

Wanneer is een toets goed genoeg?

Een lijst met 47 (zelf)evaluatievragen

Bron: Robert Coe <https://www.cem.org/blog/would-you-let-this-test-into-your-classroom/>

Vertaling: Dominique Sluijsmans | sluijsmans.net

Thema	Vraag	Hoe ziet de eigen praktijk uit?
1. Constructdefinitie	1. Wat beoogt de toets te meten (het construct, het doel)?	
	2. Wat zeggen de uitkomsten (scores, cijfers, niveau)? Hoe kunnen ze worden geïnterpreteerd? Kun je kwalitatief beschrijven wat het verschil is tussen een student met een hoge score en een student met een lage score?	
	3. Welke beslissingen mag je nemen op basis van de uitkomsten van de toets (maar welke ook niet)?	
	4. Als er sprake is van subschalen binnen de toets, hoe moeten deze worden geïnterpreteerd?	
	5. Als er prestatiedrempels zijn gespecificeerd (zak/slaaggrens; onvoldoende, voldoende, goed), is het helder wat deze betekenen en hoe deze zijn bepaald?	
	6. Hoe helder zijn de interpretaties en het gebruik van de toetsresultaten? Geeft de toetsontwerper expliciete begeleiding in welke interpretaties en beslissingen worden ondersteund?	
	7. Is er passende begeleiding bij voorkomende, maar onbedoelde en onjuiste interpretaties van scores?	
	8. Voor welke studentpopulaties of contexten zijn de interpretaties en gebruik van toetsresultaten geschikt?	
2. Constructvaliditeit	9. Hoe verhoudt het te toetsen construct zich tot andere toetsen/constructen, volgens de theorie?	
	10. In welke mate worden deze verwachte relaties (of het gebrek eraan) ondersteund door bewijs?	

Thema	Vraag	Hoe ziet de eigen praktijk uit?
	11. Hoe goed correspondeert de toets met toetsen gericht op dezelfde of gerelateerde constructen?	
3. Inhoudsvaliditeit	12. Zijn de toetsitems/taken geschikt voor het te toetsen construct?	
	13. Is de structuur van de toets/toetstaken geschikt voor het te toetsen construct (toetslengte, balans in verschillende toetsmodaliteiten)?	
	14. Is de wijze van beoordelen/scoren geschikt?	
	15. Vallen de te beoordelen aspecten binnen het beoogde construct?	
	16. Kunnen de toetsresultaten worden beïnvloed door onechte karakteristieken/vaardigheden?	
	17. Dekt de toets de volledige reikwijdte van het beoogde constructdomein, zowel in termen van inhoud als niveau? Zijn er zaken die niet worden gevat in de toetsresultaten?	
4. Criteriumvaliditeit	18. Wat voorspelt (als het al iets voorspelt) het toetsresultaat?	
	19. Hoe goed voorspelt het toetsresultaat latere prestaties (bijvoorbeeld op landelijke, gestandaardiseerde toetsen)?	
	20. Hoe sterk hangen toetsscores samen met andere metingen over hetzelfde construct?	
	21. Voor welke studentpopulaties/contexten is deze samenhang aangetoond? Met welk tijdsinterval tussen toetsen? Komt dit overeen met vraag 8?	
5. Betrouwbaarheid	22. Geven herhaalde afnames van de toets consistente resultaten? Welke <i>test-hertest</i> correlaties zijn bekend? Bij welke studentpopulaties, contexten, steekproeven en tijdsintervallen tussen toetsen? Komt dit overeen met vraag 8?	
	23. In welke mate is het toetsresultaat afhankelijk van de specifieke toetsitems? Welke <i>interne consistentie</i> waarden zijn bekend (bijvoorbeeld Cronbach's alpha)? Bij welke studentpopulaties, contexten, steekproeven? Komt dit overeen met vraag 8?	
	24. Als toetsresultaten (deels) afhankelijk zijn van de beoordelaar(s), in welke mate zijn deze dan afhankelijk van wie beoordeelt? Welke <i>inter-</i>	

