

# Inzicht in de kwaliteit van (digitale) toetsen

Een descriptief onderzoek door een literatuurstudie en interviews met toetsexperts

## ■ Sjoerd Lindenburg MA

Sjoerd Lindenburg MA is onderzoeksstagiair bij het onderwijsinnovatieprogramma Educate-it van de Universiteit Utrecht. Email: s.lindenburg@uu.nl

## ■ Rik Vangangelt MA

Rik Vangangelt MA is projectmedewerker en stagecoördinator bij het onderwijsinnovatieprogramma Educate-it van de Universiteit Utrecht. Email: r.h.a.vangangelt@uu.nl

## ■ Dr. Gemma Corbalan

Dr. Gemma Corbalan is onderzoeker, adviseur en trainer bij de afdeling Onderwijsadvies en Training van de Universiteit Utrecht. Email: g.corbalan@uu.nl

Een van de nieuwe ontwikkelingen op het gebied van toetsen in het hoger onderwijs is de digitale toetsafname. De Universiteit Utrecht (UU) beschikt sinds maart 2015 over een zeer grootschalige digitale toetsvoorziening voor formatieve én summatieve toetsen. Deze studie streeft ernaar meer inzicht te krijgen in de kwaliteit van (digitale) toetsen door middel van een literatuurstudie en interviews met toetsexperts.

### Introductie

De Universiteit Utrecht (UU) beoogt de kwaliteit van het onderwijs voortdurend te verbeteren, bijvoorbeeld met behulp van innovatie en technologie. Eén van de nieuwe ontwikkelingen is de digitale toetsafname. Het universiteitsbrede onderwijsinnovatieprogramma Educate-it ondersteunt docenten bij het gebruik van het digitale toetssysteem Remindo. Uit een eerdere evaluatie (Corbalan, de Kleijn & Manrique, 2016) blijkt dat zowel studenten als docenten positief tegenover digitale toetsen en Remindo staan. Daarentegen

zijn beide groepen niet van mening dat de mogelijkheden van digitaal toetsen (bv. grote variëteit aan vraagtypen) helpen om toetsen beter aan te laten sluiten op de leerdoelen. Educate-it wil meer inzicht in de kwaliteit van digitale toetsen. Daarom wordt in dit artikel de volgende onderzoeksvraag beantwoord:

Welke (invloeden op) kwaliteitscriteria van (digitale) toetsen zijn te vinden in de literatuur en welke zijn er volgens de toetsexperts?

### Methodologie

#### Literatuurstudie

De eerste stap van deze literatuurstudie was het zoeken naar relevante zoektermen en selectiecriteria op het gebied van (digitale) toetsen in het hoger onderwijs. Dit gebeurde in de literatuur en werd bevraagd in de interviews met toetsexperts. Op basis hiervan is gezocht via een vijftal databases: Google Scholar, ScienceDirect.com, Scopus, ERIC en Web of Science. Alle hits zijn gefilterd op niet-werkende links, dubbele hits en het voldoen aan de selectiecriteria.

In totaal zijn er 49 relevante artikelen met betrek-

king tot de kwaliteit van (digitale) toetsen in het hoger onderwijs gevonden. Deze literatuurlijst bestond uit één talk en 48 artikelen, waarvan zeventien verslagen, veertien theoretische papers, vier literatuurstudies, vijf non-experimentele studies, drie overzichten, vier experimentele studies en één rapportage. 41 artikelen gingen direct over digitale toetsen en acht artikelen over toetskwaliteit in het algemeen.

### Interviews

Vier toetsexperts uit verschillende hogeronderwijsinstellingen in Nederland zijn geïnterviewd. De interviewleidraad was onderverdeeld in drie thema's:

1. De kwaliteitscriteria van (digitale) toetsen;
2. Invloeden op de kwaliteit van digitale toetsen;
3. Digitale voordelen die kunnen leiden tot kwaliteitsverbetering van toetsen.

## Resultaten

### Kwaliteitscriteria bij (digitale) toetsen

Tijdens de interviews kwamen bij iedere toets-expert dezelfde kwaliteitscriteria naar voren, overeenkomend met de kwaliteitscriteria voor toetsen in het algemeen (bv. Gerritsen-van Leeuwenkamp et al, 2017). Deze kwaliteitscriteria kunnen als volgt worden gedefinieerd:

- Validiteit: In hoeverre de toets meet wat bedoeld is te meten;
- Betrouwbaarheid: In hoeverre er vertrouwen kan zijn in de toetscore als meting, ongeacht de inhoud van de toets;
- Transparantie: Alle informatie is voorhanden. (Van Berkel & Bax, 2014).

