Kan met de portfoliomethodiek betrouwbaar en valide summatief getoetst worden in het mbo?

Beantwoord door Maartje Kouwenberg en Christa Teurlings (kennismakelaars Kennisrotonde).

24 maart 2022

KR. 1381

Kort antwoord

Er is onvoldoende onderzoek en daarom onvoldoende bewijs dat een portfolio een betrouwbare en valide instrument is voor summatieve toetsing bij mbo-studenten. Aannemelijk is dat de betrouwbaarheid en validiteit sterk afhankelijk zijn van de manieren waarop opleidingen portfolio's samenstellen en beoordelen.

Opleidingen kunnen deze betrouwbaarheid en validiteit op verschillende manieren versterken. Bijvoorbeeld door te werken met portfolio's die een mix aan opdrachten en taken bevatten en door meerdere getrainde beoordelaars de portfolio's met eenduidige criteria te laten beoordelen.

Dit weten we vooral uit studies naar programmatisch toetsen (in het hoger onderwijs). Dit is een integrale aanpak waarin portfolio's worden ingezet om zowel formatief als summatief te toetsen. Een overzichtsstudie hiernaar bij universitaire zorgopleidingen geeft een positief beeld, maar studies binnen met mbo ontbreken vooralsnog.

Antwoord

Middelbare beroepsopleidingen (mbo's) vinden de bestaande wijze van examineren niet altijd aansluiten bij hun onderwijsconcept. Zo zien opleidingen met een grote praktijkcomponent het examineren als dubbelop of ze denken dat het maatwerk in de weg zit (Willemse & van der Neut, 2020). Daarom kijken ze naar andere manieren van toetsen, zoals naar portfolio's.

Portfolio's kunnen verschillen in doelen

Portfolio's zijn er in verschillende soorten en maten en kunnen verschillende doelen hebben. Er zijn twee doelen waarmee opleidingen gebruik kunnen maken van portfolio's om te toetsen (Roberts et al., 2014).

Ten eerste kunnen opleidingen portfolio's inzetten om *formatief* te toetsen. Zij gebruiken de portfolio's om het leerproces van studenten in kaart te brengen en te stimuleren. Dit worden ontwikkelingsgerichte portfolio's genoemd.

Ten tweede kunnen opleidingen portfolio's inzetten om te beoordelen of studenten leerdoelen hebben behaald. Dit worden beoordelingsportfolio's genoemd. Hier worden portfolio's dus ook voor *summatieve*

toetsing ingezet. Dergelijke portfolio's bestaan vaak uit verzamelingen van bewijstukken waarmee studenten aantonen dat zij de competenties van een opleiding beheersen (Van Berkel et al, 2017). Het zijn bijvoorbeeld beschrijvingen van relevante ervaringen, een overzicht van opgedane competenties, cijfers voor toetsen en verslagen van stagebegeleiders (Sanders, 2016).

Naast deze twee doelen kunnen studenten of professionals portfolio's inzetten om aan te tonen dat de uitspraken die zij (tijdens sollicitaties) over hun capaciteiten doen aannemelijk zijn (Sanders, 2016). Dit heet een presentatieportfolio, maar deze is niet bedoeld om te toetsen.

Portfolio's voor summatieve toetsing dienen betrouwbaar en valide te zijn

Summatieve toetsen moeten aan verschillende kwaliteitscriteria voldoen om geschikt te zijn voor examinering. Twee belangrijke criteria voor summatieve toetsen zijn betrouwbaarheid en validiteit (Sanders, 2016).

Een toets is betrouwbaar als de beoordelingen accuraat zijn of als herhaalde metingen dezelfde beoordelingen opleveren (Van Berkel et al., 2017). Ook als verschillende beoordelaars tot eenzelfde oordeel komen is de betrouwbaarheid hoog. Er is dan sprake van een hoge interbeoordelaarsbetrouwbaarheid.

Validiteit heeft te maken met de betekenis, de bruikbaarheid en de juistheid van de conclusies die opleidingen uit scores trekken (Van Berkel et al., 2017). Een toetsinstrument moet immers meten wat die moet meten.

Er zijn verschillende soorten validiteit. Zo is er de mate waarin de inhoud van een toetsinstrument representatief is voor wat men wil meten (inhoudsvaliditeit) en de mate waarin een toetsinstrument het begrip of construct meet dat men beoogt te meten (begripsvaliditeit). Het is bijvoorbeeld niet valide om studenten hun inzicht te toetsen door hen te vragen naar een definitie die zij uit het hoofd hebben geleerd (Van Berkel et al., 2017).

