

Testen in het onderwijs

Nut, gevaren en voorwaarden

Tom Luken¹

Dit artikel gaat over het gebruik van tests in het onderwijs. Wat is het mogelijk nut ervan? Welke gevaren zijn eraan verbonden? Onder welke voorwaarden kan men het nut maximaliseren en de gevaren vermijden?

Deze vragen zijn actueel gezien de snel toenemende mogelijkheden van online testen via internet². Een tweede reden voor de actualiteit van het onderwerp ligt in ontwikkelingen in het onderwijs. In het algemeen voortgezet onderwijs en nog sterker in het beroepsonderwijs zijn van hoog tot laag loopbaanontwikkeling en zelfsturing centrale thema's geworden. In dit artikel zal worden betoogd dat juist binnen deze ontwikkelingen tests een bijzondere bijdrage kunnen leveren.

Alvorens in te gaan op de gestelde vragen, moeten eerste enkele andere vragen beantwoord worden. Wat zijn tests? En wat zijn hun sterkte punten en beperkingen?

Wat zijn tests?

Het Nederlands Instituut van Psychologen (NIP) definieert een test als *“een verzameling opdrachten of vragen, waarmee men bepaalde soorten gedrag onder gestandaardiseerde omstandigheden wil oproepen om zo vergelijking tussen personen mogelijk te maken.”*

In deze definitie zijn twee basiskennmerken van tests opgenomen. In de eerste plaats standaardisatie. De test moet voor de ene persoon hetzelfde zijn als voor een ander en moet zo veel mogelijk onder dezelfde omstandigheden worden afgenomen (goed licht, geen verstoringen, zelfde instructies, niet de één meer hulp geven dan de ander, etc.) Pas als de test gestandaardiseerd is komt het tweede basiskennmerk goed tot zijn recht. Een test heeft 'normen', met andere woorden er is minimaal één normgroep, waardoor men (via een normstelsel) kan zeggen in hoeverre de uitslag van een persoon hoog, laag of gemiddeld is.

¹ Tom Luken is arbeids- en organisatiepsycholoog en is als adviseur bij loopbaan- en beoordelingsvraagstukken werkzaam te Amsterdam bij Stichting NOA en Luken Loopbaan Consult. E-mail adres: t.luken@noa-vu.nl. Een ouder, maar uitgebreider artikel over hetzelfde onderwerp met meer wetenschappelijke onderbouwing is te vinden op <http://home.planet.nl/~tluken/tests.htm>

² In de afgelopen maanden zijn minstens drie sites in Nederland online testen gaan aanbieden: <http://www.noa-vu.nl/> (klik op NOA Online testen), <http://www ldc.nl/content/portal/portal.asp> (klik op Tests) en <http://www.123test.nl/> (klik op Zakelijk Testgebruik).

Beide bovenstaande kenmerken kan men beschouwen als minimumeisen. Als er niet aan is voldaan, kan men niet van een test spreken.

Daarnaast bestaan er de volgende kwaliteitseisen. Als daaraan is voldaan kan men van een *goede* test spreken.

1. Een helder gebruiksdoel en stevige theoretische basis (goed gedefinieerde begrippen)
2. Een goede testuitvoering (bijv. een objectief scoringssysteem, geen racistische of anderszins kwetsende inhoud, heldere handleiding)
3. Betrouwbaarheid: de test meet zo nauwkeurig mogelijk, met andere woorden de rol van het toeval is zo klein mogelijk
4. Validiteit: de test meet werkelijk wat hij pretendeert te meten. Of: de test beantwoordt aan zijn doel.

Anekdote over het verschil tussen betrouwbaarheid en validiteit.

Francis Galton probeerde tegen het eind van de negentiende eeuw als een van de eersten een intelligentietest te maken. Hij ging uit van het idee dat intelligentie gedefinieerd kan worden als het kunnen opnemen van informatie. Informatie opnemen gaat via de zintuigen, redeneerde hij. 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? Etc. Hoe beter je waarneemt, des te beter 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....

