelal alleen gedacht aan schriftelijke tentamens of examens. Toetsen is echter een brede onderwijsterm voor alles waarmee getoetst kan worden. Getoetst óf kennis aanwezig is, getoetst óf er geleerd is, getoetst óf iemand aanwezig of afwezig is. Nadenken over toetsen of over de kwaliteit van toetsen en dan specifiek over toetsen op afstand veronderstelt dan ook een verdere uitwerking van het begrip toetsen. In dit essay ga ik in op de summatieve en formatieve functies van toetsen in het onderwijs en de manier waarop deze functies een rol spelen in het nadenken over toetsen op afstand en het monitoren van studievoortgang.

## De summatieve functie van toetsen

Tot aan het begin van deze eeuw was de aandacht voor kwaliteit van toetsen vooral gericht op het toetsinstrument zelf en de wijze waarop dat toetsinstrument ingezet werd in de onderwijsprocessen van het onderwijs. Een toets moest zo betrouwbaar en valide mogelijke toetsscores opleveren om beslissingen te kunnen nemen over leerlingen en studenten. De definitie die Eggen in 2009 in zijn oratie gaf aan toetsen sluit het meest aan bij een definitie van de summatieve functie van toetsen: 'Toetsen is het meten dat leidt tot een score die uitdrukt of een bepaalde kwalificatie wel of niet aan een persoon kan worden toegekend of in welke mate de persoon een bepaalde eigenschap heeft' (p.5). In het hoger onderwijs worden toetsen ingezet om beslissingen te nemen, zoals selectie, plaatsing, classificatie of beheersing. De beheersingsbeslissing vindt meestal plaats aan het einde van een onderwijsperiode om vast te stellen of de stof beheerst wordt (Van der Linden, 1990). Een voldoende op de toets geeft aan dat er sprake is van voldoende beheersing. Een uitspraak over voldoende of onvoldoende beheersing van de stof veronderstelt dat er sprake is van een kwalitatief goede toets, dat wil zeggen dat de toetsscore een betrouwbaar en valide beeld geeft van de werkelijkheid. Onvoldoende kwaliteit van toetsing heeft mogelijk grote consequenties voor de student doordat het de geschiktheid, accuraatheid en geloofwaardigheid van de verzamelde informatie om een beslissing te nemen beïnvloedt (Gerritsen-van Leeuwenkamp, Joosten-ten Brinke, & Kester, 2017). Uit het onderzoek van Gerritsen-van Leeuwenkamp et al. blijkt dat standaardisering door te werken met vaste procedures, beoordelingsstandaarden of checklists kan bijdragen aan het verhogen van de kwaliteit van toetsen. De NRO-review van Maassen, den Otter, Wools, Hemker, Straetmans en Eggen (2015) laat zien dat artikelen die ingaan op de summatieve functie van toetsing nadrukkelijk aandacht besteden aan aspecten als normering en cesuurbepaling. Dit verklaren de auteurs doordat er bij toetsen met een summatieve functie de nadruk ligt op het nemen van min of meer belangrijke beslissingen op basis van de toetsscores. Daarnaast wordt de nadruk gelegd op de validiteit van toetsscore. De reden hiervoor is dat bij toetsen met een summatieve functie het belangrijk is dat een juiste interpretatie gedaan kan worden en een juiste (valide) beslissing wordt genomen op basis van de toetsscore.

Echter geen enkele toets levert 100% betrouwbare en valide scores op. Van der Vleuten en Schuwirth (2005) en andere onderzoekers maakten daarom duidelijk dat het belangrijk is om niet alleen naar de kwaliteit van een toets te kijken, maar om te kijken naar een toetsprogramma van een opleiding in zijn geheel. Baartman (2008) en Dijkstra (2014) lieten in hun proefschriften zien aan welke kwaliteitseisen zo'n toetsprogramma zou moeten voldoen en op basis van welke richtlijnen je zo'n toetsprogramma zou kunnen vormgeven. Een belangrijke conclusie op basis van onderzoeken uit die tijd is dat alleen kijken naar de summatieve functie van toetsen te beperkt is. Meer aandacht voor de formatieve functie van toetsen zou ervoor kunnen zorgen dat studenten meer betrokken worden bij hun eigen leerproces en daardoor uiteindelijk ook beter gaan presteren.