Overigens zijn er verschillende visies over kwaliteitscriteria voor summatieve toetsen, zoals van portfolio's. Zo noemt Sanders (2016) bijvoorbeeld nog dat, naast betrouwbaar en valide, summatieve toetsen in het mbo transparant moeten zijn, dat ze qua inhoud en moeilijkheidsgraad overeen moeten komen met de beroepspraktijk (authenticiteit), en dat kwaliteitseisen vooraf vastgesteld moeten zijn (standaardisatie). Bij deze Kennisrotonde-vraag gaat het om de criteria betrouwbaarheid en validiteit.

Portfolio's zijn niet altijd betrouwbaar en valide

Onderzoek naar portfolio's en hun beoordelingen in het onderwijs levert geen helder inzicht op in de mate waarin ze betrouwbaar en valide zijn. Soms zijn de portfolio's wel betrouwbaar of valide, in andere gevallen weer niet. Het is daarom aannemelijk te veronderstellen dat de validiteit en betrouwbaarheid van portfolio's afhankelijk is van de manieren waarop opleidingen ze samenstellen en beoordelen. Daarmee is het waarborgen en vaststellen van de betrouwbaarheid en validiteit van portfolio's iets voor opleidingen zelf (Gadbury-Amyot et al., 2014)

Betrouwbaarheid

Zo komt er uit de literatuur geen eenduidig beeld naar voren of (de beoordeling van) portfolio's aan het kwaliteitscriterium betrouwbaarheid voldoen. Sanders (2016) concludeert bijvoorbeeld dat portfoliobeoordelingen binnen het mbo er niet aan voldoen. Andere onderzoeken beamen dat de betrouwbaarheid van portfoliobeoordelingen een heikel punt is (Johnston, 2004; Roberts et al., 2014). Maar er is ook onderzoek dat laat zien dat verschillende taken die studenten opnemen in hun portfolio wel een betrouwbare vorm van toetsing kunnen zijn (e.g., Bok et al., 2018). Dit onderzoek vond plaats bij 972 Nederlandse studenten diergeneeskunde.

Daarnaast laat onderzoek wisselende resultaten zien of beoordelaars tot eenzelfde beoordeling komen (interbeoordelaarsbetrouwbaarheid). Bijvoorbeeld Sanders (2016) concludeert dat de interbeoordelaarsbetrouwbaarheid van portfoliobeoordelingen onvoldoende is. Ook een review over het gebruik van portfolio's in medische opleidingen laat zien dat de beoordelaarsbetrouwbaarheid gemiddeld laag was (Driessen et al., 2007).

Daartegenover laten twee recentere onderzoeken zien dat de betrouwbaarheid tussen verschillende beoordelaars wel voldoende kan zijn (O'Brien et al, 2016; Favier et al., 2019). Dit bleek uit zowel onderzoek met 574 portfolio's van Nederlandse studenten diergeneeskunde (Favier et al., 2019) als uit onderzoek met 156 portfolio's van Amerikaanse studenten geneeskunde (O'Brien et al., 2016).

Validiteit

Het is vaak moeilijk aan te tonen of portfolio's valide zijn, omdat ze veelal beschrijvend van aard zijn en beoordelaars ze moeten interpreteren (Johnston et al. 2004). Daarnaast hangt de validiteit van een portfolio af van de context van een opleiding, zoals van de doelstellingen van specifieke opleidingen en hun onderwijsvisie (Gadbury-Amyot, 2014).

Roberts en collega's (2014) concludeerden bijvoorbeeld dat de portfoliobeoordelingen in een medische opleiding inhoudsvaliditeit hadden, omdat ze heel doelgericht voor de opleiding waren ontwikkeld om ermee te toetsen. Begripsvaliditeit werd in datzelfde onderzoek aangetoond doordat de beoordelingen een verschil lieten zien tussen studenten met verschillende ervaringsniveaus. Maar dit zegt iets over de validiteit van portfoliobeoordelingen in die opleiding en niet over validiteit van portfoliobeoordelingen in het algemeen.

In die opleiding – en ook in ander onderzoek waar de validiteit (en betrouwbaarheid) voldoende was - waren wel bepaalde manieren ingezet om ervoor te zorgen dat de portfoliobeoordelingen eraan voldeden.

Meerdere manieren voor versterking van betrouwbaarheid en validiteit van portfoliomethodiek

Opleidingen die bepaalde maatregelen treffen als zij werken met de portfoliomethodiek, blijken meer betrouwbaar en valide hun studenten te kunnen beoordelen. Er zijn enkele aanwijzingen dat de volgende maatregelen belangrijk zijn.

De inhoud van het portfolio bevat een mix aan toetstaken

Opleidingen die gebruikmaken van een mix aan taken of toetsen kunnen de betrouwbaarheid en validiteit van de summatieve toets versterken (Bok et al., 2018; Gulikers et al., 2014; Sluijsmans et al, 2008). Door deze mix kunnen opleidingen de competenties van studenten beter vaststellen (Baartman et al., 2017), zeker als studenten meerdere keren een mix aan authentieke taken in meerdere contexten moeten uitvoeren (Sluijsmans et al., 2008).