In Nederland beoordeelt de COTAN (Commissie Testaangelegenheden van het NIP³) de kwaliteit van tests.

Van goede tests kan men stellen, dat het efficiënte, nauwkeurige en valide meetinstrumenten zijn van relevante psychische kenmerken. Zij kunnen een relatief scherp en betrouwbaar beeld geven van kenmerken als cognitieve capaciteiten, beroepsinteressen, persoonlijkheidskenmerken (bijv. extraversie), motivatie en leerstijl.

Beperkingen van tests

Dat klinkt mooi, de slotzin van de vorige paragraaf. Maar het is niet allemaal rozegeur en maneschijn. Lang niet alle tests die in Nederland gebruikt worden, zijn beoordeeld door de COTAN. De tests die wel beoordeeld zijn, halen lang niet altijd op alle onderdelen voldoende. Kortom: lang niet alle tests zijn goed. Maar zelfs als tests wel goed zijn, wat betekenen dan de begrippen 'betrouwbaarheid' en 'validiteit' concreet?

Ook bij betrouwbare tests speelt toeval een rol, bijvoorbeeld ten gevolge van raden. Stel dat een IQ test een betrouwbaarheid heeft van '.93'. Deze krijgt van de COTAN zonder meer de beoordeling 'goed'. Hoe nauwkeurig meet deze test? Wat betekent dit bijvoorbeeld voor een leerling die een IQ-score van 110 heeft behaald?

Uit het betrouwbaarheidsgetal kan men een 'standaardschattingsfout' afleiden. Deze bedraagt in dit geval vier IQ-punten. Dit betekent, dat de 'ware score' van de betreffende leerling

³ Zie www.cotan.nl

waarschijnlijk (68% kans) ligt tussen 106 en 114. Hoogstwaarschijnlijk (95% kans) ligt de ware score tussen 102 en 118. Met andere woorden: van de 100 leerlingen die een score van 110 behalen, zijn er vijf die een ware score onder de 102 of boven de 118 hebben. Als men zich nu bedenkt dat 102 een normaal IQ-niveau is voor het VMBO en 118 voor het VWO... Conclusie: zelfs bij tests met een betrouwbaarheid die goed wordt geacht, gaat het meten niet erg nauwkeurig.

En hoe staat het met de validiteit? Bij het bepalen van de validiteit gaat men dikwijls uit van de correlaties van de test met eerder gemaakte tests. Met andere woorden: als de scores van mensen op een nieuwe test, bijvoorbeeld voor ruimtelijk inzicht, duidelijk samenhangen met de scores van dezelfde mensen op een bestaande test, dan beschouwt men dit als een argument voor de validiteit van de nieuwe test. Nadeel van deze benadering is, dat men eeuwig kan blijven voortborduren op gemaakte fouten. Als de eerdere test 'niet klopt', klopt de nieuwe evenmin, terwijl toch, omdat ze 'met elkaar kloppen' de validiteit positief wordt beoordeeld.

Een moeilijkere proef voor validiteit is de voorspellende variant. Als bijvoorbeeld de scores op een in 2006 afgenomen intelligentietest samenhangen met de schoolresultaten in 2009, dan spreekt men van voorspellende validiteit. Persoonlijkheidstests en interesstests hebben in het algemeen niet zo veel voorspellende waarde. De voorspellende waarde van cognitieve capaciteitentests is daarentegen behoorlijk. Echter: ook hier geldt dat toekomstige prestaties slechts voor een klein deel voorspeld kunnen worden met tests. De reden is dat allerlei factoren (bijv. motivatie, gezondheid en ontmoetingen) medebepalend zijn voor succes.