## De formatieve functie van toetsen

De aandacht voor formatief toetsen is echter al veel ouder. Scriven sprak al in 1967 over 'formative evaluation' naast 'summative evaluation' om te verwijzen naar continue evaluatie om het onderwijsproces te verbeteren (Scriven, 1967). Bloom (1969) verwoordde twee jaar later formatieve evaluatie als toetsen met als doel het geven van feedback in elke stap van het onderwijsleerproces. Formatieve evaluatie werd daarbij gezien als het gebruik van korte toetsen door docenten als hulpmiddel voor het leren. Daarna werd in de jaren tachtig de directe rol van de studenten bij het leren benoemd door Sadler en werd een eerste onderscheid gemaakt tussen assessment of learning en assessment for learning (Sadler, 1989). In het laatste decennium wordt formatief toetsen gekoppeld aan de effecten van formatief toetsen op het verbeteren van leerprocessen, verhogen van de motivatie van studenten en het ontwikkelen van (zelf)regulatievaardigheden (Clark, 2012; Meusen-Beekman, 2015; Sluijsmans, Joosten-ten Brinke & van der Vleuten, 2013; Schildkamp et al, 2014). Door middel van formatief toetsen kunnen instructiemethoden en zelfregulerende leerstrategieën door studenten gecombineerd worden (Clark, 2012). In het onderzoek naar formatief toetsen zien we een meer instrumentele benadering (zie bijvoorbeeld Sluijsmans et al, 2013) en een meer procesgeoriënteerde benadering (zie bijvoorbeeld Gulikers & Baartman, 2017). In de instrumentele benadering, ook wel gezien als een smalle definitie van formatief toetsen, is het onderzoek gericht op specifieke formatieve methoden die ingezet kunnen worden, zoals het geven van (peer)feedback, het organiseren van toetsdialogen of klassengesprek, self-assessments, inzetten van exit-tickets of het werken met rubrics. In deze smalle definitie gaat het ook om het inzetten van formatieve toetsen als toetsen die leraren en leerlingen informatie geven over het leerproces en onderdeel zijn van het leerproces. Een formatieve toets is dan een fysieke, digitale of papieren toets met open en/of gesloten vragen die tussentijds gebruikt wordt om het leren te bevorderen (Dirkx, Joosten-ten Brinke, & Camp, 2019). In de brede definitie van formatief toetsen gaat het meer om het proces van formatief toetsen. De Engelse term 'assessment for learning' wordt daarbij gebruikt: "… het proces van zoeken, aggregeren, interpreteren van informatie die studenten en docenten gebruiken om te bepalen waar studenten staan in hun leerproces, waar zij naartoe moeten en op welke manier" (Assessment Reform Group, 2002). In dit proces gaat het er vooral om wat er gedaan wordt met informatie die met verschillende instrumenten verzameld wordt (Wiliam & Leahy, 2015). De betrokkenheid van studenten en de voortdurende toetsing van effectiviteit van strategieën gedurende het leerproces zijn hierbij zeer belangrijk (William, 2007). Echter om die studenten daarin te trainen, is de rol van docenten belangrijk. Gulikers & Baartman (2017) hebben die rol van de docent uitgewerkt in hun formatieve toetscyclus. In deze cyclus is de docent bezig met de volgende vijf fasen: (1) verwachtingen verhelderen, (2) studentreactie(s) ontlokken, (3) studentreactie(s) analyseren en interpreteren, (4) communiceren over resultaten met studenten en (5)

vervolgacties ondernemen voor onderwijs en leren. Deze fasen staan niet los van elkaar, maar worden in relatie met elkaar doorlopen (Gulikers & Baartman, 2017).

