

Validering en betrouwbaarheid

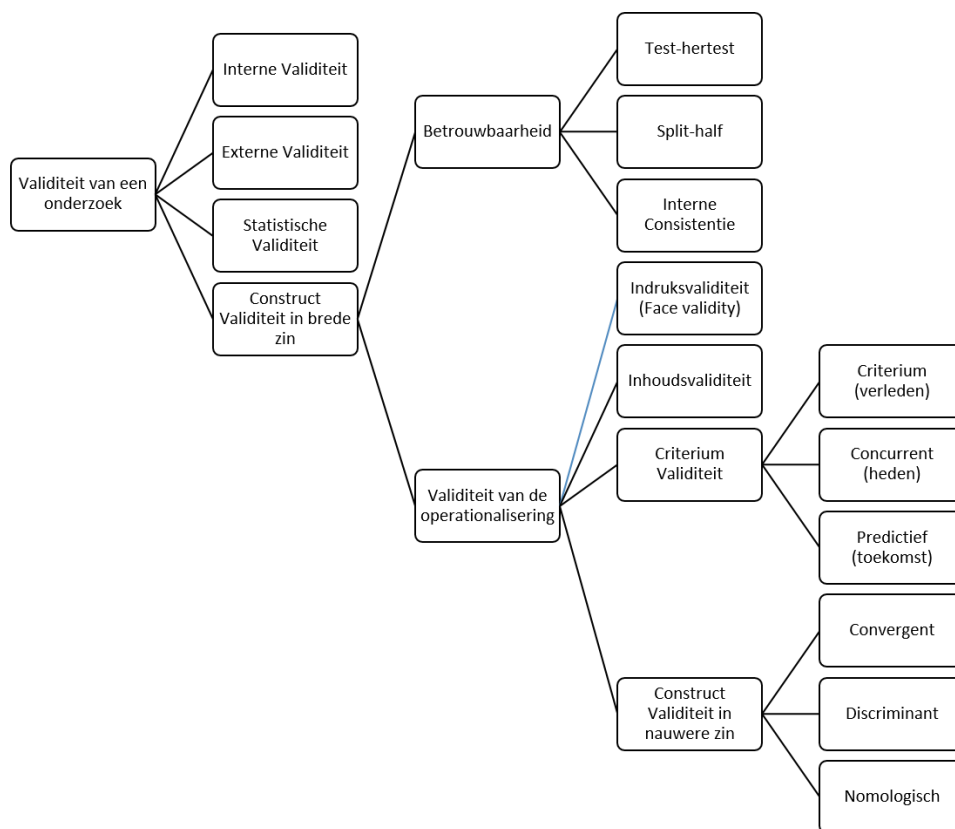
Inleiding

De kwaliteit van een sociaalwetenschappelijk onderzoek berust in de basis op twee pijlers: de validiteit en de betrouwbaarheid. Validiteit verwijst naar de accuratesse van een onderzoek. Een valide onderzoek geeft een goed antwoord op de vraag: “meet je wat je wilt meten?”.

Betrouwbaarheid verwijst naar de consistentie van de resultaten. Een betrouwbaar onderzoek geeft een voldoende antwoord op de vraag: “meet je steeds hetzelfde?”.

Er zijn meerdere aspecten toe te kennen aan zowel de betrouwbaarheid als de validiteit van een onderzoek. Vaak zijn de aspecten ook onderling verbonden. Zo betekent bijvoorbeeld een vermindering van de split-half betrouwbaarheid, dat de resultaten minder generaliseerbaar zijn, wat betekent dat de externe validiteit lager uit valt.

Hieronder worden de verscheidene validiteitsvormen kort behandeld.



