

Testgebruik in de loopbaanadvisering

Is meten weten?¹

Amsterdam
Luken Loopbaan Consult
T.P. Luken²

1. Inleiding
2. Begripsbepaling
3. Soorten tests: een indeling en enkele voorbeelden
4. Aan welke eisen moeten tests voldoen en in hoeverre voldoen zij daadwerkelijk aan die eisen?
5. Welke waarde kan men hechten aan testcores?
6. Consequenties voor de praktijk van het testgebruik
7. Consequenties voor testconstructeurs en eisen aan testgebruikers
8. Vergelijking van tests met andere beoordelingsmiddelen
9. De waarde van tests in de loopbaanadvisering: slotbeschouwing, balans en conclusie

¹ Dit artikel is een bewerking van een hoofdstuk uit het Handboek Studie- en Beroepskeuzebegeleiding (Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1997, p. 8960-1 t/m 8960-29)

² Tom Luken is als adviseur bij loopbaan- en beoordelingsvraagstukken werkzaam bij Stichting NOA en Luken Loopbaan Consult te Amsterdam.

1. Inleiding

De test is bepaald geen onomstreden fenomeen.

"Toen mijn zusje twaalf was, is ze getest. Ze zeiden dat de MAVO erg moeilijk voor haar zou worden, maar vorig jaar is zij cum laude afgestudeerd in de biologie."

"Ik kreeg een test bij een Adviesbureau voor Opleiding en Beroep (AOB). De uitslag klopte precies. Die mevrouw bracht me op een heel nieuw spoor. Ik heb er veel aan gehad."

"Bij dat selectiebureau werd ik als een nummer behandeld. Er werden totaal irrelevante vragen gesteld. Ik voelde me gesloopt na die dag. Op de afwijzing voor de functie kreeg ik geen toelichting..."

Dit zijn zo maar wat geluiden van studenten die in 1996 aan de Akademie Mens-Arbeid begonnen aan het vak beoordelingsmethodiek. Het valt vaak op dat mensen geneigd zijn tot een wat extreem standpunt wanneer het gaat om testen. Tests zijn waardeloze momentopnames en fungeren als een vijgeblad waarmee de adviseur zijn onwetendheid verbergt. Of tests zijn nuttige instrumenten, waarmee men een helder en wetenschappelijk verantwoord inzicht krijgt in de capaciteiten, eigenschappen en voorkeuren van de persoon. Opvallend is dat de discussies hierover dikwijls meer op basis van emotionele en anekdotische dan van zakelijke en exacte argumenten worden gevoerd.

Wat zijn tests nu exact waard? Wat heb je er aan in de keuzebegeleiding? Wat kun je er nu wel mee en wat niet? Hoe moet je ze gebruiken? Beantwoording van deze vragen is het voornaamste doel van dit artikel. Het zal duidelijk zijn dat het niet uitsluitend over "de test" kan gaan. Ook de testgebruiker zal in beeld komen, alsmede wat er gebeurt tussen hem/haar en de cliënt.

We beginnen echter met het verschijnsel test op zichzelf beschouwd. *Wat is een test? Wat onderscheidt een test van andere beoordelingsmiddelen. Wat voor soorten tests zijn er?*

Het betoog zal voor een groot deel worden opgebouwd rond de vragen: *Aan welke eisen moet een test voldoen, in hoeverre voldoen de gangbare tests aan die eisen en wat zijn de consequenties van een en ander?*

Om alle genoemde vragen van een antwoord te kunnen voorzien, zullen enige technisch/exacte onderwerpen, met name in paragraaf 5, niet vermeden kunnen worden. De lezer die vooral geïnteresseerd is de consequenties voor de praktijk, kan de voetnoten desgewenst overslaan.

2. Begripsbepaling

"Een test is een verzameling opdrachten of vragen, waarmee men bepaalde soorten gedrag onder gestandaardiseerde omstandigheden wil oproepen om zo vergelijking tussen personen mogelijk te maken." Zo definieert het Nederlands Instituut van Psychologen de test (NIP, 1988:3).

Een test kan omschreven worden als een instrument waarmee getracht wordt (psychische) hoedanigheden te meten, bijvoorbeeld intelligentie, rekenvaardigheid, extraversie, interesse in techniek enz.

De tests die in de loopbaanadvisering gebruikt worden, bestaan doorgaans nog steeds uit een testboekje met vragen, antwoordformulieren, nakijkmallen en een handleiding, waarin normtabellen zijn opgenomen.

In toenemende mate wordt per computer getest. De vragen, antwoordmogelijkheden,

"sleutel" en normen zijn dan allemaal in de software verwerkt.

3. Soorten tests: een indeling en enkele voorbeelden³.

De catalogus van Swets & Zeitlinger, bijna-monopolist op het gebied van de testuitgeverij, telt zo'n 240 tests. De Documentatie voor Tests en Testresearch in Nederland (NIP, 1992), waarin ook een aantal researchinstrumenten en Belgische tests zijn opgenomen, omvat beoordelingen en beschrijvingen van 372 tests.

Psychologen hebben zich veel moeite gegeven om deze tests op een logische manier in te delen. Dit heeft geleid tot vrij gecompliceerde schema's (zie Drenth en Sijtsma, 1990; NIP, 1992). Op deze plaats kan de volgende, eenvoudige en in het kader van de loopbaanadvisering vaak gehanteerde driedeling volstaan.

1. Intelligentie- en capaciteitentests.

In deze categorie worden de tests gerangschikt die gaan om het *kunnen* van de cliënt/kandidaat.

"Intelligentie" kan daarbij gedefinieerd worden als het (veronderstelde, algemene) denk-, begrips- of leervermogen of het vermogen om nieuwe problemen op te lossen.

"Capaciteiten" is een breder begrip; het staat in het meervoud en verwijst dan ook niet naar één algemeen vermogen, maar naar afzonderlijke vermogens (bijvoorbeeld kunnen rekenen, ruimtelijk inzicht, spelling e.d.).

Intelligentie wordt dikwijls gezien als "pure" aanlegfactor. Het woord capaciteiten kan naast aanleg/inzicht ook slaan op de kennis en vaardigheden die men heeft aangeleerd. Het intelligentiebeprip is voorwerp van steeds terugkerende discussies (zie bijv. Admiraal-Hilgeman, 1994a). Bovendien zijn intelligentietests vooral gericht op het geven van een indicatie voor het algemene niveau. Capaciteitentestseries geven informatie over iemands sterke en zwakke kanten en worden tegenwoordig in de loopbaanadvisering vaker gebruikt dan algemene intelligentietests.

De meest gebruikte test in deze categorie is zonder twijfel de Differentiële Aanleg Testserie (DAT '83, Evers en Lucassen, 1992). Deze bestaat uit negen onderdelen. Enkele hiervan hebben betrekking op taalbeheersing (Woordenlijst, Woordbeeld en Zinnen), enkele op algemeen begripsvermogen (Analogieën voor het verbale gebied en Figuren Reeksen voor het non-verbale), drie kunnen als "exacte" tests worden aangeduid (Ruimtelijk Inzicht, Technisch Inzicht en Rekenvaardigheid) en één onderdeel heeft betrekking op een in zekere zin administratieve vaardigheid (Snelheid en Nauwkeurigheid). Voor het afnemen van de gehele serie is ongeveer vier uur nodig.

2. Belangstellings- of interessetests

Bij de tests in deze categorie gaat het om de vraag wat mensen leuk, prettig of interessant

³ Dit artikel biedt slechts ruimte voor globale informatie en weinig voorbeelden. Een uitvoerig, zij het helaas niet meer op alle punten actueel overzicht biedt De Zeeuw (1986). Veel beknopter en voor een breder publiek geschreven is Van Minden (1987). De artikelen van Admiraal-Hilgeman (1994 a en b) en Luken (1995) hebben specifiek betrekking op testgebruik bij keuzebegeleiding en loopbaanadvisering.

vinden: wat trekt de aandacht, waar zijn ze graag mee bezig?

De bekendste tests in deze categorie zijn de Beroepen Interesse Test (BIT: Evers e.a., 1987) en, voor de hogere niveaus, de Amsterdamse Beroepen Interesses Vragenlijst (ABIV: Evers, 1992).

De BIT bestaat, althans in de officiële versie, uit een formulier waarop de cliënt een route volgt en aldoende steeds (144 maal) op kruispunten komt waar hij⁴ moet kiezen tussen vier beroepsactiviteiten of bezigheden (bijvoorbeeld Porselein beschilderen, Verhalen schrijven, Verzekeringsuitkeringen regelen, Een sterke papiersoort ontwikkelen). Via nakijkmallen worden de keuzes van de cliënt geteld en vervolgens worden de resulterende ruwe scores omgezet in normscores. Deze worden op hun beurt doorgaans uitgezet in een profiel van de beroepsinteressen dat laat zien in welke mate de persoon geïnteresseerd is in negen verschillende interesserichtingen (Technisch Handwerk, Ambachtelijke Vormgeving, etc.)

Bij de ABIV hoeft de cliënt niet steeds te kiezen tussen een aantal mogelijkheden, maar kan zij steeds vrij aangeven of zij beroepen of bezigheden (260 in aantal) aantrekkelijk, onaantrekkelijk of neutraal vindt. Hier resulteert een profiel van 18 verschillende interesserichtingen (Exact-wetenschappelijk, Technisch, Literair etc.)