### De samenhang tussen de formatieve en summatieve functie van toetsen

De summatieve functie van toetsen zorgt er dus voor dat vastgesteld kan worden of er sprake is van beheersing, de formatieve functie van toetsen zorgt ervoor dat een continue ontwikkeling wordt gerealiseerd. Het is daarom van belang om direct bij het ontwerpen van het onderwijsprogramma al na te denken over de toetsing. Hoe zorg je ervoor dat er voldoende informatie over de ontwikkeling van studenten verzameld wordt om adviezen te geven én om beslissingen te nemen. Nadenken over toetsing is daarmee nadenken over het toetsprogramma van een opleiding, een 'bewust samengesteld geheel van toetsen' (Baartman, 2007). De relatie tussen een toetsprogramma en het onderwijs is hierbij een belangrijk criterium, maar het hebben van een goed toetsprogramma garandeert nog niet dat er een integratie is van toetsen en leren ten behoeve van de ontwikkeling van de student. Deze integratie van toetsen en leren en de competentieontwikkeling van de student staan centraal bij programmatisch toetsen (Baartman, Van Schilt-Mol, & Van der Vleuten (2020). Bij programmatisch toetsen wordt bewust niet meer gesproken over de formatieve en summatieve functie van toetsen, maar over low-stakes en high-stakes beslissingen. De zwaarte van een beslissing die genomen moet worden is leidend voor de wijze waarop informatie over een student (een datapunt genaamd) verzameld wordt. Een mix van datapunten zorgt voor inzicht in de ontwikkeling en biedt de mogelijkheid om beslissingen te nemen.

### Het ontwerp van je toetsprogramma

2020 heeft ons duidelijk gemaakt dat wanneer we beslissingen over studenten willen nemen uitsluitend op basis van toetsen die aan het einde van een onderwijsperiode plaatsvinden, we voor haast onoverkomelijke problemen komen te staan. Het ineens aanpassen van een toetsvorm naar een toetsvorm die thuis zonder toezicht afgenomen kan worden én nog aansluit bij de leerdoelen is niet altijd mogelijk. Het verplaatsen van een toets die gepland was voor een afname op locatie een-op-een verplaatsen naar een locatie op afstand is mogelijk indien de toets al digitaal was en de onderwijsinstelling beschikt over software voor online proctoring. Echter het inzetten van online proctoring levert nog verschillende dilemma's op met betrekking tot privacy en eerlijkheid. Vragen over de mogelijkheden voor toetsen op afstand en de risico's die daarbij optreden maakten duidelijk dat het onderwijs niet voorbereid was op online onderwijs en online toetsen. Deze constatering maakt duidelijk dat onderwijs *ontworpen* moet worden en niet ad hoc kan worden bedacht. De locatie en de vorm van afname is namelijk een van de randvoorwaarden waar je bij het ontwerpen van je onderwijs en in dit geval het ontwerpen van de toetsing al rekening mee moet houden. Een tweede constatering is dat in het onderwijs veel gewicht gehangen wordt aan een toets die aan het einde van een onderwijsperiode aan studenten wordt voorgelegd. Op basis van zo'n toets worden belangrijke beslissingen genomen. Het wegvallen van zo'n toets doordat de randvoorwaarden zijn veranderd heeft daardoor grote consequenties. Zowel voor de studenten, maar ook voor het hele onderwijssysteem. Ik wil dit essay dan ook eindigen met een paar aanbevelingen. Zorg bij het ontwerpen van het toetsprogramma van een opleiding dat de informatie die over de ontwikkeling van de student en zijn of haar beheersing van de stof beter verspreid over het jaar verzameld wordt. Maak hierbij gebruik van toetsen die zowel een summatieve als formatieve functie vervullen. Kijk bij het toetsontwerp ook direct naar de digitale mogelijkheden voor toetsing. Onze ervaring is dat het betrekken van de studenten bij dit ontwerp bijdraagt aan acceptatie van bijvoorbeeld het surveilleren op afstand

(Welling, 2020). Ook het geven van ruimte en vertrouwen aan docenten om uit te gaan van mogelijkheden en niet uit te gaan van problemen leveren mooie oplossingen op (Welling, 2020). De randvoorwaarde bij het ontwerpen van onderwijs en toetsing is vanaf dit jaardat je ervanuit moet gaan dat onderwijs én toetsen soms uitvallen. Studenten mogen daar niet de dupe van zijn.

