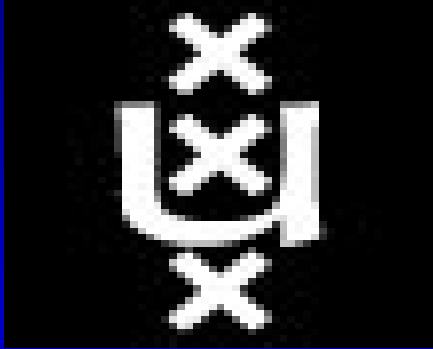




De beoordeling van tests en
toetsen door de COTAN:
Meetinstrumenten de maat
genomen

Arne Evers

RCEC Conferentie 19 november 2008



Arbeids- & Organisationspsychologie
Universiteit van Amsterdam
E-mail: a.v.a.m.evers@uva.nl

Coördinator Testbeoordelingen
van de Commissie Test-
Aangelegenheden Nederland
(COTAN) van het NIP
www.cotan.nl



Inhoud van deze presentatie

- Wat is de COTAN?
- Beoordeling van de kwaliteit van tests
- Stand van zaken in Nederland op het gebied van testkwaliteit
- De kwaliteit van de Nederlandse schooltoets
- Nieuwe ontwikkelingen

Wat is de COTAN?

Het Nederlands Instituut van Psychologen (NIP) heeft in 1959 de Commissie Testaangelegenheden Nederland (COTAN) geïnstalleerd.

De missie van de COTAN is: **het bevorderen van een beter gebruik van betere tests.**

De COTAN probeert dit te realiseren door voorlichting aan testgebruikers en testconstructeurs over o.a. de kwaliteit van tests.

Samenstelling van de COTAN:

- » elke Universiteit met een Afdeling Psychologie heeft een vertegenwoordiger
- » elke sectie binnen het NIP heeft een vertegenwoordiger
- » enkele vertegenwoordigers van grote bedrijven (CITO, KLM)