Afname van BIT of ABIV kost ongeveer een half uur.

3. Persoonlijkheidstests, waarden- en behoeften- en overige tests.

Deze categorie kan opgedeeld worden in twee subcategorieën, nl. enerzijds tests die met motivatie te maken hebben (waarden, d.w.z. wat mensen belangrijk vinden, en behoeften, d.w.z. wat mensen nodig hebben) en anderzijds tests die "andere aspecten van persoonlijkheid of functioneren" betreffen (bijvoorbeeld faalangst, zelfvertrouwen, extraversie). In het kader van de loopbaanadvisering zijn bekende voorbeelden de Guilford LTP Temperament Survey (GLTS, Akkerman en Buijk, 1994), de Nederlandse Persoonlijkheidsvragenlijst (NPV, Luteijn, Starren en Van Dijk, 1979) en de Edwards Personal Preference Schedule (EPPS, Tjoa, 1993).

Bij de laatstgenoemde test moet de cliënt steeds kiezen welk van twee antwoordalternatieven (die sociaal ongeveer even wenselijk of onwenselijk zijn) op hem of haar van toepassing zijn, bijv.: "Ik vind het prettig mensen te kunnen helpen met hun moeilijkheden" of "Ik stel er een eer in alles wat ik doe zo goed mogelijk te doen". Beantwoording van 225 van dergelijke vragen leidt tot een persoonlijkheidsprofiel dat laat zien in hoeverre sprake is van vijftien op behoeften gebaseerde persoonlijkheidstrekken (bijv. ambitie, variatie, etc). De GLTS en NPV bestaan uit vragen waarbij de cliënt moet aangeven in hoeverre de tekst van toepassing is, bijv.: "Ik ben tevreden met het werk dat ik doe. Juist/?/Onjuist."

4. Aan welke eisen moeten tests voldoen en in hoeverre voldoen zij daadwerkelijk aan die eisen?

Een in Nederland gezaghebbend instituut op het gebied van de test is de COTAN (Commissie Testaangelegenheden van het Nederlands Instituut van Psychologen). De eisen die de COTAN aan tests stelt (NIP, 1978 en 1988) kunnen in de volgende zes punten geresumeerd worden. De drie eerste eisen zijn basiseisen. Als daaraan niet is voldaan kan

⁴ Waar 'hij' staat kan in de regel ook 'zij' gelezen worden en vice versa.

volgens de COTAN niet van een test gesproken worden. De laatste drie eisen kunnen beschouwd worden als kwaliteitseisen. Hieraan moet voldaan zijn om van een *goede* test te kunnen spreken.

1. Een voorgeschreven en *gestandaardiseerde* vorm van testopgaven, instructie en testprocedure.

Bij deze eerste eis gaat het er om, dat iedere cliënt/kandidaat dezelfde test op dezelfde manier voorgeschoteld krijgt en dat de omstandigheden waarin dat gebeurt zo goed mogelijk "standaard" zijn. Als aan deze basisvoorwaarde niet is voldaan, is er geen sprake van een meetinstrument. Wanneer de één de vragen op een andere manier krijgt uitgelegd dan de ander, of als de één, om wat voor reden dan ook, een beetje hulp krijgt en de ander niet, dan kan men de uitslagen niet meer goed met elkaar vergelijken. Daarom behoort de testhandleiding een nauwkeurige instructie te bevatten die beschrijft hoe de test moet worden afgenomen.

Alle gangbare tests schrijven in de handleiding inderdaad een gestandaardiseerde instructie en testprocedure voor. Als deze voorschriften worden opgevolgd, krijgen de cliënten, zoals het hoort, dezelfde vragen op dezelfde manier voorgeschoteld. In hoeverre dit altijd gebeurt is nog de vraag. De voorschriften zijn soms uiterst summier, soms pietluttig, soms voor meer dan één uitleg vatbaar en soms zelfs onuitvoerbaar. In dergelijke gevallen moet de testassistent of adviseur bij de testafname wel improviseren.

Een ander probleem is dat de situatie waarin de test wordt afgenomen niet altijd volgens de standaard is (boren bij de burens, ordeproblemen in de klas, brandweerauto's in de straat...)

Een volgend probleem is dat soms niet-officiële versies van tests (bijvoorbeeld de BIT⁵) worden gebruikt, waarbij een instructie ontbreekt.

Een laatste probleem is dat de test niet altijd door gekwalificeerd personeel wordt afgenomen. De moeilijkheidsgraad van dit werk wordt wel eens onderschat, waardoor het soms overgelaten wordt aan niet voldoende geïnstrueerde en ingewerkte krachten.

2. De tweede eis die aan tests gesteld wordt is een voldoende *objectieve* procedure van verwerking, scoring, beoordeling of classificatie van de antwoorden of reacties.

De antwoorden of reacties van de geteste dienen door verschillende testafnemers op dezelfde wijze verwerkt en beoordeeld te worden. Dat wil zeggen dat hetzelfde antwoord niet door de één goed en door de ander fout wordt gerekend of van de één twee punten en van de ander nul punten krijgt.

Vanwege deze eis zijn bij moderne tests multiple choice vragen gangbaar. De objectiviteit is daarbij optimaal. Bij open vragen of opdrachten, zoals gebruikelijk bij "projectietests" (bijvoorbeeld de Rorschach inktvlekkentest) en bij enkele oudere intelligentietests, worden uitvoerige scoringsinstructies opgenomen, die soms meerdere of zelfs vele bladzijden tekst voor één vraag of opdracht beslaan. Echter: hoe meer uitleg nodig is, des te meer kans dat het subjectieve oordeel van de testafnemer een rol gaat spelen.

Vanwege de problemen op het gebied van de objectieve scoring (en andere, ermee samenhangende problemen op het gebied van de betrouwbaarheid en validiteit) is het gebruik van projectietests (Rorschach, Wartegg, boomtest, Zinnen-Aanvul test) in de professionele loopbaanadvisering verminderd. De meeste nu gebruikte tests voldoen aan de eis van

⁵ Zie bijlage 1 voor de betekenis van afkortingen van bij keuzebegeleiding gebruikte tests.

objectieve "scoring" van de antwoorden. Subjectiviteit van de testafnemer speelt in de regel geen rol.

De COTAN publiceert met enige regelmaat officiële beoordelingen van in Nederland verschenen tests (NIP, 1992). Het overzicht van 1992 omvat 372 tests. Vier op de vijf van deze tests worden op beide bovenstaande punten als voldoende beoordeeld. De onvoldoendes betreffen veelal wat oudere tests in categorieën die bij de loopbaanadvisering niet gangbaar zijn.

3. De derde eis stelt dat de test officieel via een uitgever (of anders informeel via de betreffende constructeur of instantie) *beschikbaar* moet zijn.

Dit is een kwestie van definitie. Een zelf ontwikkeld instrument dat in de bureaulade van een dekaan of adviseur ligt, maar niet in de handel of anderszins verkrijgbaar is, is volgens het NIP geen test.

Bekend is dat er van instrumenten als de BIT en het BZO talloze niet-officiële, min of meer "verkrijgbare" versies bestaan. Afgezien hiervan gebruiken loopbaanadviseurs allerlei eigen of op lokaal niveau beschikbare instrumenten. Het is hierbij van belang dat de gebruiker de beperkingen aangeeft en dat bij de cliënt niet de indruk ontstaat dat het om (wetenschappelijk verantwoorde) tests gaat.

Tot voor enkele jaren waren tests op grond van een "gentlemans agreement" van het NIP en Swets & Zeitlinger vrijwel uitsluitend beschikbaar voor psychologen en psychodiagnostisch geschoolde pedagogen. Tegenwoordig verkoopt Swets talloze tests (waaronder alle interestests) zonder enige beperking. In het navolgende zal nog duidelijk worden dat dit niet zonder risico is.

4. De vierde eis houdt in dat er duidelijke *normgegevens* op gedefinieerde groepen zijn.

Een ruwe testuitslag (bijvoorbeeld "5 vragen goed") zegt niets als men niet weet wat een "normale" uitslag is. Er moet goed vergelijkingsmateriaal zijn, een "meetstandaard". Pas dan kan men weten of de uitslag hoog of laag is. Daarom zijn "normen" nodig. De normgroepen moeten goed gedefinieerd zijn. Als men zegt dat een bepaalde uitslag hoog is vergeleken met scholieren, zegt dat nog maar weinig. Het zegt veel meer als men kan zeggen dat het gaat om een vergelijking met een normgroep van bijvoorbeeld derde klassers MAVO. Bij voorkeur moet men dan ook nog weten of het om een landelijke groep gaat of een groep uit een bepaalde provincie, hoe groot de groep was, of hij uit jongens en/of meisjes bestond en in welk jaar het vergelijkingsmateriaal was samengesteld.

