

Oefenvragen bij Hoofdstuk 8

1. Leg uit waarom niet in absolute zin gesproken kan worden van de validiteit van een test.
2. Dienen medische procedures ook een valideringsonderzoek te ondergaan? Beargumenteer uw antwoord.
3. Hoe zou men predictieve validering kunnen opvatten als onderdeel van begripsvalidering?
4. Waarom geven de auteurs van het boek toch de voorkeur aan gescheiden behandeling van predictieve validiteit en begripsvaliditeit?
5. Wat is inhoudsvaliditeit?
6. Noem een zwak punt van inhoudsvaliditeit.
7. Leg uit wat het effect is van de betrouwbaarheid van een criterium op de voorspelling van dit criterium met een test.
8. Bepreek het probleem dat een te voorspellen criterium verschillende graden van abstractie kan hebben, en hoe het komt dat de beter voorspelbare aspecten van een criterium vaak ook de minder relevante zijn.
9. Bedenk wat het uiteindelijke criterium bij voorspelling met de CITO Eindtoets Basisonderwijs zou kunnen zijn. Doe eveneens voorstellen voor een tussentijds en een onmiddellijk criterium.
10. Met welk doel trekt men een gestratificeerde steekproef?
11. Het komt nogal eens voor dat tests gevalideerd op gemakkelijk voorhanden zijnde steekproeven, zoals psychologiestudenten. Wanneer is dit geen probleem voor de validiteit?
12. Noem twee redenen waarom items meestal niet op basis van hun moeilijkheid in een test of vragenlijst zouden moeten worden geselecteerd.
13. Welke items dient met in een test of vragenlijst te selecteren om een testscore met een hoge betrouwbaarheid te verkrijgen?
14. In welke situatie zoekt met vooral zijn toevlucht tot 'concurrent validity' ter vervanging van predictieve validiteit?
15. Wat zijn in een onderzoek ter bepaling van de predictieve validiteit meestal de oorzaken van de geringere representativiteit van de steekproef die tot 'restriction of range' leiden?
16. Wat is de contaminatie van het criterium? Geef een voorbeeld. Welke invloed heeft dit verschijnsel op de validiteit van een selectieprocedure?
17. Voordat men de correlatie tussen een testscore en een criteriumscore bepaalt, dient eerst de 'scatter plot' van de empirisch gevonden relatie te worden bestudeerd. Waarom is dit nodig?
18. Wat is een heteroscedastische relatie? Is de product-momentcorrelatie geschikt om deze relatie te beschrijven? Beargumenteer uw antwoord.
19. Wat is een suppressorvariabele? Geef een ander inhoudelijk voorbeeld dan in dit hoofdstuk wordt gegeven.
20. Wat is een moderatorvariabele? Geef ook hier inhoudelijk voorbeeld, anders dan in dit hoofdstuk wordt gegeven.

21. Beschrijf voor de predictoren die uiteindelijk in een meervoudig regressiemodel worden opgenomen de onderlinge correlaties en de correlaties met het criterium.
22. Noem de vijf mogelijke oorzaken van doorgaans lage validiteitscoëfficiënten.
23. Waarom zien de auteurs weinig heil in het blind selecteren, dus alleen op basis van correlaties (of andere statistische gegevens), van de beste voorspellende tests in een voorspellingsmodel ten behoeve van een bepaald criterium?
24. Wat is een nomologisch netwerk? Welke rol spelen tests hierin?
25. Welke rol spelen confirmerende validiteit en discriminante validiteit bij het onderzoek van hypothesen?
26. Wat is de multitrek-multimethodematrix? Hoe geeft deze matrix informatie over zowel betrouwbaarheid als validiteit?