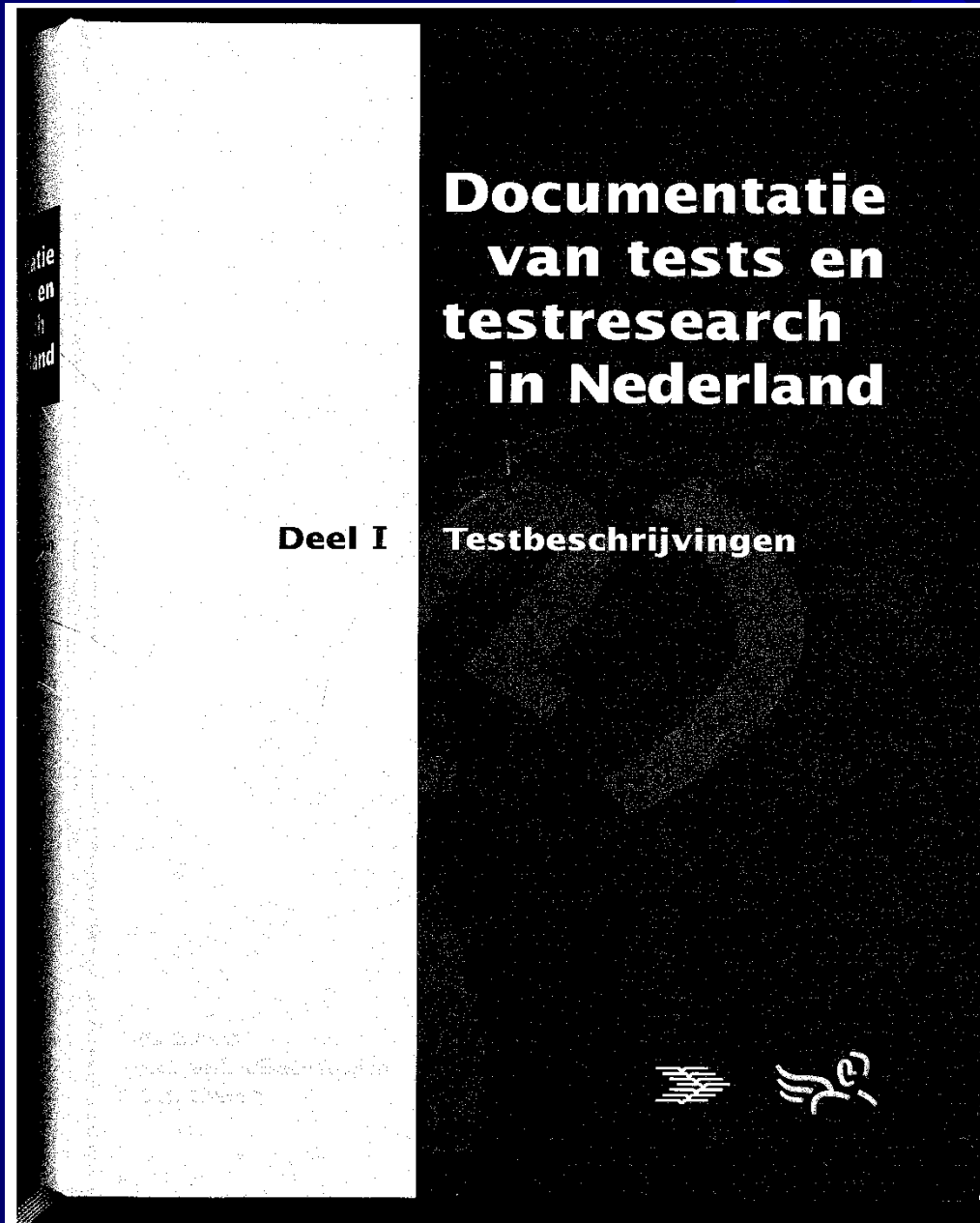


De beoordeling van testkwaliteit

De belangrijkste activiteit van de COTAN is het beoordelen van de kwaliteit van tests, toetsen en vragenlijsten (= tests).

Het doel van de testbeoordelingen is:

- voorlichting aan testgebruikers
- feedback aan testauteurs/uitgevers
- hopelijk leidend tot: verbetering van de kwaliteit van tests



Deze beoordelingen worden gepubliceerd in de

Documentatie van Tests en Testresearch

en in de losbladige

Aanvullingen
(BOOM)

Kwaliteitscriteria:

- Theoretische uitgangspunten
- Kwaliteit van het testmateriaal
- Kwaliteit van de handleiding
- Normen
- Betrouwbaarheid
- Begripsvaliditeit
- Criteriumvaliditeit

De beoordeling op deze zeven criteria kan zijn:
onvoldoende, voldoende, of goed.

Beoordelingsprocedure

1. Selectie te beoordelen test
2. Keuze 2 (onafhankelijke) beoordelaars door stafmedewerker COTAN
3. Vergelijking beoordelingen door Coördinator Testbeoordelingen
4. Consistentie-overleg
5. Verzending beoordeling aan auteur(s) door stafmedewerker
6. Reactie door auteur(s)
7. Reactie COTAN en vaststelling eindoordeel
8. Publicatie beoordeling.

Aandachtspunten bij procedure

- Bij keuze beoordelaars wordt gelet op expertise en niet-betrokkenheid (noch collega, noch concurrent)
- Beoordelaars blijven anoniem
- Na een periode van 1 jaar is herbeoordeling mogelijk op grond van nieuw materiaal/nieuwe gegevens.

Conclusie

Binnen de wereld van het testgebruik neemt Nederland op het gebied van de zorg voor testkwaliteit een vooraanstaande positie in.



Het COTAN-systeem vormt daarbij een gewaardeerd instrument. Uitbreiding met 'review' wenselijk.

Testkwaliteit in Nederland

- Eerst zullen de meest recente totaalgegevens (2008) worden gepresenteerd.
- Vervolgens zal de kwaliteit van de gedocumenteerde tests in 1982, 1992, 2000 en 2008 worden vergeleken.

Kwaliteit in 2008

In het overzicht worden de gegevens van 544 tests vermeld.

Dit is exclusief researchinstrumenten en van oorsprong Belgische tests.

Testkwaliteit in 2008

| | GOED | VOLD | ONVOLD | NVT |
|---------------------|------|------|--------|-----|
| Uitgangspunten | 63% | 25% | 12% | - |
| Testmateriaal | 72% | 21% | 7% | - |
| Handleiding | 51% | 27% | 22% | - |
| Normen | 13% | 29% | 56% | 2% |
| Betrouwbaarheid | 29% | 40% | 30% | - |
| Begripsvaliditeit | 19% | 48% | 33% | - |
| Criteriumvaliditeit | 7% | 22% | 57% | 14% |

Conclusie kwaliteit 2008

- Te weinig goed opgezet normeringsonderzoek en onderzoek naar de criteriumvaliditeit.
- Ook de betrouwbaarheid en het onderzoek naar de begripsvaliditeit laten te wensen over, maar tweederde van de tests is toch minstens 'voldoende'.
- Circa éénvijfde van de tests heeft géén of een slechte handleiding.



Wordt de Nederlandse test beter?

- Vergelijking tussen de jaren waarin een Documentatie werd uitgegeven (1982, 1992 en 2000) en de huidige stand van zaken (2008).
- Sinds 2000 zijn meer dan 170 tests beoordeeld (nieuw of gereviseerd).
- Vergelijking op basis van het percentage tests dat als 'onvoldoende' wordt beoordeeld op elk van de zeven criteria.

Percentages 'onvoldoende'

| CRITERIUM | 1982 N=238 | 1992 N=296 | 2000 N=373 | 2008 N=544 |
|---------------------|---------------|---------------|---------------|---------------|
| Uitgangspunten | 22% | 18% | 13% | 12% |
| Testmateriaal | 19% | 13% | 9% | 7% |
| Handleiding | | | 28% | 22% |
| Normen | 69% | 66% | 59% | 56% |
| Betrouwbaarheid | 47% | 38% | 33% | 30% |
| Begripsvaliditeit | 63% | 50% | 35% | 33% |
| criteriumvaliditeit | | | 57% | 57% |

Voorgaande tabel levert een onvolledig beeld op, omdat alleen naar % 'onvoldoendes' wordt gekeken.

