

Over toetskwaliteit in het onderwijs

Huh, moesten we dat ook leren?

Toetsing is aan de orde van de dag. De uitslagen van toetsen kunnen zeer bepalend zijn voor leerlingen. Het is dus belangrijk dat toetsen van goede kwaliteit zijn. Maar wat is goede toetskwaliteit? En hoe kunt u als leraar zorgen voor een goede toetskwaliteit?

TEKST: NATHALIE MAASSEN EN DORIEN DEN OTTER

We kunnen zeggen dat er in Nederland 16 miljoen deskundigen zijn op toetsgebied. Iedereen heeft ervaring met toetsen en een beeld van wat goede en slechte toetsen zijn (Eggen, 2009). Zoveel mensen als er zijn, zoveel meningen bestaan er over het begrip kwaliteit. Het is belangrijk om meer inzicht te krijgen in het begrip toetskwaliteit, want voor leerlingen hangt er veel van af: bijvoorbeeld of ze overgaan naar de volgende klas. Daarnaast gebruiken leraren toetsresultaten voor de precieze invulling van het onderwijsleerproces. Er is dan ook veel onderzoek gedaan naar toetskwaliteit. Dit onderzoek hebben wij geordend. We hebben hiervoor systematisch 242 wetenschappelijke artikelen, artikelen uit praktijkvakbladen en boeken over toetskwaliteit bestudeerd. Resultaten van deze verschillende artikelen komen grotendeels overeen met wat iedere leraar eigenlijk wel weet. Het is goed om te zien dat het literatuuronderzoek dit bevestigt. In dit artikel leest u de belangrijkste resultaten aan de hand van herkenbare uitspraken van leerlingen.

'Hij krijgt vast een hoger cijfer omdat u hem aardig vindt.'

Uit het onderzoek blijkt dat betrouwbaarheid van een toets vaak wordt genoemd, vooral bij het afnemen en beoordelen van toetsen. Hoe kunt u als leraar zorgen voor een betrouwbare toets? Een belangrijk aspect is het objectief nakijken van een toets (o.a. Grainger, Purnell & Zipf, 2008). Het kan verleidelijk zijn om bepaalde leerlingen sneller voldoende te beoordelen omdat u hun dat nou eenmaal gunt vanwege hun grote inspanning. Op dat moment bent u echter niet meer objectief. Manieren om objectiviteit te bereiken zijn (o.a. Rom, 2011):

- het gebruik van een beoordelingsschema;
- het nakijken van de toets per vraag in plaats van per leerling;
- de inzet van een tweede beoordelaar.

'Telt deze toets mee voor ons rapport?'

Het doel van de toets blijkt bepalend te zijn voor de kwaliteitseisen van een toets. In veel gevallen zult u de leerling willen toetsen om een eindbeoordeling van zijn of haar niveau te

geven (summatief toetsen). Bij summatieve toetsen bepaalt u nauwkeurig de normering en cesuur: wanneer is een leerling geslaagd? Het kan echter ook zo zijn dat u wilt weten welke onderdelen van de leerstof een leerling beheerst en aan welke onderdelen meer aandacht moet worden besteed (formatief toetsen). Bij formatieve toetsen is het belangrijk dat u feedback geeft aan leerlingen en uw onderwijs afstelt op de toetsresultaten.

'Ik dacht dat we wel een rekenmachine mochten gebruiken!'

Soms komen leerlingen tijdens een toets voor vervelende verrassingen te staan. Misschien hebben ze tijdens de instructie niet goed opgelet, maar het kan ook gebeuren dat u iets vergeten bent te vermelden. In een toetsituatie spreken we van transparantie als leerlingen alle informatie hebben die nodig is om zich optimaal te kunnen voorbereiden op de toets en de toets te kunnen maken (o.a. Strijbos, 2013). Dit blijkt een veelgenoemd aspect van toetskwaliteit te zijn. Stel uw leerlingen dus op de hoogte van de inhoud, vorm en het moment van de toets. Geef ook informatie over de manier van beoordelen: hoeveel punten krijgt een leerling bij een goed antwoord? En geeft u minpunten bij spelfouten?