Antwoorden Oefenvragen Hoofdstuk 8

1. Er kan niet in absolute zin gesproken worden over validiteit, omdat het begrip nooit een eenduidige betekenis heeft gehad. Dit heeft ertoe geleid dat men de term te pas en te onpas, en in vele niet nadere aangeduide betekenissen dooreen gebruikt.
Betekenissen als juistheid, nuttigheid, accuraatheid in het gebruik, interpreteerbaarheid, al of niet gecorrigeerd voorspellend vermogen, en waarde bij beslissingen.
2. De vraag naar validiteit kan gesteld worden met betrekking tot ieder proces van concluderen tot andere dan waargenomen feiten en gebeurtenissen. In de geneeskunde is het bijvoorbeeld de vraag in hoeverre iemands bloeddruk iets zegt over de kans dat hij of zij een hartkwaal ontwikkelt. Niet die bloeddruk zelf – zeg maar, het testgedrag – is van belang, maar de mate waarin het iets zegt over andere verschijnselen.
3. De relatie tussen de predictor en het criterium, alsmede de operationalisering van de onderliggende begrippen, vormen een onderdeel van het theoretisch netwerk van het door de test geoperationaliseerde begrip.
4. De reden daarvoor is dat vooral in de context van het onderwijs en de personeelspsychologie op grote schaal gebruik gemaakt wordt van het voorspellend vermogen van een test. Met het vaststellen van deze voorspellende kracht komen methodologische overwegingen en principes aan de orde die minder kenmerkend zijn voor het proces van begripsvalidering.
5. Er wordt een schatting verkregen door te beoordelen hoezeer de inhoud van de test een geheel van situaties, kennisinhouden of vaardigheden representeert, waaruit met betrekking tot de onderzochte persoon conclusies moeten worden getrokken. Dit begrip is vooral van toepassing op het terrein van onderwijskundig meten, waarbij conclusies getrokken worden over het geheel van taken of kennisinhouden waaruit de test een steekproef vormt.
6. Een belangrijk bezwaar tegen het begrip is dat empirisch onderzoek, waarin gebruik wordt gemaakt van gegevens anders dan afkomstig van vergelijkbare toetsen, vaak ontbreekt. De reden is dat het domein van alle mogelijke vragen doorgaans denkbeeldig is, en hooguit bestaat in de vorm van een definitie en een groot, maar toch beperkt aantal items. Het itemdomein is niet veel meer dan een groot aantal toetsen die ieder voor zich niet gevalideerd zijn. Inhoudsvaliditeit is dus vooral een kwestie van oordelen.
7. De betrouwbaarheid van de maten is vaak omgekeerd evenredig aan de relevantie. Er moeten dus criteriummaten gevonden worden die nog zo betrouwbaar zijn dat ze de test niet op voorhanden invalideren, maar niet zo irrelevant zijn dat ze het conceptuele criterium niet meer weerspiegelen.
8. Zie antwoord vraag 7.
9. Met het uiteindelijke criterium wordt het alomvattende, uiteindelijke doel van een procedure bedoeld. Bij de CITO Eindtoets Basisonderwijs kan dit bijvoorbeeld plaatsing op het vwo zijn. Bij tussentijdse criterium gaat het niet om het einddoel, maar om een tussendoel. Dit kan dus

een toets halverwege het jaar zijn. Met een onmiddellijk criterium wordt bijvoorbeeld het totaal aantal behaalde toetsen tot die tijd bedoeld.

10. Het doel van een gestratificeerde steekproef is de proportie die voorkomt in de populatie terug te laten komen in de steekproef. Op die manier wordt een representatieve steekproef getrokken.
11. Het probleem is dat de groep niet representatief is voor de hele populatie. Het is geen probleem wanneer je populatie bestaat uit personen met een gemiddelde hoge intelligentie, een relatief grote interesse in psychologie, grotendeels bestaat uit vrouwen en een leeftijd hebben tussen gemiddeld 18 en 28 jaar.
12. Er kan een test ontstaan die inhoudelijk heterogeen is of waarvan de score onbetrouwbaar is. Dat is zo omdat de moeilijkheid van een item op zich niets heeft te maken met wat het item meet of hoe betrouwbaar het item dit doet.
13. Een homogene test verkrijgt men door bijvoorbeeld factoranalyse uit te voeren op de scores van de k kandidaten voor de test. Met factoranalyse worden items die onderling hoog correleren in groepen bij elkaar genomen. Een andere mogelijkheid is dat de onderzoeker zelf items rechtstreeks in dezelfde test selecteert op basis van hun onderlinge correlaties.
14. In een kliniek, waar het vaak gaat om een vergelijking met een gelijktijdig criterium, zoals een psychiatrische diagnose of een score op een soortgelijke test, zal dit niet veel tijd vergen. Daarom wordt in een dergelijke situatie concurrent validity boven predictieve validiteit gekozen.
15. De oorzaken van de geringe representativiteit die tot 'restriction of range' leiden zijn de uitvallers. Degene die afhaken of afvallen zijn vaak minder geschikt. En degenen die na selectie verdwijnen omdat ze hogerop gaan, zijn vaak de zeer geschikte kandidaten. Het gevolg daarvan is de restriction of range.
16. Als een beoordelaar zijn oordelen door kennis van testuitslagen laat beïnvloeden en zijn oordelen worden als criteriummaat in het valideringsonderzoek gebruikt, dan is duidelijk dat testscore en criteriumscore niet meer onafhankelijk zijn vastgesteld. Dit verschijnsel heet contaminatie van het criterium.
17. Een scatterplot wordt gebruikt om te kijken hoe de correlatie tussen testscore en criteriumscore is.
18. Bij een heteroscedactische relatie bestaat er wel een verband tussen de lagere waarde van twee andere variabelen X en Y, maar niet voor de hogere waarden. Het is zelfs zo dat, naarmate X toeneemt, de spreiding van mogelijke waarden voor Y toeneemt. Het is niet verstandig om de product-momentrelatie kritiekloos toe te passen.
19. Suppressor variabelen spreken door hun merkwaardige rol wel tot de verbeelding. Taalvaardigheid kan bijvoorbeeld een rol spelen in een toets voor rekenvaardigheid waarmee een technisch beroep wordt voorspeld. Door de scores op de rekenvaardigheid te corrigeren voor taalvaardigheid kan een betere voorspelling worden verkregen.