De COTAN is kritisch wanneer zij de normen van tests beoordeelt. Maar liefst 72% van de tests krijgt op dit punt een onvoldoende! Gangbare problemen bij de normen zijn: gebrekkige representativiteit van de steekproef voor de doelgroep, te kleine normgroepen (kleiner dan circa 250) en onduidelijke omschrijvingen van de kenmerken van de steekproef. Het grootste probleem, althans in de ervaring van veel testgebruikers, ligt echter in de gebrekkige actualiteit van de normen. Dit probleem speelt sterk bij interestests, omdat interessen zich, zeker bij jongeren, snel kunnen ontwikkelen niet alleen vanwege ontwikkelingspsychologische oorzaken, maar ook vanwege soms snel veranderende "modes" (Luken, 1995). Een hernormering steeds na maximaal vijf jaar zou bij deze tests bepaald geen overbodige luxe zijn. In werkelijkheid komt het regelmatig voor dat normen worden gebruikt van tien jaar oud of nog ouder. Geen wonder dat veel testgebruikers dan afzien van het gebruik van normen en slechts de ruwe scores interpreteren, waardoor feitelijk niet meer van

een test gesproken kan worden.

5. Als vijfde worden positieve gegevens geëist over de *betrouwbaarheid*.

Bij het begrip "betrouwbaarheid" gaat het om vragen als: is de test een goed meetinstrument? Als de ene persoon de test afneemt, komt er dan inderdaad hetzelfde uit als wanneer een andere persoon het doet? Als de cliënt de test over een poosje nog eens doet, behaalt hij dan hetzelfde resultaat? Is de testuitslag eenduidig? Meet de test nauwkeurig? Wanneer deze vragen met ja kunnen worden beantwoord, is de test betrouwbaar. Als de testuitslagen echter voor een groot deel door irrelevante invloeden of door het toeval worden bepaald, is de test onbetrouwbaar.

Een voorbeeld. Wanneer een groep werknemers twee maal een bepaalde interestest maakt binnen een maand en de uitslagen komen bij alle werknemers de tweede keer goed overeen met de eerste keer, dan spreekt men van een goede betrouwbaarheid.

De betrouwbaarheid van een test wordt uitgedrukt in een getal tussen 0 en 1, de zogenaamde betrouwbaarheidscoëfficiënt. Deze coëfficiënt staat voor de "correlatie van de test met zichzelf". Het kan bijvoorbeeld gaan om de correlatie tussen een eerste en een tweede afname of tussen de ene en de andere helft van de test⁶.

Als de testuitslag puur door het toeval bepaald zou worden, dan is de betrouwbaarheid nul. De test meet dan niets. Er is geen enkele correlatie tussen een eerste en een tweede afname (of tussen de ene en de andere helft etc.)

Indien het toeval bij de uitslag geen enkele rol speelt, is de betrouwbaarheid 1: de meting is perfect.

Bij tests vinden we nooit betrouwbaarheden van 0 of 1, ze liggen er altijd ergens tussenin. De COTAN stelt dat de betrouwbaarheid eigenlijk altijd boven de 0,80 zou moeten liggen om een test in de praktijk te kunnen gebruiken. Dit niveau worden lang niet altijd gehaald. De COTAN beoordeelt 41% van de tests op dit punt als onvoldoende. De betrouwbaarheden van de onderdelen van de meest gebruikte capaciteitentest, de DAT, variëren tussen 0,73 (voor het onderdeel Woordbeeld) en 0,88 (Ruimtelijk Inzicht). Die van de BIT variëren tussen 0,72 (voor de schaal Handel) tot tussen de 0,85 en 0,97 voor de andere schalen.

6. De zesde en laatste eis betreft de *validiteit*.

Bij de validiteit gaat het niet zo zeer om de vraag of de test goed meet, maar meer of de test het goede meet. Meet de test inderdaad wat hij moet meten? Beantwoordt de test aan zijn doel?

Een kleine anekdote, ontleend aan Vroon (1980) kan het lastige, maar in de psychometrie belangrijke onderscheid tussen betrouwbaarheid en validiteit verduidelijken. Francis Galton probeerde ruim een eeuw geleden een intelligentietest te construeren. Hij ging uit van het idee dat intelligentie gedefinieerd kan worden als het kunnen opnemen van informatie. Informatie opnemen gaat via de zintuigen. Op basis van deze gedachte ontwikkelde Galton een, wat wij nu een soort waarnemingstest zouden noemen: kun je goed verschillen zien tussen objecten? Onderscheid je kleuren? Hoor je tonen goed? Etcetera. Hoe beter je waarneemt, des te beter

⁶ Hier wordt gerefereerd aan de stabiliteit of test-hertestbetrouwbaarheid en aan de "split-half-betrouwbaarheid". Er zijn nog diverse andere methoden om de betrouwbaarheid van een test vast te stellen. Zie voor een beknopt overzicht Admiraal-Hilgeman, 1991, of voor een meer volledig overzicht Drenth en Sijtsma, 1990.

je informatie opneemt, des te intelligenter je bent, dacht Galton. Hij probeerde zijn test uit op de wereldtentoonstelling in 1884 te Londen. De test bleek voldoende betrouwbaar. Zo scoort iemand die een eerste keer laag scoort, bij een tweede afname ook laag. Maar helaas voor Galton: eminente wetenschappers bleken niet hoger uit te komen dan de gewone man of vrouw in de straat. Met andere woorden: de test was wel betrouwbaar, maar niet valide. Althans uitgaande van de veronderstelling dat eminente wetenschappers intelligenter zijn dan de gemiddelde mens....

Bij de validiteit gaat het dus om de vraag of de test inderdaad dat meet wat hij pretendeert te meten. Om dit te kunnen bepalen is, in tegenstelling tot bij de betrouwbaarheid, een criterium nodig dat buiten de test ligt. Om vast te stellen of een nieuwe intelligentietest valide is, is een criterium nodig voor intelligentie, bijvoorbeeld een andere intelligentietest waarvan de validiteit reeds bewezen is, of bijvoorbeeld schoolprestaties, uitgaande van de veronderstelling dat intelligente leerlingen beter presteren op school⁷.

Ook de validiteit van een test wordt uitgedrukt d.m.v. een coëfficiënt. Ook hier gaat het om getallen tussen 0 en 1⁸. In dit geval gaat het om de correlatie van een test met het (externe) criterium.

Omdat het veel moeilijker is om de validiteit van een test aan te tonen dan de betrouwbaarheid, stelt de COTAN t.a.v. de validiteit minder exacte en aanzienlijk lagere eisen: "Bruikbare validiteiten liggen gewoonlijk tussen 0,40 en 0,70". Niettemin wordt van 56% van de Nederlandse tests de validiteit onvoldoende geacht. Validiteiten liggen zelden hoger dan 0,50.

De onderstaande tabel resumeert de beoordelingen van de COTAN voor tests die in de loopbaanadvisering vaak gebruikt worden. Een mogelijke conclusie is dat deze tests kwalitatief beter zijn dan de "gemiddelde" test.

	DAT	BIT	ABIV ⁹	PMT-K	SVL
uitgangspunten	goed	onvoldoende	goed	goed	voldoende
uitvoering	goed	goed	goed	goed	goed
normen	goed	voldoende	voldoende	goed	goed
betrouwbaarheid	voldoende	voldoende	goed	goed	voldoende
validiteit	goed	onvoldoende	goed	goed	voldoende

⁷ Hier wordt gerefereerd aan "construct-" en "concurrente validiteit". Ook voor de bepaling van de validiteit bestaan veel methoden. Zie voor nadere documentatie de vorige voetnoot.

⁸ In theorie, en soms ook in de praktijk, zijn ook negatieve getallen mogelijk, namelijk wanneer de test negatief correleert met het criterium. Als de leerlingen met een hoog IQ het op school minder goed zouden doen dan de leerlingen met een laag IQ, dan zou de betreffende intelligentietest een negatieve validiteitscoëfficiënt hebben.

⁹ De beoordeling betreft de versie uit 1979. De ABIV '92 zal naar verwachting minstens even positief beoordeeld worden.

Tabel 1: beoordelingen van de COTAN (NIP, 1992) voor enkele bij loopbaanadvisering gangbare tests

5. Welke waarde kan men hechten aan test scores?

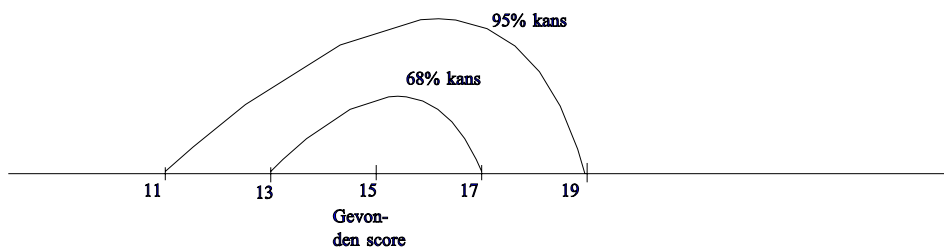
Het al dan niet voldaan zijn aan de eerste twee van de hiervoor genoemde eisen (gestandaardiseerde afname en objectieve scoring) manifesteert zich in de betrouwbaarheid. Als de testafname niet standaard verloopt of als de beoordeling niet objectief is, dan zal de test in de regel niet betrouwbaar zijn. Als de test op zijn beurt niet betrouwbaar is, zal hij ook niet valide zijn. Betrouwbaarheid is een noodzakelijke, maar niet voldoende voorwaarde voor validiteit.

Deze paragraaf handelt over de vraag welke waarde nu aan test scores gehecht kan worden. Om deze vraag te kunnen beantwoorden zullen wij ons richten op een "vertaling" van de begrippen betrouwbaarheid en validiteit, zodat een interpretatie voor praktische situaties mogelijk wordt.