Van elke test kan men de totale kwaliteit berekenen door de oordelen over de vijf (1982, 1992) of zeven (2000, 2008) criteria te middelen.

Deze waarden kan men vervolgens van alle tests op elk van de vier meetmomenten middelen.

De resulterende waarden (op een schaal van 1 t/m 3) zijn:

| | |
|------|----------------|
| 1982 | 1.84 (N = 238) |
| 1992 | 1.94 (N = 296) |
| 2000 | 2.01 (N = 373) |
| 2008 | 2.03 (N = 544) |

Conclusies



Sinds 1982 belangrijke verbetering in kwaliteit van 'de' Nederlandse test.

Afvlakking van de verbetertrend na 2000.

De kwaliteit van de Normen en die van het onderzoek naar Criterium-validiteit baart nog steeds zorgen.



Mogelijke verklaring verbetering

Het COTAN-effect:

Het zichtbaar maken van kwaliteit leidt tot meer inspanningen om de kwaliteit te verbeteren.



Het onderzoekshulpmiddeleneffect:

De wetenschap ontwikkelt zich verder en steeds meer (technische) hulpmiddelen zijn beschikbaar, onderzoek wordt makkelijker uitvoerbaar.



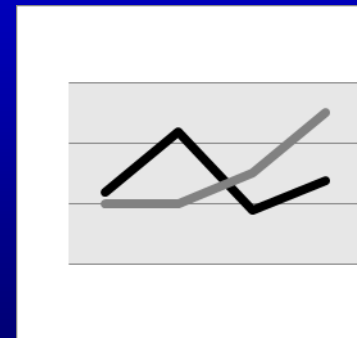
Mogelijke verklaring afvlakking

Bij elke nieuwe Documentatie worden in onbruik geraakte verwijderd, dat is 2008 (nog) niet



tests bij de telling in gebeurd.

Het bestand aan tests wordt bovendien steeds groter, dus relatieve bijdrage van nieuwe tests wordt steeds kleiner.



Kwaliteit schooltoetsen

De percentages 'onvoldoende' op de zeven criteria worden vergeleken van:

- Schooltoetsen en alle andere typen tests
- 'Oude' en 'nieuwe' schooltoetsen
 - oude toetsen: tot en met 2000 (N=58)
 - nieuwe toetsen: 2000-2008 (N=62)

Testkwaliteit van schooltoetsen vergeleken met totaal van 'andere' tests (% 'onvoldoende')

| | Schooltoetsen | Andere tests |
|---------------------|---------------|--------------|
| Uitgangspunten | 10% | 12% |
| Testmateriaal | 2% | 9% |
| Handleiding | 13% | 24% |
| Normen | 49% | 58% |
| Betrouwbaarheid | 34% | 29% |
| Begripsvaliditeit | 39% | 31% |
| Criterionvaliditeit | 33% (85%) | 63% |

Wordt de Nederlandse schooltoets beter? (% 'onvoldoende')

| CRITERIUM | 2000 | 2008 |
|---------------------|------|------|
| Uitgangspunten | 12% | 10% |
| Testmateriaal | 2% | 2% |
| Handleiding | 16% | 10% |
| Normen | 52% | 50% |
| Betrouwbaarheid | 36% | 34% |
| Begripsvaliditeit | 47% | 35% |
| Criteriumvaliditeit | 28% | 42% |

Conclusie schooltoetsen

- De kwaliteit van schooltoetsen steekt gunstig af bij dan van andere typen tests op de meer praktische kwaliteitscriteria, maar ongunstig op de meer psychometrische criteria.
- De kwaliteit van toetsen is tussen 2000 en 2008 toegenomen. Deze verbetering loopt parallel met die van andere typen tests.
- Er is een grote activiteit op het gebied van toetsontwikkeling; het aantal toetsen is sinds 2000 meer dan verdubbeld.
- Dit heeft geleid tot een aantal goede nieuwe instrumenten, maar ook tot veel onderontwikkelde toetsen.

Nieuwe ontwikkelingen (1)

- Momenteel wordt gewerkt aan een grondige herziening van het beoordelingsstelsel.
- Hierin worden onderdelen van de aanpassingen van Keuning (2004) met betrekking tot *Computer Based Tests* en Wools (2007) met betrekking tot *Competentie assessment* opgenomen.
- Verschijnt april 2009.

Nieuwe ontwikkelingen (2)

- In 2009 zal eveneens een nieuwe versie van de Documentatie van Tests en Testresearch verschijnen.
- Deze editie zal digitaal zijn en via een abonnementsysteem zijn te raadplegen.
- Elke maand zal een update plaatsvinden.
- Bij nieuwe beoordelingen zal ook het commentaar van de beoordelaars worden gepubliceerd.

Nieuwe ontwikkelingen (3)

- De COTAN is betrokken bij een project om een internationale ISO-norm te ontwikkelen voor assessment procedures. De definitieve tekst van de norm wordt waarschijnlijk eind 2009 vastgesteld.