Eén toetsexpert gaf aan dat vooral transparantie complexer ligt bij digitale toetsen. Daarnaast gaven twee experts aan dat de focus niet alleen op de summatieve eindtoets moet liggen, maar juist ook op het hele leerproces van de student. In de geselecteerde literatuur zijn dezelfde drie kwaliteitscriteria te herkennen (tabel 1).

Transparantie is het minst vaak genoemde criterium in de literatuur – volgens de toetsdeskundigen komt dit waarschijnlijk doordat digitale toetsystemen nieuw en anders zijn, waardoor ze juist minder transparant zijn. In de literatuur werd geopperd dat studenten moeten kunnen oefenen met digitale toetsystemen (McKenna & Bull, 2000), dat er een duidelijke syllabus moet zijn (Buijns & Kok, 2014), evenals een toetsprotocol (Peet, 2017) en rubrics (Evertse, 2014a).

(Digitale) toetsen moeten volgens de artikelen en toetsexperts ook valide en betrouwbaar zijn (bv. Conole & Warburton, 2005; Peet, 2017). Helaas werd niet duidelijk hoe deze criteria werden beïnvloed. Evertse (2014a) suggereert bijvoorbeeld dat toetsen objectiever kunnen worden nagekeken met een digitale toetsstelsel, maar maakt hierbij geen link naar 'betrouwbaarheid'.

### Randvoorwaarden bij digitale toetsen

Volgens twee toetsexperts bestaan er ook randvoorwaarden (beveiliging, infrastructuur en databeheer) die aan de basis van goede digitale toetsen staan. Uit de geselecteerde artikelen kwamen dezelfde drie randvoorwaarden naar voren (tabel 2).

Tabel 1. Kwaliteitscriteria bij digitale toetsing

	Artikelen over digitale toetsing	Artikelen over toets-kwaliteit in het algemeen	Totaal
Validiteit	24	7	32
Betrouwbaarheid	30	8	39
Transparantie	4	7	11

Tabel 2. Randvoorwaarden bij digitale toetsing

	Aantal artikelen
Beveiliging	21
Infrastructuur	8
Databeheer	2

Digitale toetsystemen en het netwerk van hogeronderwijsinstellingen moeten goed worden beveiligd zodat alleen bevoegden bij de toetsgegevens kunnen. Dit kan met behulp van bijvoorbeeld een securitymanager (Zakrzewski & Steven, 2003). Ook wordt een goede infrastructuur aanbevolen om digitale toetsen mogelijk te maken, zoals het beschikbaar stellen van genoeg computers (Draaijer et al., 2004) of ad-random kunnen toetsen (Draaijer et al., 2014). Daarnaast benadrukken twee studies het belang van goed databeheer, mede vanwege de beveiliging van gevoelige toetsinformatie (Buzetto-More & Alade, 2006; Doukas & Andreatos, 2007).

**Factoren met directe invloed op de kwaliteit van digitale toetsen**

Uit deze studie blijken ook een aantal kwaliteitsfactoren. Deze kwaliteitsfactoren hebben volgens



**Digitale toetsen moeten ad-random kunnen plaatsvinden**

de experts invloed op de verschillende kwaliteitscriteria (tabel 3) en kunnen daarmee invloed hebben op de toetskwaliteit. De meeste van deze

Tabel 3. Directe kwaliteitsfactoren bij digitale toetsing

Invloed		Factoren
Invloed op validiteit	Psychometrische analyse	Vragenbanken
	Toetsmatrijs	Aansluiting op de praktijk
	Multimedia	Vraagtypen
	Samenwerking tussen docenten	Standaardisatie
		Interactiviteit
Invloed op betrouwbaarheid	Interbeoordelaarschap	Psychometrische analyse
	Anoniem nakijken	Leesbaarheid
	Standaardisatie	Horizontaal nakijken
	Rubrics	Editen van vragen tijdens het proces
Invloed op transparantie	Vraagtypen	Gebruikersvriendelijkheid
	Toetsevaluatie door studenten	Fairness
	Standaardisatie	Studentondersteuning

kwaliteitsfactoren kunnen ook ingezet worden bij papieren toetsen, maar zijn volgens toetsexperts makkelijker en sneller in te zetten bij digitale toetssystemen.

Veruit de meeste kwaliteitsfactoren hebben invloed op de toetsvaliditeit, al moet hier volgens de toetsexperts wel op een didactisch verantwoorde manier mee omgegaan worden. Het toevoegen van een video aan een toets zonder duidelijke link tussen de video en de toets(vraag) kan voor verwarring zorgen. Volgens één toetsexpert is ook de grote variëteit aan vraagtypen daarom minder wenselijk voor summatieve toetsen.