Betrouwbaarheid

Een begrip dat in de praktijk meer houvast biedt dan betrouwbaarheid is de zogenaamde "standaardmeetfout". Hiermee kan men een schatting maken van de te verwachten foutenmarge. Bij recentelijk ontwikkelde tests staat de standaardmeetfout doorgaans vermeld in de testhandleiding. Als dit niet het geval is, kan men met een eenvoudige formule de standaardmeetfout berekenen uit gegevens die wel altijd in de handleiding staan¹⁰.

De standaardmeetfout geeft aan binnen welke marge rond de gevonden score de "ware" score zich waarschijnlijk moet bevinden. Deze marge wordt ook wel betrouwbaarheidsinterval genoemd. Met 68% zekerheid kan men stellen dat de ware score niet verder dan één standaardmeetfout van de gevonden score af ligt. Met 95% zekerheid kan men stellen dat de ware score maximaal 2 standaardmeetfouten van de gevonden score af ligt. Dit laatste interval wordt echter zelden toepast, omdat dit in de praktijk tot zeer brede marges leidt.



Figuur 1: illustratie standaardmeetfout

Figuur 1 laat de toepassing van de standaardmeetfout zien. Stel dat een cliënt op een test ruwe

¹⁰ $s_e = s_x \sqrt{1 - r_{xx}}$

Hierin is s_e de standaardmeetfout, s_x de standaarddeviatie en r_{xx} de betrouwbaarheid.

score 15 heeft behaald en dat de standaardmeetfout van de betreffende test 2 is. Dan ligt de "ware score" van deze cliënt waarschijnlijk (d.w.z. 68% kans) tussen 13 en 17. Vrijwel zeker (95% kans) ligt de ware score tussen 11 en 19.

In welke orde van grote de marges doorgaans liggen, laat het volgende, redelijk representatieve voorbeeld zien. Het is ontleend aan de handleiding van de DAT (Evers en Lucassen, 1991). Een cliënt heeft bij het onderdeel Woordenlijst bij 46 van de 75 items de juiste betekenis van het opgegeven woord gekozen. De ruwe score 46 wordt via een normtabel omgezet in een staninescore. Dit is een normsysteem dat loopt van 1 tot 9, waarbij 5 precies gemiddeld is. De staninescore van de betreffende cliënt is 4. Wanneer men de standaardmeetfout toepast, moet men concluderen, dat de ware score ook 3 of 5 had kunnen zijn. Met een zekerheid van 68% kan men stellen dat de ware score 3, 4 of 5 is. Dit betekent dat er 32% kans is, dat de ware score lager dan 3 of hoger dan 5 is.

Validiteit

De waarde van tests hangt uiteindelijk af van de validiteit. Zoals eerder opgemerkt drukt men de validiteit van tests uit in validiteitscoëfficiënten. Dit zijn correlaties tussen de test enerzijds en een criterium anderzijds¹¹. Bij de DAT hebben de auteurs bijvoorbeeld onderzocht in hoeverre de scores van in het najaar van 1982 afgenomen tests samenhangen met de schoolcijfers van het paasrapport in maart 1983 (Evers en Lucassen, 1991). De meeste van de gevonden correlaties bleken te liggen tussen 0,00 en 0,30.

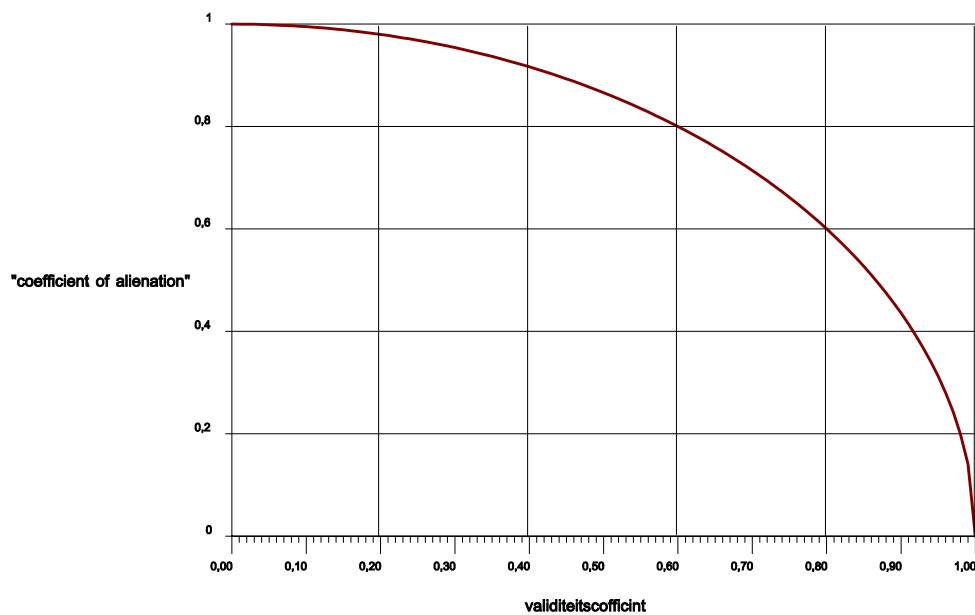
Maar hoe moeten dergelijke correlaties geïnterpreteerd worden?

Zeker veel "leken", maar helaas soms ook wel adviseurs zijn geneigd de betekenis van correlaties te overschatten. Men denkt bijvoorbeeld wel eens dat een correlatie van 0,30 staat voor een samenhang van 30%. Dit is ver bezijden de waarheid. In verschillende boeken over testtheorie (bijv. Magnusson, 1967) en statistiek (bijv. Spitz, 1968) kan men illustraties vinden zoals onderstaand, die laten zien hoe groot het deel van het voorspelde criterium is, dat *niet* bepaald is door de voorspellende test. Dit wordt uitgedrukt in de zogenaamde "coëfficiënt of alienation"¹².

¹¹ Bij "voorspellende" of "predictieve" validiteiten ligt het criterium in de toekomst. Men berekent bijvoorbeeld de correlatie tussen de scores op een test voor ruimtelijk inzicht en de cijfers voor wiskunde een jaar later. Bij "gelijktijdige" of "concurrent" validiteiten ligt het criterium in het heden. Correlaties worden bijvoorbeeld berekend van de resultaten op een prestatiemotivatietest met actuele proefwerk cijfers. Voorspellende validiteiten liggen doorgaans op een lager niveau dan gelijktijdige validiteiten. Overigens gebruikt men ook bij gelijktijdige validiteit vaak het woord voorspellen: men "voorspelt" met de test welke cijfers de leerlingen *nu* behalen (of op het laatste rapport behaald hebben).

¹² De "coëfficiënt of alienation" (ca) is de standaarddeviatie van het criterium bij een bepaalde testscore gedeeld door de standaarddeviatie van het criterium los van de voorspellende test. "Deze coëfficiënt geeft aan hoeveel van de variatie in de ene variabele vreemd is aan de variatie in de andere variabele, dus moet worden toegeschreven aan interfererende variabelen." (Spitz, 1968:275) De ca kan berekend worden als de wortel uit 1 min het kwadraat van de validiteitscoëfficiënt. In formule:

$$ca = \sqrt{1 - r^2_{xy}}$$



Figuur 2: De "coefficient of alienation" als functie van de validiteitscoëfficiënt

De illustratie laat zien dat zelfs bij validiteitscoëfficiënten die als zeer hoog worden beschouwd (bijvoorbeeld 0,60), verreweg het grootste deel (80%) van de variatie op het criterium door andere factoren dan de voorspellende test wordt bepaald. In de regel wordt nooit meer dan 10% van de variatie in schoolprestaties door capaciteitentests voorspeld. Andere factoren (zelfvertrouwen, doorzetten, gezondheid, sociale steun, toeval...) bepalen minstens 90%¹³

¹³ Lofquist en England (1961) hebben een tabel opgesteld waaruit blijkt tot welke foutenmarges bij voorspellen een en ander leidt. De tabel is opgesteld in percentielscores. Het percentielscoresysteem gaat uit van een schaal van 1 tot 100. Percentielscore 69 wil zeggen dat 69% van de mensen in de normgroep een lagere (of zelfde) score heeft behaald. Andersom geredeneerd: betreffende cliënt zit bij de 31% hoogste scores.

Een voorbeeld uit de tabel van Lofquist en England: iemand die bij een capaciteitentest met een (vrij hoge) validiteit van 0,50 percentielscore 69 behaalt, zal op het voorspelde criterium (bijv. een schoolvorderingentoets) met 68% zekerheid een percentielscore behalen tussen 27 en 86. Dan is er dus nog 32% kans dat de betreffende leerling lager dan 27 of hoger dan 86 uitkomt...

Deze grote onzekerheidsmarges zijn overigens niet alleen aan de test te wijten, maar ook, of zelfs nog sterker, aan de onbetrouwbaarheid en het gebrek aan validiteit van het criterium. De meettechnische kwaliteiten van proefwerk- en rapportcijfers laten veel te wensen over. De onbetrouwbaarheid van test en criterium is in bovenstaande gegevens ingecalculleerd en de onzekerheid die de onbetrouwbaarheid met zich meebrengt, moet er dus niet nog eens bij opgeteld worden.

6. Consequenties voor de praktijk van het testgebruik

1. Wanneer een test niet volgens de *standaard* wordt afgenomen, is het eigenlijk geen test meer. Een vergelijking met de normgroep is niet meer goed mogelijk. Wat de test als belangrijkste kwaliteiten te bieden heeft (een relatief betrouwbare meting en het bieden van geschikt vergelijkingsmateriaal) valt dan weg.