'We krijgen alleen maar meerkeuzetoetsen.'

Bij een luistertoets is het mogelijk meerkeuzevragen te gebruiken, terwijl dat bij een spreektoets juist niet goed kan. De vraagvorm moet passen bij de vaardigheid die u meet. Als u wilt weten of iemand goed kan presenteren of een muur kan metselen, dan is een praktische opdracht het best. Wilt u kennis meten, dan is een schriftelijke toets juist wel vaak prima. Laat het toetsdoel dus bepalend zijn voor de keuze van de toetsvorm (o.a. Ter Beek & Rijken, 1998). Op deze manier meet u wat u wilt weten. Dit noemen we validiteit (o.a. Kane, 2004): een kwaliteitsaspect dat in het onderzoek veel voorkwam.

'De toets was echt veel moeilijker dan normaal...'

Toetsvragen bestaan in alle soorten en maten. Een veelge-



bruikte indeling in niveaus is kennis-, toepassing- en inzichtvragen. Hulpwerkwoorden helpen u om vragen op de gewenste niveaus te stellen (Van Berkel & Draaijer, 2011), zie tabel. Met behulp van een toetsmatrijs kunt u er vervolgens voor zorgen dat op alle gewenste cognitieve niveaus wordt getoetst. Bovendien garandeert u met een toetsmatrijs dat de toets dekkend is voor de te toetsen stof en dat er geen vraag buiten de leerstof valt (o.a. Ter Beek et al., 1998). Een relevante checkvraag is: kan de leerling die maar een klein gedeelte van de stof beheerst een voldoende halen? Als u deze vraag met 'nee' kunt beantwoorden, dan heeft de toets een grotere selecterende functie.

Niveau	Voorbeelden van hulpwerkwoorden
Kennis	Benoemen, beschrijven, uitleggen
Toepassing	Berekenen, demonstreren, oplossen
Inzicht	Bekritisieren, vergelijken, interpreteren

Tot slot

Wij hopen dat de uitspraken van leerlingen hebben laten zien dat u onbewust bijna dagelijks met toetskwaliteit in aanraking komt. Ook hopen wij dat dit artikel u een aantal aangrijpingspunten geeft om rekening mee te houden bij het maken, afnemen en beoordelen van toetsen.

Wilt u meer weten?

Kijk op www.nro.nl voor:

- De wetenschappelijke rapportage of de publieksbrochure van het onderzoek.
- Het hulpmiddel 'Eerste Hulp Bij Toetsen', uitgebracht voor leraren in het voortgezet onderwijs. Dit boekje behandelt kort en bondig belangrijke onderwerpen over toetskwaliteit. ●

► **Nathalie Maassen, MSc.** en **Dorien den Otter, MSc.** hebben dit onderzoek vanuit het RCEC (Universiteit Twente) uitgevoerd.

- Beek, A. ter, & Rijken, R. (1998). Meetinstrument Kwaliteit Schooltoetsen. 's-Hertogenbosch: KPC Onderwijs Innovatie Centrum.
- Berkel, H. J. M. van, & Draaijer, S. (2011). Gids voor toetsontwikkeling. *Examens*, 1, 1-8.
- Eggen, T. J. H. M. (2009). De kwaliteit van toetsen. *Oratie Universiteit Twente*, 9 april 2009.
- Grainger, P., Purnell, K., & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33(2), 133-142. doi:10.1080/02602930601125681
- Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2(3), 135-170. doi:10.1207/s15366359mea0203_1
- Rom, M. C. (2011). Grading more accurately. *Journal of Political Science Education*, 7(2), 208-223. doi:10.1080/15512169.2011.564916
- Strijbos, J. (2013). Een kwaliteitsborgingssysteem van toetsing: Van beleid naar praktijk. *Examens*, 3, 22-26.