Thema	Vraag	Hoe ziet de eigen praktijk uit?
	<i>ratercorrelaties</i> zijn bekend? Bij welke studentpopulaties, contexten, steekproeven? Komt dit overeen met vraag 8?	
	25. Welke zaken kunnen voor ruis in de scores hebben gezorgd (bijvoorbeeld de toetsafname, de selectie van items, de beoordelaar, etc.)? Is de relatieve grootte van hun bijdrage aan de onzekerheid in scores nauwkeurig beoordeeld?	
	26. Met welke precisie geven de toetsresultaten een inschatting van het onderliggende construct? Zijn bijvoorbeeld standaardafwijkingen/meetfouten bekend? Bij welke studentpopulaties, contexten, steekproeven? Komt dit overeen met vraag 8?	
	27. Als er prestatieniveaus worden gebruikt voor het nemen van beslissingen, maakt de toets dan een adequaat onderscheid tussen deze niveaus? Welke deel van de studenten wordt juist ingedeeld?	
6. Vrij van bias	28. Is er kans dat de toets voor sommige studenten tot gunstigere resultaten kan leiden dan andere studenten (bijvoorbeeld <i>geslacht</i> , <i>SES</i> , <i>etniciteit</i>)?	
	29. Worden toetsresultaten beïnvloed door factoren die niks te maken hebben met het te beoordelen construct (bijvoorbeeld <i>leesvaardigheid</i> , <i>culturele kennis</i> , <i>taalkennis</i> , <i>houdingen/persoonlijkheidstrekken</i>)?	
	30. Is er een overzicht van dergelijke factoren opgesteld (zie vraag 29)? Welk bewijs is voorhanden om te bepalen of toetsresultaten door deze factoren worden beïnvloed?	
7. Reikwijdte	31. Voor welke reikwijdte (leeftijd, vaardigheden, etc.) is de toets geschikt?	
	32. Is de toets vrij van plafond- (iedereen haalt het) en vloereffecten (niemand haalt het)?	
	33. Hoe varieert de nauwkeurigheid van de toets over de hele reikwijdte?	
8. Robuustheid	34. Zijn er mogelijkheden voor studenten (of anderen) fraude te plegen/te spieken?	

Thema	Vraag	Hoe ziet de eigen praktijk uit?
	35. Als de toets wordt gebruikt voor zwaarwegende beslissingen, is het dan mogelijk dat studenten een score behalen die niet het te beoordelen construct reflecteert (zoals gedefinieerd in thema 1)?	
	36. Leidt het verbinden van zwaarwegende consequenties aan de toetsresultaten tot een verandering in de antwoorden op eerdere vragen?	
	37. Als de toets wordt gebruikt als een uitkomstmaat in een evaluatie, is de toets dan 'objectief', in de zin dat deze niet kan worden beïnvloed door de verwachtingen of wensen van een 'ziende' beoordelaar (i.e. iemand die op de hoogte is welke studenten de toets hebben gehad)?	
9. Waarde voor leren	38. Heeft het proces van toetsing, of de feedback die het genereert, directe waarde voor docenten en studenten?	
	39. Wordt de toets als positief ervaren door de betrokkenen (studenten, docenten, ouders)?	
10. Toegankelijkheid	40. Is de toets toegankelijk voor studenten met een functiebeperking (bijvoorbeeld visuele beperkingen, doofheid, dyslexie, andere fysieke beperkingen)?	
	41. Zijn er adequate voorzieningen (inclusief adequate begeleiding) hiervoor beschikbaar?	
11. Administratieve eisen	42. Hoeveel tijd vraagt de toets van de student?	
	43. Is er extra tijd nodig om de toets klaar te zetten?	
	44. Moet de toets worden afgenomen/de afname worden gesurveilleerd door een gekwalificeerd/getraind iemand?	
	45. Is technologie nodig (bijvoorbeeld computers, toetslokaal, andere technologische middelen)?	
	46. Kan de toets worden afgenomen bij groepen of alleen bij individuele studenten?	
	47. Moeten de antwoorden worden becijferd? Hoeveel tijd is hiervoor nodig?	