Daarom is het van groot belang dat de instructies voor testafname zo goed mogelijk (hetgeen niet automatisch betekent zo letterlijk mogelijk) worden opgevolgd. Goed opleiden en inwerken van de personen die het afnemen van de tests verzorgen, is hierbij een belangrijke randvoorwaarde.

Communicatie met de testconstructeur of -uitgever is van belang, wanneer er vragen of opmerkingen zijn over de standaardinstructies. Soms kan de zaak opgelost worden, anders kan het probleem bij een volgende uitgave wellicht voorkomen worden.

Het gebruik van niet-officiële versies van tests moet in zijn algemeenheid afgekeurd, c.q. ontraden worden. Al dan niet zelfgemaakte computerversies van (interesse)tests zijn niet automatisch equivalent aan de oorspronkelijke versie (zie bijv. Akkerman, 1994). Wanneer men een niet-officiële versie van een test gebruikt, kan men zich bij problemen niet tot de officiële constructeur of uitgever wenden. Als men geen auteursrechten betaalt is het niet terecht om te klagen dat nieuwe (officiële) versies of nieuwe normen uitblijven.

Zelf hulpmiddelen maken hoeft in zijn algemeenheid natuurlijk zeker niet afgekeurd of afgeraden te worden. Men dient zich hierbij echter te realiseren aan welke eisen voldaan moet worden alvorens men van een meetinstrument kan spreken en men dient te voorkomen dat de cliënt denkt met een test van doen te hebben.

2. Omdat vrijwel alle gangbare tests van een multiple choice systeem uitgaan, levert, zoals reeds is opgemerkt, de *objectieve* verwerking en scoring in de regel geen problemen op. Wel verdient het voorkomen van fouten bij het tellen van de antwoorden en het aflezen van de genormeerde scores veel aandacht. In concreto: een goede instructie voor de testassistent (waarbij het houdingsaspect niet buiten beschouwing blijft), inwerken, (zelf)controle en het gebruik van hulpmiddelen. Wat dit laatste betreft: het computerprogramma SCORPROG¹⁴ kan de kans op fouten bij het scoren en normeren sterk verkleinen en kan bovendien een aanzienlijke tijdwinst opleveren.

3. Wat betreft de *verkrijgbaarheid*: een instrument dat men slechts voor eigen gebruikt hanteert, is volgens de definitie van het NIP geen test. Afgezien hiervan zijn er minstens twee redenen om een zelfgemaakt instrument via publicatie of anderszins aan anderen ter beschikking te stellen. Andere loopbaanadviseurs en hun cliënten kunnen er dan ook baat bij hebben en bovendien wordt controle van buitenaf en daarmee wellicht verbetering mogelijk.

4. Wat betreft de *normen*: veel tests voldoen niet aan de (strenge) eisen die de COTAN op dit punt stelt. In de praktijk is het vaak een probleem dat er geen geschikte norm voorhanden is (voor het MBO zijn bijvoorbeeld vaak geen toegespitste normen aanwezig) of dat de normen verouderd zijn. Het ontwikkelen van "lokale normen" kan een bijdrage leveren aan de oplossing van deze problemen. Testen per computer kan dit sterk vergemakkelijken.

¹⁴ Verkrijgbaar bij Swets & Zeitlinger te Lisse.

Als de achtergrond van de cliënt (bijvoorbeeld qua leeftijd, of etnische/culturele herkomst) afwijkt van die van de mensen in de normgroep, dan moet men, als men de test al afneemt, extra voorzichtig zijn bij het interpreteren van de uitslagen. Het klinkt paradoxaal, maar het gebruik van aparte normen kan raadzaam zijn, als men discriminatie of rolbevestiging wil voorkomen (zie bijv. Luken, 1995).

5. De *betrouwbaarheid* van tests is een relatief sterk punt. Betrouwbaarheidscoëfficiënten van 0,80 of hoger laten zien dat er wel degelijk iets gemeten wordt. Een cliënt die zegt dat de uitslag puur toeval is of dat hij, als hij beter had geslapen, veel hoger uitgekomen zou zijn, heeft in de regel geen gelijk. De testgebruiker zal echter niet alleen moeten waken voor onderschatting, maar ook voor overschatting van de betekenis van de testuitslag. Testgebruikers zouden eigenlijk altijd globaal op de hoogte moeten zijn van de standaardmeetfouten van de testonderdelen en deze kennis in de uitleg aan de cliënt moeten overbrengen. Een eenvoudige maatregel zou kunnen zijn om in het profielformulier waarmee de cliënt op de hoogte wordt gebracht van de uitslagen, niet één kruisje te zetten bij de behaalde score, maar de zone aan te geven waarbinnen de "ware score" waarschijnlijk (68% kans) ligt. Dit is het gebied van een standaardmeetfout boven tot een standaardmeetfout onder de gevonden score. Bij een test als de D.A.T. is dat een gebied van in totaal ongeveer drie stanines.

6. De testgebruiker is waarschijnlijk het meest geïnteresseerd in de *validiteit* van de test. Beantwoordt de test aan zijn doel? Uit de betrouwbaarheid blijkt dat de test iets meet, maar *wat*? Wat komt men met het meten nu eigenlijk te weten?

Helaas is de validiteit van tests een minder sterk punt. De COTAN stelt hier in verhouding lage eisen en die worden lang niet altijd gehaald. Wat de test precies meet en wat de betekenis daarvan is, is doorgaans niet zo duidelijk. Voor een groot deel kan dit verklaard worden uit het feit dat de psychologie een relatief jonge wetenschap is die op een moeilijk terrein opereert. Als men bestudeert hoe mensen denken en zich gedragen en hoe hun drijfveren werken, stuit men op veel complicaties. Over veel psychologische begrippen bestaat dan ook nog onduidelijkheid en/of onenigheid. Dit geldt zelfs voor zeer veel gebruikte begrippen als "persoonlijkheid", "intelligentie" (Admiraal-Hilgeman, 1994a en -b) en "interesse" (Luken, 1995). Een en ander betekent dat zelfs bij redelijk geachte validiteiten het verklarend en voorspellend vermogen van de test gering is. De betekenis van de gemeten begrippen is niet helder. En te veel andere onbekende en/of niet of nauwelijks meetbare en voorspelbare factoren spelen een rol. Wat zijn bijvoorbeeld de krachten in de sociale omgeving van de persoon? En wat is het effect van de testuitslag op de cliënt? Dezelfde uitslag kan bij de één leiden tot ontmoediging en bij de ander tot een besluit "ik zal ze eens laten zien!"

Een consequentie van dit alles voor de testgebruiker is, zoals ook het NIP aanbeveelt in een richtlijn (NIP, 1988:97) de cliënt te waarschuwen *niet* alleen op testcores af te gaan bij het nemen van loopbaan/beroepskeuzebeslissingen en de cliënt aan te moedigen ook via andere wegen verkregen informatie over bijvoorbeeld persoonlijke en sociale vaardigheden, interesses, waarden etc. bij de keuze te betrekken.

Een andere consequentie is dat de testgebruiker terughoudend opereert bij het geven van adviezen op grond van testresultaten. Uit het voorafgaande is gebleken hoe weinig men met testresultaten kan voorspellen. Hoe graag de cliënt ook heldere antwoorden wil krijgen ("moet ik het nu wel of niet doen, wat zegt de test?") en hoe graag de adviseur deze ook wil geven, op basis van tests alleen is dit niet op verantwoorde wijze mogelijk. Een cliënt die zich afvraagt of een avond HTS een haalbare kaart is, kan men in de regel niet met ja of nee

antwoorden. Bij gebrek aan onderzoeksgegevens kan men helaas zelfs weinig zeggen over de kansen op het diploma. Wel kan men bijvoorbeeld zeggen: je verstandelijke capaciteiten, voorzover gemeten, liggen op hetzelfde niveau als die van de gemiddelde HTS-leerling, dus wat die capaciteiten betreft zou je het moeten kunnen halen. En vervolgens kan men overgaan tot een bespreking van de achtergrond van haar twijfel en van andere factoren die bepalend zijn voor succes (interesse, zelfvertrouwen, doorzettingsvermogen, gezondheid, omstandigheden etc.)

Bij een ander zal men misschien moeten zeggen dat zijn aanlegtestresultaten lager dan het gemiddelde HTS niveau liggen, maar dat bijvoorbeeld 25% van de HTS-studenten testuitslagen op hetzelfde niveau hadden. Waarschijnlijk zijn diens kansen geringer dan die van de gemiddelde student, maar hoe gering? In hoeverre is compensatie nodig en mogelijk met extra inspanning? In ieder geval moet de adviseur oppassen opleidingen of beroepen op grond van uitsluitend capaciteitentestresultaten te ontraden. Het VWO wordt succesvol afgerond door talloze leerlingen met aanlegtestresultaten op MAVO-niveau. Op de MAVO zitten talloze leerlingen¹⁵ die qua aanleg net zo goed op het VWO hadden kunnen zitten. In de wereld van het werk is de overlap tussen niveaus nog sterker dan bij opleidingen het geval is. Er zijn advocaten met een IQ van 95 en vrachtwagenchauffeurs met een IQ van 145 (Roe, 1956).