Interne validiteit

De interne validiteit is de mogelijkheid van een onderzoek om antwoord te geven op de onderzoeksvraag. Hoe sterker de interne validiteit, hoe zekerder men is van de antwoord(en) op de onderzoeksvraag. Veelal zal dit verwijzen naar de mogelijkheid van een onderzoek om een causaal verband aan te tonen. Wanneer een onderzoek pretendeert een causaal verband aan te kunnen tonen, maar slechts correlaties kan overleggen, wordt dit onderzoek als intern niet-valide beoordeeld.

Externe validiteit

De externe validiteit is het vermogen van een onderzoek om de resultaten te generaliseren. Onderzoeken worden meestal op basis van steekproeven uitgevoerd. De externe validiteit verwijst naar de mogelijkheid om resultaten op basis van een steekproef te vertalen (lees: generaliseren) naar de populatie. Bijvoorbeeld: als een steekproef te klein is, dan kan deze niet gebruikt worden om de resultaten ook voor de totale populatie te omschrijven. In een dergelijk geval is er sprake van onderzoek dat extern niet-valide is.

Statistische validiteit

De statistische validiteit verwijst naar het correct gebruik van alle statistische methoden. Dit behelst de keuze van de juiste analysestrategie in de juiste situatie. Ook dienen de juiste conclusies te worden getrokken uit een analyse.

Constructvaliditeit (in brede zin)

De constructvaliditeit in brede zin bekijkt de precisie van een meetinstrument. Deze vorm van validiteit behelst het idee dat het meetinstrument moet meten wat het wil meten. Een vragenlijst die allerlei vragen bevat over de depressieve kenmerken van een patiënt, kan niet worden gebruikt als intelligentietest. Wanneer dit het geval is, is er in de brede zin sprake van lage constructvaliditeit. Hieraan zijn een aantal verschillende aspecten te verbinden, welke hieronder verder uiteengezet worden.

Betrouwbaarheid

De betrouwbaarheid verwijst naar de consistentie van de resultaten die het meetinstrument levert. Als het instrument op twee verschillende momenten wordt gebruikt, dient er zoveel mogelijk eenzelfde antwoord te worden gevonden. Er zijn drie manieren om dit te waarborgen.

- Test-hertest

De test-hertest betrouwbaarheid wordt gemeten door bij dezelfde groep respondenten dezelfde test tweemaal af te nemen. Idealiter zijn alle resultaten gelijk van beide tests. In de praktijk zal dit vaak niet zijn, waardoor wordt gekeken naar de correlatiecoëfficiënt r tussen de eerste en tweede meting. Hoe hoger de correlatie, hoe sterker test-hertest betrouwbaarheid.

- Split half

De split-half betrouwbaarheid meet de betrouwbaarheid binnen een vragenlijst. Dit wordt gedaan door de test in twee (idealiter gelijke) delen te verdelen. De correlatie tussen de eerste helft en de tweede helft wordt berekend en deze bepaalt de split-half betrouwbaarheid.

- Interne Consistentie (Cronbach's Alfa)

Interne consistentie is de maat waarmee kan worden geïndexeerd in hoeverre gedurende de gehele vragenlijst vergelijkbare antwoorden worden gegeven. Hoe hoger de coëfficiënt hiervoor, hoe sterker de interne consistentie en dus betrouwbaarheid van een vragenlijst. Veel voorkomende maat om dit te meten is de Cronbach's α .

Inhoudsvaliditeit

Bij inhoudsvaliditeit gaat het om hoeverre de inhoud van de test het gehele kennis-, gedrags-, of vaardigheidsdomein representeert. De vraag "meet mijn instrument alles?" onderzoekt de inhoudsvaliditeit. Als bijvoorbeeld een tentamen inhoudsvalide is, dekken de vragen de gehele stof.

De vragen in het tentamen zijn in essentie een steekproef van alle mogelijke vragen die je kan stellen over de stof. Deze steekproef van vragen dient representatief te zijn voor de gehele populatie aan vragen. Deze representativiteit is de mate van inhoudsvaliditeit.

De bepaling van de inhoudsvaliditeit valt onder te verdelen in een viertal fases.