Hierbij is wel een aantal zaken belangrijk. Zo dient de mix aan taken uit een portfolio voorafgaand aan de toetsing ontwikkeld te zijn en de vastgestelde beoordelingsnormen te dekken (zie Baartman et al., 2017). De opleiding dient vooraf duidelijk te maken in welke mate individuele taken meewegen (Baartman et al., 2017; Schut et al, 2021). Maar er is hierbij geen 'magisch aantal beoordelingstaken' dat in een portfolio opgenomen moet worden (Schut et al., 2021).

Voor studenten dienen de doelen voor het werken met portfolio's helder te zijn. Zij moeten duidelijke informatie krijgen over alle toetsen, beoordelingscriteria, normering en uitslagen, zodat zij weten wat er van hen verwacht wordt (Van Berkel et al., 2017). Opleidingen dienen studenten ook goed voor te bereiden en in te lichten om met portfolio's te werken (Davis et al., 2001, Roberts et al., 2002; Tartwijk & Driessen, 2009). Wel dienen opleidingen voorzichtig te zijn om studenten te verplichten welke taken ze moeten opnemen in hun portfolio's (Schut et al., 2021). Hierdoor bestaat de kans dat het een checklist wordt en het zijn doel voorbijschiet.

Meerdere getrainde beoordelaars beoordelen op eenduidige wijze de portfolio's

Wanneer twee of meer getrainde beoordelaars onafhankelijk van elkaar de portfolio's beoordelen aan de hand van eenduidige beoordelingscriteria kan de betrouwbaarheid worden vergroot. Dit bleek uit onderzoek met meerdere beoordelaars (Driessen et al., 2007; Gadbury-Amyot et al., 2014; Oudkerk et al., 2018), met getrainde beoordelaars (Royal & Hecker, 2016; Schut et al., 2021; Sluijsmans et al., 2008), en wanneer eenduidige beoordelaarscriteria werden gebruikt (Roberts et al., 2002; Straetmans et al, 2006).

De rol van beoordelaar is nadrukkelijk anders dan die van mentor/docent. De mentor/docent ondersteunt de student in het maken van het portfolio en geeft studenten rijke en narratieve feedback. Sommige onderzoeken geven aan dat de rol van mentor en die van beoordelaar strikt gescheiden moeten zijn om met portfolio's betrouwbaar summatief te toetsen (Baartman, 2017; Gadbury-Amyot et al., 2014). Ander onderzoek ziet die rol niet als strikt gescheiden, maar geeft wel aan dat er een balans moet zijn tussen beide rollen (Driessen et al., 2005).

Portfolio's zijn vaak onderdeel van programmatisch toetsen

Bovenbeschreven inzichten leiden we deels af uit studies naar het gebruik van portfolio's binnen programmatisch toetsen. Programmatisch toetsen is een integrale benadering van toetsen waarbij de focus ligt op het optimaliseren van de leer- en beslisfunctie (Baartman et al., 2020).

Bij programmatisch toetsen volgt de opleiding de ontwikkeling van studenten via zogeheten datapunten die studenten opnemen in een portfolio. Voorbeelden van datapunten zijn opdrachten, presentaties, feedback en uitkomsten van kennistoetsen. Een enkel datapunt wordt niet gebruikt om een beslissing te

nemen, maar alleen om studenten feedback te geven. De datapunten uit het portfolio bij elkaar worden wel gebruikt om een (summatieve) beslissing te nemen (Van Berkel et al., 2017).

Een overzichtsstudie naar programmatisch toetsen bij universitaire zorgopleidingen liet een aantal positieve resultaten zien die erop wijzen dat een opleiding met portfolio's summatieve beslissingen kan nemen (Schut et al., 2021). In hoeverre die uitkomsten voor de mbo-sector gelden is nog onduidelijk.

Geraadpleegde bronnen

Baartman, L., & Gulikers, J. (2017). Assessment in Dutch vocational education: Overview and tensions of the past 15 years. *Enhancing Teaching and Learning in the Dutch Vocational Education System*, 245-266.

Baartman, L., van Schilt-Mol, T., & van der Vleuten, C. (2020). *Programmatisch toetsen: voorbeelden en ervaringen uit de praktijk*. Hogeschool Utrecht.

Bok, H. G. J., De Jong, L. H., O'Neill, T., Maxey, C. & Hecker, K. G. (2018). Validity evidence for programmatic assessment in competency-based education. *Perspectives on Medical Education*, 7 (6), p. 362 – 372.

Davis, M. H., Ben-David, M. F., Harden, R. M., Howie, P., Ker, J., McGhee, C., ... & Snadden, D. (2001). Portfolio assessment in medical students' final examinations. *Medical Teacher*, *23*(4), 357-366.