Wat hier gezegd is over capaciteitentests geldt mutatis mutandis ook, zelfs nog sterker, voor interesstests en persoonlijkheidstests. Ook deze bieden onvoldoende basis om bij individuele vraagstellingen verantwoorde voorspellingen te kunnen doen over de opleidingen en beroepen waarin de cliënt succesvol en tevreden zal zijn.

Welke bijdrage kunnen tests leveren in de loopbaanadvisering?

In het voorafgaande zijn de beperkingen van de test ruimschoots aan de orde geweest. Onder meer is geconcludeerd dat het niet verantwoord is om duidelijke adviezen te geven aan individuele cliënten op basis van uitsluitend testuitslagen. Sommigen zouden wellicht op grond hiervan ertoe over willen gaan het testgebruik in de loopbaanadvisering af te schaffen. Men zou dan echter voorbijgaan aan wat men wel allemaal kan met tests. Hier volgt een inventarisatie.

- Tests bieden een goed uitgangspunt voor een gesprek. Zij helpen de dialoog te structureren met behulp van begrippen die bij loopbaankeuzes belangrijk zijn.
- Tests geven concrete en exacte informatie die moeilijk op een andere manier te verkrijgen is. Dat cliënt A vergeleken met een landelijke normgroep bij de beste 20% op test X uitkomt, is geen toeval. Het is een feit. De betekenis van dit feit zal echter bij voorkeur door de cliënt zelf, daarbij geholpen door de adviseur die waakt voor het overdrijven of onderschatten van de betekenis, moeten worden geïnterpreteerd.
- De feiten die tests opleveren kunnen bijdragen aan de verheldering en het realistisch maken van het zelfconcept. Zoals bekend draagt helderheid en realisme van het zelfconcept sterk bij tot de kwaliteit van (school)loopbaankeuzes (zie bijv. Taborsky,

¹⁵ Ongeveer één op de tien MAVO leerlingen heeft een "doorsnee" VWO aanleg en vice versa, als men afgaat op de verbale tests of op Rekenvaardigheid van de DAT. Maar liefst ongeveer 50% van de VWO leerlingen heeft een "doorsnee" MAVO aanleg en vice versa, als men uitgaat van de scores op de onderdelen Ruimtelijk en Technisch Inzicht. (Afgeleid uit normtabellen DAT handleiding, Evers en Lucassen, 1991)

1987 of Luken, 1990). Temidden van alle subjectieve en vervormde oordelen van anderen en van de persoon zelf, bieden testresultaten een van de weinige mogelijkheden voor concreet, objectief houvast.

- Bij goed gebruik stimuleren testresultaten tot het formuleren van zinvolle vragen, het leggen van verbanden en het verder ontwikkelen van de zelfkennis in relatie tot opleidings-/arbeidsmogelijkheden, d.w.z. tot adequaat leergedrag met betrekking tot de eigen (school)loopbaan.
- Voor interesstests geldt dat zij een systematische inventarisatie leveren van de mate waarin de cliënt zich al dan niet voelt aangetrokken tot beroep(sactiviteiten)en. Al doende krijgt de cliënt een systematisch overzicht van wat er aan richtingen in de wereld van het werk "te koop" is.
- Tests kunnen cliënten bevestigen in hun ideeën, waardoor zij zekerder worden en tot uitvoering overgaan.
- Tests kunnen cliënten aan het twijfelen brengen als het gaat om niet goed onderbouwde plannen en daarmee kunnen slechte keuzen voorkomen worden.
- De loopbaanadviseur of de cliënt zelf kan aan de hand van een uitslag op geheel nieuwe ideeën komen, waardoor de cliënt op een nieuw spoor gezet wordt. Met name verrassende uitslagen kunnen nieuwe ideeën genereren, maar dit komt in verhouding minder vaak voor dan dat het *combineren* van verschillende, op zich niet verrassende resultaten tot nieuwe ideeën leidt. Bij cliënten die "vastzitten" in bepaalde denkpatronen waar zij zelf niet tevreden over zijn, kan dit zeer waardevol zijn.
- Tests kunnen licht werpen op de achtergronden van iemands twijfel. Soms kan de adviseur of cliënt tot een zinvolle interpretatie op een dieper niveau komen.
- Tests kunnen een situatie begrijpelijk maken, bijvoorbeeld waarom het slecht gaat in bepaalde vakken of waarom een cliënt geen zin heeft in mogelijkheid X.

De bijdrage van de test ligt dus niet zo zeer in het testresultaat *sec*, maar in het interpretatieproces in het licht van het loopbaanvraagstuk. De interpretatie komt tot stand in een dialoog, waarbij cliënt en adviseur ieder een eigen inbreng en verantwoordelijkheid hebben. De cliënt brengt veel deskundigheid mee rond de eigen persoon en situatie en is zo veel mogelijk zelf verantwoordelijk voor de uiteindelijke beslissingen.

De adviseur brengt deskundigheid in rond de gemeten begrippen, de kwaliteiten en tekortkomingen van de metingen en de mogelijke relaties met de wereld van opleidingen, werk en bezigheden. De adviseur is verantwoordelijk voor goede loopbaanadvisering. Dit betekent dat hij/zij er voor zorgt dat de betekenis van de testuitslagen onderschat, noch overschat wordt. Het grootste gevaar schuilt waarschijnlijk in cliënten die bang zijn voor (de verantwoordelijkheid voor) de te maken keuzen, die zich passief opstellen en een appel doen op de adviseur om als goeroe op te treden of als wijze wetenschapper die de waarheid in pacht heeft. Essentieel is dat de cliënt zelf denkt, actief verbanden legt en interpreteert en dat deze gestimuleerd wordt om verder te exploreren en (mede) verantwoordelijkheid te nemen voor de beslissingen in het loopbaanproces.

*Aanwijzingen voor het gebruik van tests in de praktijk van de loopbaanadvisering*¹⁶

Vóór de testafname:

- formuleer in overleg met de cliënt (en, in het geval dat niet dezelfde is, met de opdrachtgever) een zo duidelijk mogelijke vraag- en/of doelstelling
- zorg daarbij dat de cliënt (en opdrachtgever) reële verwachtingen hebben van wat de test vereist (tijd, inzet) en wat de test oplevert
- betrek de cliënt bij de beslissingen t.a.v. de af te nemen tests

Deze punten zijn van belang voor een optimale betrokkenheid van de cliënt bij de tests. Dit bevordert de kwaliteit van de testafname en vergroot de kans dat de cliënt de testresultaten goed zal verwerken en gebruiken.

Bij de bespreking van de resultaten:

- recapituleer de vraag-/doelstelling
- informeer hoe de cliënt de testafname heeft ervaren (in sommige gevallen, bijv. wanneer er belangrijke storende omstandigheden of ervaringen waren, zal men extra voorzichtig moeten zijn met het interpreteren van de uitslagen)
- leg de uitslagen uit; probeer hierbij:
 - de technische aspecten (bijv. het gebruik van een normgroep en de betekenis van een stanine-score) op een efficiënte en begrijpelijke manier duidelijk te maken
 - verbanden te leggen met andere informatiebronnen (merk je bij kaartlezen iets van je ruimtelijk inzicht? klopt het dat je weinig geeft om tuinieren? heb je ook op je werk moeite met cijfers?)
 - ruimte te geven voor verwerking van de uitslagen, zeker wanneer er sprake is van tegenvallende resultaten
 - steeds samen met de cliënt te bezien wat de uitslagen betekenen in het licht van de vraagstelling
- resumeer alle bevindingen en trek samen met de cliënt conclusies

Wellicht ten overvloede wordt hier nogmaals de aandacht gevestigd op het belang van een evenwichtige inbreng van beide gesprekspartners, de adviseur vanuit de positie van deskundige op het gebied van tests, psychologie en loopbanen, de cliënt als deskundige op het gebied van de eigen persoon en situatie. Een specifiek aandachtspunt voor de adviseur is dat de cliënt de betekenis van de uitslagen niet onderschat ("dat zegt toch niets dat ik bij de 5% meest dominanten uit de bus kom!"), maar ook niet overschat ("ik scoor beneden het gemiddelde, dus hoef ik die opleiding niet te proberen").

¹⁶ Met loopbaanadvisering wordt hier bedoeld op "zuivere" loopbaanadvisering, waarbij het belang van de cliënt minstens even zwaar meetelt als het belang van de opdrachtgever (indien deze een ander is dan de cliënt). De gegeven aanwijzingen zijn niet bedoeld voor situaties waarin het gaat om loopbaanadvisering in de vorm van (interne) selectie.