- 1e fase. Definiëren van het domein.
- 2e fase. Selectie van deskundigen
- 3e fase. Deskundigen matchen items in het domein
- 4e fase. Resultaten van matching worden geanalyseerd.

Criteriumvaliditeit

Criteriumvaliditeit geeft de relatie tussen testcores en een concreet criterium. Dit is een gedrag of prestatie dat niet met de test wordt gemeten, maar wel data van beschikbaar is. Het gaat hierbij om een concreet criterium. Dit zijn aantallen en voorkomingen van bepaalde soorten gedrag en prestaties. Denk hierbij aan aantal kinderen, aantal glazen water per dag etc. Hierbij kunnen de testcores vergeleken worden met een criterium in het verleden, heden (concurrente validiteit) of toekomst (predictieve validiteit).

Men meet de criteriumvaliditeit met een correlatiecoëfficiënt. Deze correlatie noemt men de (criterium)validiteitcoëfficiënt van de test. Hoe hoger de correlatie, des te hoger is de validiteit van de test. Inhoudelijk hoeft de test niet verwant te zijn met het criterium. Zo kan bijvoorbeeld een theoretische test (bijv. theoretisch rijexamen) een goede voorspeller zijn voor actief gedrag (bijv. rijvaardigheid).

Voor het beoordelen van de criteriumvaliditeit gaat men na of theoretisch veronderstelde (sterke) samenhang ook in geobserveerde gegevens wordt teruggevonden. Een goede voorspelling of schatting van een criterium op basis van testcores is mogelijk bij een sterk verband tussen de testcores en het criterium. Bij een geobserveerd zwak verband tussen testscore en criterium beoordeelt men de test als niet criteriumvalide.

Constructvaliditeit in nauwere zin

De constructvaliditeit in nauwere verwijst naar de correcte meting van het bedoelde begrip door het opgestelde instrument. In het geval van Kwaliteitscholen gaat het hierbij om de verschillende vragenlijsten. Er worden hieronder twee verschillende vormen van constructvaliditeit onderscheiden:

- Convergent

Convergente validiteit geeft aan in hoeverre verschillende instrumenten om hetzelfde begrip te meten, hetzelfde opleveren. Wanneer men zelf een instrument opstelt om depressie te meten, moeten de resultaten overeenkomen met die van reeds gevalideerde instrumenten zoals Beck Depression Inventory.

- Discriminant

Dit is tegenovergesteld van discriminant validiteit. Een instrument moet het begrip dat het pretendeert te meten ook onderscheiden van andere begrippen. Zo moet een zelf opgesteld instrument op depressie te meten, zich kunnen onderscheiden van een instrument pannenkoekenvoorkeur meet.

De constructvaliditeit in nauwere zin wordt in kaart gebracht aan de hand van Principiële Componenten Analyse (PCA), Exploratieve Factor Analyse (EFA) en/of Confirmatieve Factor Analyse (CFA).

Oppervlaktevaliditeit (face validity)

Oppervlaktevaliditeit is een wat vreemde vorm van validiteit, doch zeker nuttig. Dit is de vorm van validiteit die wordt beoordeeld op het eerste gezicht. Dit bevat een aantal trivialiteiten en onderdelen die niet onder de andere vormen van validiteit vallen. Soms wordt nog naar deze vorm van validiteit verwezen als gezond verstand. Voorbeeld ter illustratie is van een onderzoeker die een wetenschappelijk zeer sterk onderbouwde vragenlijst heeft opgesteld om ADHD te meten onder jonge kinderen. De vragenlijst is op basis van uitvoerig onderzoek zorgvuldig geselecteerd en heeft een zeer hoge betrouwbaarheid. Dit lijkt een zeer bruikbare vragenlijst, maar wanneer het drie uur kost om de vragenlijst in te vullen, is het nog maar de vraag of deze wel echt valide is.