



Interpretatie van klinische resultaten

HET INTERPRETEREN VAN DE RESULTATEN VAN
CLIËNTEN UIT EEN KLINISCHE GROEP

White paper

Interpretatie van klinische resultaten

Het interpreteren van de resultaten
van cliënten uit een klinische groep

J.E. Dek MSc

Product Developer Educational Psychology, Pearson Assessment and Information B.V.

M.S.L. Braak MSc

Product Developer Educational Psychology, Pearson Assessment and Information B.V.

Drs. A.P. Kooij

Uitgever, Pearson Assessment and Information B.V.

PEARSON

Inhoud

1	Inleiding	5
2	Normgerichte interpretatie	6
3	Criteriumgerichte interpretatie	9
4	Valideringsonderzoek	10
5	Conclusie	11
	Referenties	12
	Summary	13

1 Inleiding

Voor elk psychologisch meetinstrument rijst de vraag in hoeverre het instrument bruikbaar is bij klinische doelgroepen. Voor welke doelgroep is het instrument geschikt en welke problemen komt men tegen bij het afnemen van het instrument bij een persoon met een stoornis of beperking. Kan het instrument bij hen op gestandaardiseerde wijze afgenomen worden en wat zeggen de resultaten over een individu met een specifieke problematiek?

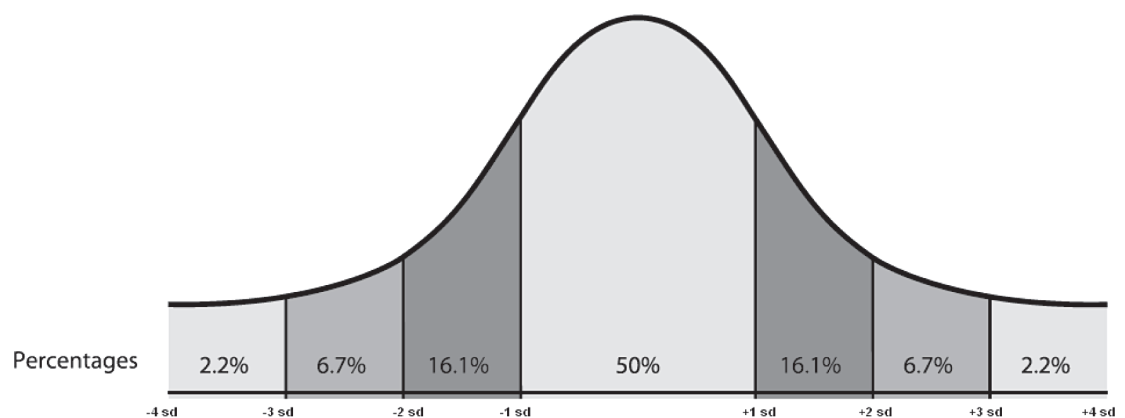
Het is bij het testen van een cliënt uit een klinische groep van belang om een onderscheid te maken tussen een normgerichte en een criteriumgerichte interpretatie van de resultaten. Bij normgerichte interpretatie wordt de ruwe score vergeleken met de scoreverdeling van één of meer gestandaardiseerde normgroepen (ook wel referentiegroepen genoemd). Het doel is om de geteste persoon af te zetten tegen een groep anderen met wie een zinvolle vergelijking gemaakt kan worden. Ook de relatieve afstand ten opzichte van de normgroep kan hierbij aangegeven worden. Zo kan bijvoorbeeld aangeduid worden of een persoon ondergemiddeld of bovengemiddeld scoort. Bij criteriumgerichte interpretatie wordt de score vergeleken met een absolute norm, bijvoorbeeld grensscores. Deze grensscores worden afgeleid uit de onderzoeksgegevens en kunnen het scorebereik opdelen in twee of meer categorieën (APA, 1999). In dit artikel zal voor het gebruik in een klinische groep van beide type normen de voordelen en nadelen gegeven worden en handvatten geboden worden voor een betere interpretatie.