Het mogelijke nut van tests

Het is een paradox. Of iemand goed zal presteren in zijn studie, goed zal functioneren in haar baan of tevreden zal zijn in het leven is niet of nauwelijks te voorspellen op basis van testresultaten. En toch bestaan er geen betere voorspellers. Alleen de assessment center methode⁴ doet het soms nog iets beter, met name als het gaat om sociale en communicatieve vaardigheden. Maar verder, of men nu interviewtechnieken gebruikt, kennistoetsen, grafologie, astrologie, antecedentenonderzoek of het natrekken van referenties: tests doen het vrijwel altijd beter. Kortom: de meetnauwkeurigheid en voorspellende waarde van tests zijn beperkt, maar betere meetinstrumenten van psychische kenmerken zijn niet voorhanden. Het is dan ook niet verwonderlijk dat tests vaak worden ingezet bij selectiebeslissingen. Er bestaat duidelijk bewijs dat als men 50 kandidaten moet kiezen uit 100 sollicitanten, tests een nuttige en rendabele rol kunnen spelen. Ook in het onderwijs kan dit aan de orde zijn, bijvoorbeeld bij de selectie voor opleidingen met een 'numerus fixus' of bij de toelating tot verkorte trajecten.

In dit artikel gaan we nader in op de rol die tests kunnen spelen bij het 'nieuwe leren'. Over welke vorm van nieuw leren men het ook precies heeft, het veronderstelt altijd dat de leerling of student zelf een intrinsieke motivatie ontwikkelt en zelf keuzes maakt ten aanzien van lange termijn doelen (arbeidsidentiteit) en korte termijn leerobjecten. Om dit te kunnen is zelfkennis nodig. Voor volwassenen is dit al moeilijk, voor jongeren geldt dit nog sterker. Veel leerlingen in voortgezet en beroepsonderwijs missen zelfkennis. Zij hebben geen helder, realistisch en stabiel beeld van wat zij kunnen en evenmin van wat zij willen. Toekomstperspectieven veranderen steeds of zijn afwezig. Velen missen intrinsieke motivatie

⁴ Hier wordt bedoeld: methodische observatie en beoordeling van de uitvoering van realistische opdrachten in gesimuleerde of reële werksituaties.

en hun gedrag is calculerend. Leerloopbanen zijn vaak ontstellend richtingloos en inefficiënt. Er is veel verspilling van talent, levensvreugde en geld.

Veel leerlingen hebben dus hulp nodig bij het ontwikkelen van hun loopbaan. Er zijn verschillende manieren waarop effectief hulp gegeven kan worden⁵. Blijkens recent onderzoek in het beroepsonderwijs⁶ ontvangen jongeren hier echter slechts sporadisch de noodzakelijke hulp. Onder meer blijkt dat slechts 11% (volgens de leerlingen) tot 13% (volgens de docenten) een test krijgt. Ongeveer 88% van de leerlingen wordt dus ‘het beste meetinstrument voor psychische kenmerken dat voorhanden is’ onthouden.

Hoe zouden tests kunnen bijdragen aan de zelfkennis die nodig is voor het nieuwe leren? Mensen in het algemeen, maar wellicht adolescenten in het bijzonder, zijn doorgaans zeer geïnteresseerd in informatie waarmee zij zichzelf beter kunnen leren kennen en begrijpen en zichzelf kunnen vergelijken met anderen. Mening en gedachten genoeg, maar concrete, objectieve gegevens zijn hierbij moeilijk te vinden. Tests kunnen op dit punt een belangrijke bijdrage leveren. Zij zijn op het gebied van zelfkennis een van de weinige leveranciers van betrouwbare feiten. Bijvoorbeeld bij een interestest: *“Vergeleken met anderen heb je vaak ja gezegd op vragen die te maken hebben met elektrotechniek.”* Of bij een capaciteitentest: *“Ik zit bij het slechtste kwart van de Nederlandse bevolking als het gaat om Woordenschat”*. Gezien de kwaliteiten van tests kan geconstateerd worden dat dergelijke bevindingen niet op toeval berusten.