7. *Consequenties voor testconstructeurs en eisen aan testgebruikers*

Op de criteria van betrouwbaarheid en validiteit is in de afgelopen negentig jaar vooruitgang geboekt, maar de vooruitgang wordt steeds marginaler. Het "plafond" lijkt vrijwel bereikt te zijn. Hoe kunnen testconstructeurs er toe bijdragen dat tests in de loopbaanadvisering optimaal gebruikt worden? Hier volgen enkele suggesties:

- frequenter hernormeren; steun en stimulans bieden bij het verzamelen van lokale normgegevens
- meer en beter communiceren met gebruikers van testmaterialen (bijvoorbeeld een feedback-antwoordkaart in het testmateriaal opnemen, telefoontjes/brieven aanmoedigen, enquête onder testafnemers, studiedagen organiseren...)
- niet uitsluitend uitgaan van wetenschappelijke criteria bij het ontwikkelen van tests, maar ook meer praktische criteria in aanmerking nemen, bijvoorbeeld zijn de gehanteerde begrippen functioneel in de communicatie, stimuleert de test de cliënt tot exploreren?
- duidelijker handleidingen schrijven, ook leesbaar voor niet-wetenschappers (wat dit betreft zijn enkele auteurs met meer recente handleidingen op de goede weg, zie bijv. Evers en Lucassen, 1991; Compaan en Kooreman, 1995)
- bij uitslag- en profielformulieren de standaardmeetfout laten zien
- verder onderzoek en ontwikkelwerk met betrekking tot de problemen bij het testen van mensen uit diverse culturen
- verder werken aan theoretische kaders en operationalisaties, bijvoorbeeld met betrekking tot nieuwe ideeën over intelligentie (Gardner, 1985; Sternberg e.a., 1995), het persoon-situatie debat (in hoeverre wordt het gedrag van een persoon bepaald door de situatie?) en de cultuurbepaaldheid van tests
- verder uitwerken en implementeren van de item-respons-theorie (zie bijv. Eggen en Sanders, 1993) en van de mogelijkheden om te profiteren van de computer
- "verwachtingstabellen" ontwikkelen en aanbieden (dit zijn tabellen waaruit meer concrete gegevens over "kansen" zijn af te lezen, zodat uitspraken mogelijk worden als: "van de meisjes die met een zelfde uitslag op de capaciteitentest als jij aan de MTS begonnen, heeft 75% zonder doublures de eindstreep gehaald." - zie bijv. Bergland, 1974)

Voor een optimaal gebruik van tests in de loopbaanadvisering zal de testgebruiker over een aantal deskundigheden en vaardigheden moeten beschikken. Met name:

- inzicht hebben in de (discussies over de) begrippen die door tests gemeten worden (intelligentie, motivatie, interesse, faalangst, etc.)
- de ins en outs van normsystemen kennen en genormeerde scores in begrijpelijke termen aan de cliënt kunnen uitleggen
- weten hoe nauwkeurig tests meten en dit aan de cliënt kunnen overbrengen (bijvoorbeeld met behulp van het begrip standaardmeetfout)
- weten wat testresultaten wel en niet kunnen betekenen
- uiteenlopende gesprekstechnieken kunnen hanteren, waaronder uitleggen, corrigeren (van verkeerde interpretaties van testuitslagen), stimuleren, non-directief counselen en (directief) adviseren of confronteren, bijvoorbeeld om een passief/receptieve houding van een cliënt te doorbreken

Essentieel hierbij is het houdingsaspect: de cliënt accepteren als deskundige en verantwoordelijke op het gebied van eigen persoon, situatie en leven.

8. *Vergelijking van tests met andere beoordelingsmiddelen*

Alvorens een conclusie te kunnen trekken over de waarde van tests bij de loopbaanadvisering moet de aandacht gericht worden op enkele alternatieve beoordelingsmethoden.

Statistische predictie (voorspellen met behulp van objectieve tests) werkt beter dan de klinische blik van de deskundige. Dat is - ook al kunnen veel deskundigen dat maar moeilijk accepteren - herhaaldelijk overtuigend vastgesteld (zie bijv. Ter Laak, 1995).

In zijn algemeenheid gesproken is er slechts één beoordelingsmethode die de vergelijking met tests kan doorstaan, namelijk de "assessment-center-methode" (zie bijv. Bloemers, 1996 of Luken, 1996). De ACM is een beoordelingsmethode waarbij cliënten/kandidaten in nagebootste praktijksituaties opdrachten uitvoeren en waarbij hun gedrag volgens een bepaalde systematiek wordt geobserveerd door getrainde "assessoren". Wanneer het gaat om het beoordelen van verstandelijke capaciteiten, dan werkt assessment ongeveer even goed als capaciteitentests. De validiteiten liggen althans ongeveer op hetzelfde niveau. Wanneer het gaat om de beoordeling van persoonlijke en/of sociale vaardigheden, dan doet assessment het zelfs beter dan tests. Er valt veel te zeggen voor de combinatie van tests met de ACM. Nadeel van assessment zijn het tijdsbeslag en de daarmee gemoeide kosten.

Vergeleken met alle andere bekende beoordelingsmethoden (gesprekken, gestructureerde interviews, brieven, curricula vitae, referenties, antecedentenonderzoek, grafologie, schooltoetsen, proefwerken, tentamens en beoordelingsformulieren) is de test steeds superieur. De betrouwbaarheden en validiteiten van tests vallen ook bij vergelijkingen op andere terreinen (bloeddrukmeting, oogmeting, arbeidsongeschiktheidsbepaling, weersvoorspelling...) bepaald niet uit de toon. Ook op zogenaamde hardere terreinen als de economie valt voorspellen niet mee. Zo viel in de Volkskrant van 19 juli 1996 te lezen: "Indien het Centraal Plan Bureau een dollarkoers voor het volgende jaar van 1,70 gulden voorspelt, dan is er 67 procent kans dat de werkelijke koers van de dollar volgend jaar tussen de 1,37 gulden en 2,02 komt te liggen." Voorspellen is kennelijk nu eenmaal moeilijk en dat beperkt zich niet tot het psychologische vlak.

Als meet- of beoordelingsinstrument functioneert de test dus behoorlijk goed in vergelijking tot andere instrumenten. Resteert de vraag: hoe groot moet de rol van een meetinstrument zijn bij de loopbaanadvisering?

9. *De waarde van tests bij loopbaanadvisering: slotbeschouwing, balans en conclusie*

Wanneer men een conclusie probeert te trekken over de waarde van tests bij de loopbaanadvisering, dan is het van belang om onderscheid te maken tussen het nut voor het betrokken individu en het nut voor "anderen". Onder "anderen" kan hier worden gelezen: school, arbeidsorganisatie, samenleving, wetenschap...

Tests zijn een rationeel en rendabel hulpmiddel, wanneer het gaat om (grote) groepen. De successtory van de test begon tijdens de eerste wereldoorlog in de Verenigde Staten, toen grote groepen (miljoenen) rekruten, die voor een groot deel geen diploma's hadden waaruit hun niveau viel af te leiden, snel naar een geschikte functie in de krijgsmacht moesten worden gealloceerd. Dit ging met tests veel beter dan zonder tests. Ook bij selectie is testgebruik rendabel, althans wanneer het om groepen gaat. Tien procent minder fouten bij

het selecteren van personeel levert een bedrijf op termijn een aanzienlijke kostenbesparing op. Ook in de wetenschap van de psychologie bewijzen tests hun nut omdat men er (cor)relaties tussen en met psychische hoedanigheden mee vast kan stellen.

Loopbaanadvisering is (tegenwoordig) echter iets heel anders dan selectie. Leren leren, leren kiezen, zelfsturing op gang brengen - daar gaat het vooral om. Tests kunnen in dat proces een bescheiden maar nuttige rol spelen, mits op de juiste manier gebruikt. Onjuist gebruik kan tot persoonlijke en maatschappelijke schade leiden.

De belangrijkste van de in dit hoofdstuk behandelde positieve en negatieve aspecten van tests worden hier in een balans gezet.

POSITIEF	NEGATIEF
<ul style="list-style-type: none"> - objectief - vergelijking met andere mensen mogelijk - in vergelijking met andere beoordelingsmiddelen behoorlijk goede betrouwbaarheid en validiteit - wetenschappelijke onderbouwing - toezicht COTAN - in verhouding efficiënt (althans voor de adviseur/adviserende instantie) - biedt mogelijkheden voor structurering van het begeleidingsgesprek, interpretatie van vraag en situatie, zelfconceptverheldering, nieuwe ideeën opdoen, doorbreken van twijfels 	<ul style="list-style-type: none"> - onnauwkeurigheid in de zin van vrij grote foutenmarges - zeer geringe voorspellende potentie - slechte tot twijfelachtige toepasbaarheid bij "bijzondere" doelgroepen of van de norm afwijkende individuen; mogelijkheid discriminatie - mogelijkheid van stigmatisering - veel aspecten die doorslaggevend zijn, zijn niet goed te meten - vaak verouderd bij gebrek aan onderhoudsinvesteringen - gevaar dat door de wetenschappelijke uitstraling te veel waarde wordt gehecht aan de uitslag

Tabel 2: Balans tussen positieve en negatieve aspecten van tests

Lezers zullen zelf moeten uitmaken hoe zwaar zij de verschillende elementen in de balans willen laten wegen. Een mogelijke conclusie zou als volgt kunnen luiden.

Tests zijn gebrekkige meetinstrumenten, maar betere zijn niet of nauwelijks voorhanden. Wat we met tests kunnen meten, beslaat slechts een klein deel van alle aspecten die belangrijk zijn bij loopbaankeuzen. In een zee van onzekerheid geven zij echter een beetje houvast en - mits op de juiste wijze gebruikt - een kans om verder te komen in het keuze- en leerproces. Een gevaar is dat men (zowel loopbaanadviseur als cliënt kunnen aan deze verleiding blootstaan) zich te veel aan dit houvast vastklampt, m.a.w. dat men schijnzekerheid verkiest boven onzekerheid. Weten is meer dan meten.

Literatuur

Admiraal-Hilgeman, Th.J. De waarde van psychodiagnostisch testgebruik bij studie- en beroepskeuze. In: Spijkerman R.M.H. e.a. (red.): *Handboek Studie- en Beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1991, p. 1450-1 t/m 24.