Het creëren van een vragenbank verhoogt de toetsrepresentativiteit, zeker in combinatie met een toetsmatrijs (Draaijer et al., 2014). Daarnaast kan een vragenbank geoptimaliseerd worden met de psychometrische analyse (Berkel & Bax, 2014) en feedback van studenten (Evertse, 2014a). Al met al zorgt dit voor meer gestandaardiseerde toetsen (Gerritsen-van Leeuwenkamp et al, 2017).

Ook wordt er door kwaliteitsfactoren invloed uitgeoefend op de transparantie en betrouwbaarheid. Voor de betrouwbaarheid zijn dit met name kwaliteitsfactoren die ingezet kunnen worden om objectiever na te kijken (bv. anoniem nakijken en

interbeoordelaarschap) (Draaijer et al., 2014). Voor de transparantie zijn dit factoren zoals een gebruikersvriendelijk systeem (Draaijer et al., 2014), studentondersteuning (Hols-Elders et al, 2009) en feedback van studenten (Escudier et al., 2014). Ook wordt de transparantie verhoogt wanneer een toets fair is (Gerritsen-van Leeuwenkamp et al., 2017).

**Factoren met indirecte invloed op de kwaliteit van digitale toetsen**

Zowel de toetsexperts als de literatuur noemden daarnaast nog kwaliteitsfactoren met betrekking tot digitale toetssystemen die indirect invloed hebben op de toetskwaliteit (tabel 4).

Voor docenten kan er tijdswinst worden behaald, zowel op de korte termijn (automatisch nakijken/snel feedback systeem; Evertse, 2014a), als op de lange termijn (via een vragenbank; Peet, 2017). Deze extra tijd kan werkdruk van docenten wegnemen (minder stress) en de herwonnen tijd kan ergens anders worden ingezet.

Ook zijn er kwaliteitsfactoren die invloed hebben op het leerproces van de student. De flexibiliteit van digitale toetssystemen verschaft de optie om werkplekonafhankelijk te kunnen toetsen (SURF, 2014). Digitale formatieve toetsen kunnen dan efficiënt in worden gezet met behulp van ad-

Tabel 4. Indirecte Kwaliteitsfactoren bij digitale toetsing

Voor wie	Factoren
Voor de docenten	Automatisch/minder nakijken Snelle/makkelijke feedback Tijdswinst/lagere werkdruk/minder stress
Voor het leerproces van de student	Flexibiliteit Ad-random vragen/vragengenerator Toetsproces Formatieve toetsing Kenmerken van de student/ICT-ervaring
Voor de hoger-onderwijsinstelling	Docentprofessionalisering/-ondersteuning Invloed van betrokkenen

random vragen, om de formatieve toets – en dus het hele toetsproces – beter aan te laten sluiten bij het individuele leerproces van iedere student (CITO, 2015; Thelwall, 2000).

Daarnaast blijkt uit zowel de literatuur als de interviews dat het ondersteunen en opleiden van docenten op hogeronderwijsinstellingen belangrijk is bij het gebruik van digitale toetsystemen (Zakrweski & Steven, 2003). Docenten gebruiken immers de systemen en maken de toetsen; hun kennis van het systeem heeft daardoor veel invloed op de toetskwaliteit. Een slechte docentondersteuning kan de toetskwaliteit zelfs negatief beïnvloeden.

## Conclusies

Het doel van dit onderzoek was om meer inzicht te krijgen in de kwaliteit van digitale toetsen. Er is daarom in de literatuur en in interviews met toetsexperts gezocht naar (invloeden op) kwaliteitscriteria van (digitale) toetsen. Uit zowel de literatuurstudie als de interviews blijkt dat dezelfde kwaliteitscriteria gelden voor digitale toetsen als papieren toetsen, namelijk validiteit, betrouwbaarheid en transparantie. Transparantie is daarbij het meest complexe criterium bij digitale toetsen ten opzichte van papieren toetsen, volgens de toetsexperts waarschijnlijk doordat digitale toetsystemen nieuw en anders zijn. Wel denken zij dat eventuele problemen op dit gebied overkomelijk zijn. Daarnaast blijken uit de literatuur en de interviews ook randvoorwaarden voor digitale toetsen, namelijk: beveiliging, infrastructuur en databeheer.

Uit deze studie blijken ook kwaliteitsfactoren die ingezet kunnen worden om de toetskwaliteit te beïnvloeden. Bijvoorbeeld vragenbanken hebben invloed op de validiteit, interbeoordelaarschap op de betrouwbaarheid en formatieve toetsen op de transparantie. De kwaliteitsfactoren kunnen over het algemeen ook worden ingezet bij papieren toetsen, maar zijn volgens de toetsexperts makkelijker en sneller in te zetten bij digitale toetsen. Andere kwaliteitsfactoren kunnen juist indirect invloed hebben op de toetskwaliteit van digitale



## Het creëren van een vragenbank verhoogt de toetsrepresentativiteit

toetsen, zoals tijdswinst voor docenten en ad-random formatieve toetsen voor de studenten. Ook wordt door zowel de experts als door de literatuur genoemd dat docenten aan de basis van iedere (digitale) toets staan en daarmee veel invloed kunnen hebben op de toetskwaliteit. Het is daarom belangrijk hun goede scholing én ondersteuning te bieden.