Tests bieden dus zeldzame feitelijke informatie. Zij bieden een spiegel die een verhoudingsgewijs helder en realistisch beeld geeft. Daarnaast bieden zij een begrippenkader dat het resultaat is van meer dan een eeuw psychologisch onderzoek. Dit heeft begrippen opgeleverd die vooralsnog aanzienlijk beter gedefinieerd en stabiel zijn dan veel begrippen die gehanteerd worden in competentiegericht onderwijs. Een interestest biedt bijvoorbeeld niet alleen inzicht in de eigen belangstelling, maar tegelijk een overzicht over de mogelijkheden in de beroepenwereld en hoe die vanuit psychologisch oogpunt van elkaar verschillen. Een persoonlijkheidstest biedt niet alleen objectief vergelijkingsmateriaal, maar ook begrippen waarmee je goed over je eigen gedrag (in vergelijking tot anderen) kunt nadenken. Bij het opbouwen van zelfkennis in relatie tot de wereld kunnen tests daarom een waardevolle bijdrage leveren.

Gevaren van tests

Verwarring van institutioneel en individueel perspectief

Testgebruik vond tot voor kort doorgaans plaats vanuit een institutioneel perspectief. Dat wil zeggen dat het belang van de onderwijsorganisatie voorop stond: deze wilde leerlingen selecteren of adviseren.

Als tests gebruikt worden in het kader van zelfkennis en zelfsturing kan men beter vertrekken vanuit een puur individueel perspectief: het belang van de individuele leerling staat voorop.

⁵ Luken, T. (2006). Zelfkennis als fundament voor loopbaanontwikkeling? Bouwen op een onbetrouwbare ondergrond. In: Huijgevoort, H. van, Admiraal, A. & Vloet, K. (red.). *Zelfkennis in loopbaanontwikkeling*, p. 11-26. Antwerpen/Apeldoorn: Garant.

⁶ Zie bijvoorbeeld Luken, T. & Newton, I. (2004). *Loopbaanbegeleiding bij de doorstroom van MBO naar HBO*. Onderzoeksrapport in opdracht van het Platform Beroepsonderwijs. Amsterdam: NOA, 2004. Of voor een grootschaliger en recenter onderzoek: Meijers, F., Kuijpers, M. & Bakker, J. (2006). *Over leerloopbanen en loopbaanleren. Loopbaancompetenties in het (v)mbo*. Concept rapport Het Platform Beroepsonderwijs, februari 2006.

Er dreigt gevaar als niet een duidelijke keus gemaakt wordt tussen beide perspectieven. Wanneer de organisatie bijvoorbeeld de leerling dwingt om ‘voor eigen bestwil’ mee te doen aan een test, dan wordt deze direct met een moeilijke paradox geconfronteerd: ik *moet* een test doen voor mijn zelfsturing. ‘Rare’ uitslagen kunnen het gevolg zijn. Als de leerling er niet zeker van is, dat hij zelf ‘eigenaar’ zal zijn van de uitslagen, kan dit evenzeer de kwaliteit van de testafname verstoren, omdat hij het gevoel heeft dat school, ouders en/of klasgenoten tijdens de testafname ‘over zijn schouder meekijken’.

Testuitslagen verkeerd interpreteren en self-fulfilling prophecies

Bij de interpretatie van testuitslagen kan veel mis gaan. Bijvoorbeeld als men er zonder meer van uit gaat dat een score van iemand uit een etnische of culturele minderheidsgroep hetzelfde betekent als die van een autochtoon.

In een waar gebeurd verhaal zag een eerstejaarsstudent op het bureau van zijn studieadviseur terloops zijn uitslag op een capaciteitentest. Hij zag het getal 98 en maakte daaruit op dat hij een IQ van 98 had. In werkelijkheid was hij in het 98-ste percentiel uitgekomen (hetgeen correspondeert met een IQ van 140). Hij zakte bijna voor het eerste trimester en zei tegen zijn ouders: *"Ik geloof niet dat ik de universiteit aankan"*. Ze gingen samen naar de studieadviseur die uitlegde wat in werkelijkheid zijn testuitslag was geweest. De student beëindigde het jaar met de kwalificatie 'uitmuntend'.

Talloos zijn de verhalen van testscores die verkeerd worden geïnterpreteerd. Bijvoorbeeld dat interesse verward wordt met geschiktheid. Of dat een zogenaamde staninescore van 5 als onvoldoende wordt opgevat (in plaats van ‘gemiddeld’). Zeker als ook het volgende gevaar aan de orde is, kunnen de consequenties ernstig zijn.