Admiraal-Hilgeman, Th.J. Plaatsbepaling van het begrip intelligentie en intelligentietests binnen de studie- en loopbaanbegeleiding. In: Spijkerman R.M.H. e.a. (red.): *Handboek Studie- en Beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1994(a), p. 1550-1 t/m 22.

Admiraal-Hilgeman, Th.J. Plaatsbepaling van het begrip persoonlijkheid en persoonlijkheidstests binnen de studie- en loopbaanbegeleiding. In: Spijkerman R.M.H. e.a. (red.): *Handboek Studie- en Beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1994(b), p. 1600-1 t/m 23.

Akkerman, A.E. en Buijk, C.A. *GLTS '94; Handleiding*. Swets & Zeitlinger, Lisse, 1994.

Akkerman, A.E. Equivalentie als uitdaging! *Psychologie en Computers*. Jrg. 11, nr. 1, mrt 1994.

Bergland, B.W. Career Planning: The use of sequential evaluated experience. In E.L. Herr (ed): *Vocational guidance and human development*. Houghton Mifflin Company, Boston, 1974, p. 350-380.

Bloemers, W. Het selectieproces. In: Kooreman, A. en Breed, A. (red.): *Psychologie in arbeid en loopbaan*. Swets & Zeitlinger, Lisse, 1996.

Compaan, E.L. en Kooreman, A. *Korte Interesse Test*. Swets & Zeitlinger, Lisse, 1995.

Drenth, P.J.D. en Sijtsma, K. *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen*. Bohn Stafleu Van Loghum, Houten/Zaventem, 1990.

Eggen, T.J.H.M. en Sanders, P.F. *Psychometrie in de praktijk*. Cito, Arnhem, 1993.

Evers, A., Lucassen, W. en Wiegersma, S. *BIT Beroepen-interesstest*. Wolters-Noordhoff, Groningen, 1987.

Evers, A. en Lucassen, W. *DAT '83: Handleiding*. Swets & Zeitlinger, Lisse, 1991.

Evers, A. *Amsterdamse Beroepen Interesses Vragenlijst ABIV 92: Handleiding*. Swets & Zeitlinger, Lisse, 1992

Gardner, H. *Frames of Mind: The theory of Multiple Intelligences*. Paladin Books, London, 1985

Inspectie van het Onderwijs en Inspectie van de Beroepskeuzevoorlichting. *Beroepskeuzevoorlichting tussen wal en schip: Een evaluatie van de Regionale Diensten Centra voor studie- en beroepskeuzevoorlichting*. Zwolle/Rijswijk, 1995.

Janssen, G., F. Meijers, R. Spijkerman en G. Wijers. *Het AOB in de Informatiemaatschappij*. Landelijk Dienstverlenend Centrum, Leeuwarden, 1993.

Laak ter, J.J.F. *Psychologische diagnostiek: Inhoudelijke en methodologische grondslagen*. Swets & Zeitlinger, Lisse, 1995

Lofquist, L.H. and England, C.W. *Problems in vocational counseling*. Minnesota, 1961.

Luken, T.P. Zelfconceptverheldering. In: Spijkerman R.M.H. e.a. (red): *Handboek Studie- en Beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1990, p. 1050-1 t/m 25

- Luken, T.P. Interesse en interessemeting. In: Spijkerman R.M.H. e.a. (red.): *Handboek studie- en beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1995, p. 1750-1 t/m 27
- Luken, T. Het assessment raadsel: het werkt, maar hoe? In Kooreman, A. en Breed, A. (red.): *Psychologie in arbeid en loopbaan*. Swets & Zeitlinger, Lisse, 1996.
- Luteijn, F., Starren, J. en Van Dijk, H. *Nederlandse Persoonlijkheidsvragenlijst; Handleiding*. Swets & Zeitlinger, Lisse, 1979.
- Magnusson, D. *Test Theory*. Addison-Wesley, Reading (Massachusetts) etc. 1967.
- Meijers, F. *Arbeidsidentiteit: studie- en beroepskeuzebegeleiding in de post-industriële samenleving*. Samsom HD Tjeenk Willink, Alphen aan den Rijn, 1995.
- Minden, J.R. van. *Alles over psychologische tests*. Veen, Utrecht/Antwerpen, 1987).
- Nederlands Instituut van Psychologen. *Richtlijnen voor ontwikkeling en gebruik van psychologische tests en studietoetsen*. Nederlands Instituut van Psychologen, Amsterdam, 1978 en 1988.
- Nederlands Instituut van Psychologen. *Documentatie van tests en testresearch in Nederland*, Amsterdam, 1992 (plus regelmatig verschijnende supplementen)
- Pere, H.M. *Tussen arbeidsmarkt en individueel welzijn; een historische analyse van de beroepskeuzevoorlichting in Nederland vanuit beroepsociologisch perspectief*. Educaboek, Culemborg, 1986.
- Roe, Anne. *The Psychology of Occupations*. Wiley, New York, 1956.
- Spitz, J.C. *Statistiek voor psychologen, pedagogen, sociologen*. Noord-Hollandsche Uitgevers Maatschappij, Amsterdam, 1968.
- Sternberg, R.J., Wagner, R.K., Williams, W.M. en Horvath, J.A. Testing Common Sense. *American Psychologist*, vol. 50 (1995), nr. 11, p. 912-927.
- Taborsky, O. Zelfconcept, zelfconceptverheldering en realiteitszin. *Handboek Leerlingbegeleiding*. Samsom H.D. Tjeenk Willink, Alphen aan den Rijn, 1987, p. 2220-1 t/m 17.
- Tjoa, S.S.H. *Edwards Personal Preference Schedule; Herziene Handleiding*. Swets en Zeitlinger, Lisse, 1993.
- Vincken, A.J.R. De schooldekaan, zijn rol, plaats en functie. In: Spijkerman R.M.H. e.a. (red.): *Handboek studie- en beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1996(b), p. 4300-1 t/m 31.
- Vincken, A.J.R. Integratie van keuzebegeleiding. In: Spijkerman R.M.H. e.a. (red.): *Handboek studie- en beroepskeuzebegeleiding*. Samsom H.D. Tjeenk Willink, Alphen a/d Rijn, 1996(b), p. 4340-1 t/m 23.
- Vroon, P. *Intelligentie*. Ambo, Baarn, 1980.
- Zeeuw, J. de. *Algemene psychodiagnostiek I: Testmethoden*. (6de druk) Swets & Zeitlinger, Lisse, 1986.

Bijlage 1:

Betekenis van afkortingen van tests, gebruikt bij loopbaanadvisering

AAT	Algemene Administratie Test
ABIV	Amsterdamse Beroepen Interesses Vragenlijst
ABV	Amsterdamse Biografische Vragenlijst
ABV-K	Amsterdamse Biografische Vragenlijst voor Kinderen
Albert-J	Alphense Beroepentest voor LBO/IBO - Jongensversie
Albert-M	Alphense Beroepentest voor LBO/IBO - Meisjesversie
APT	Applikatie Programmeur Test
BNT	Basis Niveau Test
BIO	Beroepen Interesse Onderzoek
BIT	Beroepen Interesse Test
BKT-M	Beroepskeuzetest - Middelbaar niveau
BKT-H	Beroepskeuzetest - Hoger niveau
BZO	Beroepskeuze Zelf Onderzoek
DAT	Differentiële Aanleg Test
EPPS	Edwards Personal Preference Schedule
ETAV	Examen/Toets Attitude Vragenlijst
GALO	Groninger Afsluitingsonderzoek Lager Onderwijs
GIT	Groninger Intelligentie Test
GLTS	Guilford-LTP Temperament Survey
ISI	I Intelligentie Schoolvorderingen en Interesse
KIT	Korte Interesse Test
MCT-M	Multiculturele Capaciteiten Test - Hoger Niveau
MCT-M	Multiculturele Capaciteiten Test - Middelbaar Niveau
MORINI	Motivatie Richting Niveau
MPT	Multiculturele Persoonlijkheidstest
NAT	Numerieke Aanleg Test
NPV	Nederlandse Persoonlijkheidsvragenlijst
NPV-J	Nederlandse Persoonlijkheidsvragenlijst voor Jongeren
PM	Progressieve Matrijzen
PMT	Prestatie Motivatie Test
PMT-K	Prestatie Motivatie Test voor Kinderen
SCIO	Strategic Consultancy in Organisations
SIG	Schaal voor Interpersoonlijk Gedrag
SIW	Schaal voor Interpersoonlijke Waarden
SVL	School Vragen Lijst
SSV	School en Studie Vragenlijst
TBT	Thurstone Beroepen Test
TNVA	Test voor Niet-Verbale Aanleg
VAT	Verbale Aanleg Test
VCI	Vragenlijst voor Commercieel Inzicht
VSV	Vragenlijst Studievoorwaarden
WAIS	Wechsler Adult Intelligence Scale
WISC-R	Wechsler Intelligence Scale for Children - Revised
ZA	Zinnen Aanvullen

Bijlage 2:

Betekenis van andere gebruikte afkortingen

ACM	Assessment Center Methode
AOB	Adviesbureau voor Opleiding en Beroep
COTAN	Commissie Test Aangelegenheden van het NIP
NIP	Nederlands Instituut van Psychologen