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

Desirée Joosten-ten Brinke

Decaan en Hoogleraar Leren van volwassenen

Open Universiteit

desiree.joosten-tenbrinke@ou.nl

### Referenties

Assessment Reform Group (2002). Assessment for learning: 10 principles. Research-based principles to guide classroom practice. Opgehaald via http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Afl_principles.pdf

Baartman L. J. (2008). Ontwikkeling en gebruik van kwaliteitscriteria voor Competentie Assessment Programma's. Proefschrift. Utrecht: Universiteit Utrecht.

Bloom, B. (1984). The search for methods of group as effective as one-to-one tutoring. Educational Leadership, 41 (8), 4-17.

Clark, (2012). What matters most for student assessment systems: a framework paper. Washington: the World Bank. Opgehaald via https://openknowledge.worldbank.org/bitstream/handle/10986/17471/682350wp00publ0wp10read0web04019012.pdf?sequence=1.

Dijkstra, J. (2014). Guidelines for designing programmes of assessment. Proefschrift. Maastricht: Universiteit Maastricht.

Eggen, T. J. H. M. (2009). De kwaliteit van toetsen. Oratie. Enschede: Universiteit Twente.

Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. Studies in Educational Evaluation, 55, 94-116.

Gulikers, J. & Baartman, L. J. (2017). Doelgericht professionaliseren: formatief toetsen met effect! Wat DOET de docent in de klas. NWO-PPO overzichtsstudie eindrapportage (dossiernummer 405-15-722). Den Haag: NWO. https://www.nro.nl/onderzoeksprojecten-vinden/?projectid=405-15-722-doelgericht-professionaliseren-formatieve-toetscompetenties-met-effect.

Maassen, N. A. M., den Otter, D., Wools, S., Hemker, B. T., Straetmans, G. J. J. M., & Eggen, T. J. H. M. (2015). Kwaliteit van toetsen binnen handbereik: Reviewstudie van onderzoek en onderzoeksresultaten naar de kwaliteit van toetsen. Pedagogische studiën, 92(6), 380-393.

Meusen-Beekman, K. (2015). Bridging the gap between primary and secondary education. Proefschrift. Heerlen: Open Universiteit.

Sadler, (1989). Formative assessment: revisiting the territory. Assessment in Education: Principles, Policy and Practice, 5, 77-84.

Schildkamp, K., Heitink, M., Van der Kleij, F., Hoogland, I., Dijkstra, A., Kippers, W., & Veldkamp, B. (2014). Voorwaarden voor effectieve formatieve toetsing. Een praktische review. Reviewstudie

uitgevoerd in opdracht van en gesubsidieerd door NRO-PPO: Dossiernummer 405-14-534, eindrapport 30-09-2014. Enschede: Universiteit Twente.

Scriven, M. (1967) The methodology of evaluation, in: R. W. Tyler, R. M. Gagné & M. Scriven (Eds) Perspectives of Curriculum Evaluation. Chicago: Rand McNally.

Sluijsmans, D. M. A., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2013). Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen. Den Haag: Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Van der Linden, W. J. (1990). Applications of decision theory to test based decision making. In: R. K. Hambleton, & J. N. Zaal (1990). Advances in educational and psychological testing (pp. 129-156). Boston: Kluwer.

Van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. Medical Education, 39, 309-317.

Welling, P. (2020). Hoe Fontys in 1 keer toetsen online realiseerde. Eindhoven: Fontys Hogescholen.

Wiliam, D. & Leahy, S. (2015). Embedding formative assessment: Practical techniques for k-12 classrooms. West Palm Beach, Florida: Learning Sciences International.