Docenten en docentondersteuners kunnen gebruikmaken van de onderzoeksresultaten uit deze studie om inzicht te krijgen in (de invloeden op) de kwaliteit van (digitale) toetsen. Uiteraard zal de Universiteit Utrecht onderzoek blijven doen om de kwaliteit van de (digitale) toetsen te verhogen en hun docenten optimaal te blijven ondersteunen.

## Referenties

- Berkel, H. van, & Bax, A. (2014). Toetsen: Toetssteen of dobbelsteen. In H. van Berkel,

- A. Bax, & D. Joosten-ten Brinke (Eds.) *Toetsen in het hoger onderwijs* (pp. 15-28). Houten: Bohn Stafleu van Loghum.
- Bruijns, V., & Kok, M. (2014). *Guidelines for Testing and Assessment*. Retrieved from <https://score.hva.nl/Bronnen/Guidelines%20Testing%20and%20Assessment.pdf>
  - Buzzetto-More, N., & Alade, A. (2006). Best practices in e-assessment. *Journal of Information Technology Education*, 5(1), 251-269.
  - CITO. (2015). *Prestaties op papieren en digitale examens: Wat is het verschil?* Arnhem: Stichting Cito.
  - Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *The Journal of the Association for Learning Technology*, 13(1), 17-31.
  - Corbalan, G., Klein, R. de, & Manrique, M. (2016). Ervaringen van studenten en docenten met digital toetsen bij de UU. *Examens*, 13(4), 18-23.
  - Doukas, N., & Andreatos, A. (2007). Advancing electronic assessment. *International Journal of Computers Communications & Control*, 2(1), 56-65.
  - Draaijer, S., van Gastel, L., Peeters, V., Frinking, P., & Reumer, C. (2004). Flexibilisering van toetsing. *Flexibility in Assessment*. Utrecht: Digital Universiteit.
  - Draaijer, S., Boxel, P. van, & Brunschot, M. van. (2014). Digitaal toetsen. In H. van Berkel, A. Bax, & D. Joosten-ten Brinke (Eds.) *Toetsen in het hoger onderwijs* (pp. 91-100). Houten: Bohn Stafleu van Loghum.
  - Escudier, M., Tricio, J., & Odell, E. (2014). Student acceptability of high-stakes e-assessment in dental education: Using privacy screen filters to control cheating. *Journal of Dental Education*, 78(4), 558-566.
  - Evertse, J. (2014a). Digitaal toetsen: Kansen voor het hoger onderwijs. *SURF*. Artikel gevonden via: <https://www.surf.nl/themas/onderwijsinnovatie-met-ict/digitaal-toetsen>
  - Evertse, J. (2014b). 5 kansen van Digitaal Toetsen – voor docenten. *SURF*. Artikel gevonden via: <https://www.surf.nl/themas/onderwijsinnovatie-met-ict/digitaal-toetsen>
  - Gerritsen-van Leeuwenkamp, K., Joosten-ten Brinke, D., & Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. *Studies in Educational Evaluation*, 55, 94-116.
  - Hols-Elders, W. P. M., Bloemendaal, P. M., Bos, N. R., Quaak, M. J., Sijstermans, R., & de Jong, P. G. M. (2009). Twaalf tips voor computerondersteund toetsen in het medisch onderwijs. *Tijdschrift voor Medisch Onderwijs*, 28(1), 4-12.
  - McKenna, C. & Bull, J. (2000). Quality assurance of computer-assisted assessment: Practical and strategic issues. *Quality Assurance in Education*, 8(1), 24-32,
  - Peet, A., (2017). Digitale toets- en vragenbanken in het onderwijs: praktijkexperts aan het woord over kansen en mogelijkheden. *SURF*. Artikel gevonden via: <https://www.surf.nl/themas/onderwijsinnovatie-met-ict/digitaal-toetsen>
  - SURF. (2014). Stimuleringsregeling: Gezamenlijk gebruik van toetssoftware in de cloud. *SURF*. Artikel gevonden via: <https://www.surf.nl/themas/onderwijsinnovatie-met-ict/digitaal-toetsen>
  - Thelwall, M. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education*, 34(1), 37-49.
  - Zakrzewski, S. & Steven, C. (2003). Computer-based assessment: quality assurance issues, the hub of the wheel. *Assessment & Evaluation in Higher Education*, 28(6), 609-623.