Te veel belang hechten aan testuitslagen

Kiezen bij belangrijke stappen in het leven is vaak moeilijk. Het kan verleidelijk zijn de verantwoordelijkheid ervoor te ontlopen. Mede door het wetenschappelijke aura van tests bestaat het gevaar dat te veel belang wordt gehecht aan testuitslagen. In zekere zin kiest de test dan voor je (en kun je die ook de schuld geven als het mis gaat.) Iemand met een IQ van 110 zegt bijvoorbeeld: *“Mijn IQ ligt duidelijk beneden het gemiddelde in het wetenschappelijke onderwijs. Dat betekent dat een universitaire studie te moeilijk voor mij zal zijn.”* Deze persoon bedenkt zich niet dat er talloze mensen met een IQ van 110 succesvol aan universiteiten studeren. Of een ander zegt: *“Mijn hoogste score bij de interestetest is op administratief werk, dus kan ik het best die kant op gaan.”* Deze persoon vergeet onder andere dat interesse slechts één factor dient te zijn bij loopbaankeuze. Een laatste voorbeeld: *“Ik kom bij de persoonlijkheidstest uit als weinig consciëntieus, daarom kan ik maar beter geen studie of soort werk kiezen, waarbij zorgvuldigheid en doorzetten belangrijk zijn.”* Hij vergeet dat mensen binnen zekere grenzen kunnen veranderen, vooral als ze daarvoor kiezen. Bovendien betekent de lage score niet automatisch dat hij niet consciëntieus is, maar dat hij zichzelf als zodanig beschreven heeft.

Voorwaarden aan testgebruik in het onderwijs

Als tests in het onderwijs worden ingezet, zonder dat aan een aantal voorwaarden is voldaan, zullen de opbrengsten in het gunstige geval gering, maar in het ongunstige geval schadelijk zijn. Uit ervaringen bij projecten waarbij testgebruik op grote schaal is ingevoerd bij twee grote organisaties voor beroepsonderwijs, kunnen de volgende voorwaarden worden afgeleid:

- Goede voorlichting vooraf: het moet de leerling of student glashelder zijn wat het doel van de test is en hoe en aan wie rapport zal worden uitgebracht. Ook moet globaal

duidelijk zijn wat de inhoud van de tests is en wat men er wel en niet van kan verwachten. Misverstanden hierover moeten zo veel mogelijk vooraf al worden tegengegaan.

- Gebruik van kwalitatief goede tests (zie ook de kwaliteitseisen onder het kopje ‘Wat zijn tests?’).
- Goede afnameprocedures (betrouwbare hard- en software, ongestoorde ruimte etc.)
- Verantwoorde en duidelijke rapportage.
- Procedures die garanderen dat testgegevens niet in handen komen van onbevoegden.
- Recht op bespreking van de uitslagen met een competente gesprekspartner. Dit hoeft niet per se een psycholoog te zijn. Wel moet bedacht worden dat de docent, coach, decaan of studieloopbaanadviseur in staat moet zijn om testuitslagen (inclusief de technische details!) goed uit te leggen en om op een productieve manier te communiceren over de betekenis ervan voor het zelfbeeld en de plannen van de deelnemer/student. Zij moet de leerling stimuleren om zelf actief de testuitslagen te interpreteren en ermee aan de slag te gaan, maar ook eventuele foute interpretaties signaleren en corrigeren. Zij moet inzicht hebben in de gemeten begrippen en in zaken als standaardschattingfouten en normsystemen. Aan de hand hiervan moet zij precies kunnen uitleggen wat wel en wat niet uit testcores kan worden afgeleid.
- Follow-up in het vervolg van de studieloopbaanbegeleiding (bijvoorbeeld in reflectiemethode).

Testgebruik dient te zijn ingebed in de organisatie en in het onderwijs-/leerproces. Als aan de voorwaarden is voldaan, kan de opbrengst in termen van zelfkennis, zelfsturing en daadwerkelijke competentieontwikkeling hoog zijn.