Driessen, E., Van Der Vleuten, C., Schuwirth, L., Van Tartwijk, J., & Vermunt, J. D. H. M. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical education*, *39*(2), 214-220.

Driessen, E., Van Tartwijk, J., Van Der Vleuten, C., & Wass, V. (2007). Portfolios in medical education: why do they meet with mixed success? A systematic review. *Medical education*, *41*(12), 1224-1233.

Favier, R. P., Vernooij, J. C., Jonker, F. H., & Bok, H. G. (2019). Inter-rater reliability of grading undergraduate portfolios in veterinary medical education. *Journal of Veterinary Medical Education*, *46*(4), 415-422.

Gadbury-Amyot, C. C., McCracken, M. S., Woldt, J. L., & Brennan, R. L. (2014). Validity and reliability of portfolio assessment of student competence in two dental school populations: a four-year study. *Journal of Dental Education*, *78*(5), 657-667.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Design*, *52*(3), 67–85.

Johnston, B. (2004). Summative assessment of portfolios: an examination of different approaches to agreement over Outcomes. *Studies in higher education*, 29(3):395-412

O'Brien, C. L., Sanguino, S. M., Thomas, J. X., & Green, M. M. (2016). Feasibility and outcomes of implementing a portfolio assessment system alongside a traditional grading system. *Academic Medicine*, *91*(11), 1554-1560.

Oudkerk Pool, A., Govaerts, M.J.B., Jaarsma, D.A.D.C. *et al.* (2018). From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv in Health Sci Educ, 23*, 275–287.

Roberts, C., Shadbolt, N., Clark, T., & Simpson, P. (2014). The reliability and validity of a portfolio designed as a programmatic assessment of performance in an integrated clinical placement. *BMC medical education*, *14*(1), 1-11.

Roberts, C., Newble, D. I., & O'Rourke, A. J. (2002). Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? *Medical education*, *36*(10), 899-900.

Royal, K. D., & Hecker, K. G. (2016). Rater errors in clinical performance assessments. *Journal of veterinary medical education*, 43(1), 5-8.

Sanders, P. (2016). *Toetsen op School Middelbaar beroepsonderwijs.* Cito.

Schut, S., Maggio, L. A., Heeneman, S., van Tartwijk, J., van der Vleuten, C., & Driessen, E. (2021). Where the rubber meets the road—An integrative review of programmatic assessment in health care professions education. *Perspectives on medical education*, *10*(1), 6-13.

Sluijsmans, D. M. A., Straetmans, G. J. J. M., & van Merriënboer, J. J. G. (2008). Integrating authentic assessment with competence-based learning in vocational education: The protocol portfolio scoring. *Journal of Vocational Education & Training*, *60*, 159–172.

Straetmans, G., Sluijsmans, D., Bolhuis, B. & Van Merriënboer, J. (2003). Integratie van instructie en assessment in competentiegericht onderwijshttps://www.sluijsmans.net/wp-content/uploads/2019/01/TvHO-artikel-Straetmans-Sluijsmans.pdf *Tijdschrift voor Hoger Onderwijs*, *3*, 171-197.

Van Berkel, H., Bax, A., & Joosten-ten Brinke (2017). *Toetsen in het Hoger Onderwijs.* Managementboek.

Van Tartwijk, J., & Driessen, E. W. (2009). Portfolios for assessment and learning: AMEE Guide no. 45. *Medical teacher*, *31*(9), 790-801.

Willemse, P, & Van der Neut, I. (2020). *Anders verantwoorden van de diplomabeslissing in het mbo.* IVA onderwijs.

Meer weten?

Canon beroepsonderwijs (2021). *Examineren in het mbo. Een diploma van waarde.* Canon Beroepsonderwijs.

Kennisrotonde (2019). *Wat is een correcte toetsconstructie op het gebied van verwerving van taal-, reken- en digitale vaardigheden voor volwassenen met gebruikmaking van de portfoliomethodiek? (KR 717).* Kennisrotonde.

Lerend Netwerk Kennispunt Onderwijs & Examinering MBO

Over deze vraag

Opgesteld door:	Maartje Kouwenberg en Christa Teurlings (Kennismakelaars Kennisrotonde)
Vraagsteller:	expert docent mbo instelling
Onderwijssector:	MBO
Trefwoorden:	Beoordelen, portfolio, examineren, formatief en summatief toetsen.
Referentie:	Kennisrotonde. (2022). *Kan met de portfoliomethodiek betrouwbaar en valide summatief getoetst worden in het mbo?* (KR. 1381)

Dit antwoord is gepubliceerd op Kennisrotonde.nl. De Kennisrotonde is samen met NCO en Onderwijskennis een dienst van het NRO.