20. Een moderatorvariabele is een variabele die zelf niet of nauwelijks hoeft te correleren met een criteriumscore Y, maar die wel de relatie van andere variabelen met Y beïnvloedt. Een voorbeeld is de motivatie in relatie tot intelligentiescores en schoolprestaties. Dit kan tussen mannen en vrouwen bijvoorbeeld variëren.
21. Ideaal is een batterij die onderling laag correleren en alle hoog met Y correleren. Het is dus zinloos om tests in een batterij op te nemen die ongeveer hetzelfde meten.
- 22.
- Geringe betrouwbaarheid van het criterium.
 - Miskenning van een niet-lineaire relatie tussen predictor en criterium.
 - Negeren van de complexe samenstelling van groepen.
 - Negeren van de variabele betekenis van het criterium in verschillende organisaties.
 - Onterechte vereenvoudiging van het criterium.
23. Soms zijn de verbanden van tests met criteria erg doorzichtig. Soms worden er ook relaties gevonden die zich volledig aan het oog onttrekken. Alleen selecteren op correlaties is niet verstandig om de volgende redenen. Ten eerste leidt het blind toepassen van een test niet tot noodzakelijk inzicht. Ten tweede impliceert het blind afgaan op de grootste correlaties, dat men eigenlijk alle mogelijk predictoren zou moeten onderzoeken op hun geschiktheid voor de voorspelling van het criterium. Verder kan er sprake zijn van een unieke of zeldzame situatie, waarvoor geen criterium vastgesteld zijn. Ook kan de psycholoog die zonder theorie te werk gaat in de knel komen bij de vraag de onderzochte te beschrijven. En als laatste bieden de betekenisanalyse en de begripsvaliditeit een mogelijkheid om het 'sacrosancte' criterium van zijn voetstuk te halen en te analyseren.
24. Een ruimer theoretisch kader, waarbinnen een groot aantal begrippen en relaties tussen die begrippen voorkomen, die in het ideale geval alle empirisch gefundeerd zijn. Dit complex van begrippen en relaties is een voorbeeld van een nomologisch netwerk. Het is van belang hoe de relevante informatie aangeboden wordt, vervolgens wat het type taak is en ten slotte wat de inhoud van de taak is.
25. Voorspellingen kunnen dienen om de oorspronkelijke hypothese te bevestigen, maar ook om de alternatieve hypothesen te verwerpen. De eerste soort wordt de confirmerende validering genoemd, de tweede soort wordt de discriminante validering genoemd.
26. Belangrijk bij de multitrek-multimethodematrix is het idee dat een test voor een bepaalde eigenschap zou kunnen worden vervangen door een andere test die dezelfde eigenschap pretendeert te meten. Dit lukt doorgaans niet perfect, waardoor de variatie in de test resultaten van de onderzochte groep op de bedoelde eigenschap.
- Bij deze benadering wordt dan geschat in hoeverre de variantie van de testscore wordt bepaald door de variantie in de bedoelde trek en juist niet door zowel de variantie ten gevolge van gebruikte methode als de variantie in de niet-bedoelde trekken. Om dit vast te stellen, dient de evaluatieprocedure van de test de meting van diverse trekken en het meten van

verschillende methoden te omvatten. De multitrek-multimethodematrix bevat de correlaties die op deze wijze uit alle combinaties van methoden en trekken ontstaan.