2 Normgerichte interpretatie

Het merendeel van de tests die gebruikt worden in de klinische setting maken gebruik van een normgerichte interpretatie. Dit is bijvoorbeeld het geval bij intelligentietests. De cliënt bij wie de test is afgenomen, wordt vergeleken met een normgroep. Deze normgroep moet een zinnige vergelijking bieden op basis van overeenkomsten in bijvoorbeeld leeftijd en geslacht. Met de uitslag op een dergelijke test kan geconstateerd worden in hoeverre de cliënt gemiddeld is ten opzichte van zijn/haar normgroep of een onder- of bovengemiddelde score heeft behaald. Hierbij is dan ook bekend hoeveel procent van de normgroep een dergelijke score of lager behaalde. Tevens is het mogelijk om aan te geven wat het betrouwbaarheidsinterval rond de betreffende score is, hetgeen aangeeft in hoeverre de gemeten score overeenkomt met de ware score. Bij een test is er namelijk altijd een bepaalde mate van een meetfout aanwezig; bij replicatie van de meting zijn er een aantal aan de specifieke, eenmalige testsessie gebonden factoren, waarvan de invloed over de testscore over replicaties heen niet geheel hetzelfde is (Drenth & Sijtsma, 2006). Hierbij moet men bijvoorbeeld denken aan een plotselinge hoestbui, tijdelijk concentratieverlies door invloed van de buitenwereld, slaperigheid en vele andere factoren.

2.1 Klinische personen in de normgroep

De referentie- of normgroep zoals die hiervoor is beschreven moet representatief zijn voor de bevolking waarin het instrument wordt gebruikt en de personen waarmee een vergelijking gemaakt wordt. Dit betekent dat indien er een vergelijking gemaakt wordt met de totale/gemiddelde bevolking dat er in de normgroep voor het overgrote deel mensen zitten met een normale ontwikkeling en maar een klein deel personen met een afwijkende ontwikkeling; ofwel 'klinische personen'. Net zoals dat in de totale bevolking het geval is. Het is de bedoeling dat deze steekproef een normaalverdeling (zie afbeelding 1) volgt, zodat het gemiddelde van de populatie en de standaardafwijking van het gemiddelde bepaald kunnen worden. Dit betekent dat er in een normgroep voor bijvoorbeeld een intelligentietest slechts 2.2% personen zitten met een IQ onder de 70. Het totale aantal personen uit de groep met een IQ onder de 70 die getrokken wordt, hangt af van de totale steekproef die getrokken wordt. Het aandeel van personen met een benedengemiddelde intelligentie zal echter altijd laag zijn. Op deze manier is het mogelijk om vast te stellen hoe groot de afwijking van 'normaal' (de gemiddelde populatie) is van een bepaalde geteste persoon.



Afbeelding 1 Normaal Verdeling

2.2 Klinische groep

Naast het vergelijken van een persoon uit een klinische groep met een normgroep uit de totale bevolking, zou het ook kunnen dat men wil vergelijken met andere personen uit dezelfde klinische groep. Als bijvoorbeeld een persoon met Autisme Spectrum Stoornis (ASS) een gestandaardiseerde score van 75 op een intelligentietest behaalt, is dat een ondergemiddelde score in vergelijking met de normgroep van de 'algemene' bevolking. Mogelijk is dit in vergelijking met andere personen met ASS echter een gemiddelde of bovengemiddelde score.

Dit zou ten eerste kunnen door het samenstellen van een normgroep op basis van een deel-populatie. Dit is bijvoorbeeld het geval bij een persoonlijkheidstest voor hoger opgeleiden, hier worden in de normgroep alleen hoger opgeleiden meegenomen. De behaalde standaard-score zegt in een dergelijk geval dus alleen wat ten opzichte van andere hoger opgeleiden. Ook een capaciteitentest voor psychologiestudenten zal genormeerd worden onder alleen psychologiestudenten. Deze normgroep moet dan wel een representatieve afspiegeling zijn van alle psychologiestudenten wat betreft persoonlijke- en omgevingskenmerken. Dit is vooral voor instrumenten die bedoeld zijn voor een kleine specifieke doelgroep interessant. Stel dat dit namelijk voor mensen met autisme gedaan zou worden, dan moet er voor elke leeftijdsgroep een aparte representatieve normgroep verzameld worden. Indien het instrument voor meerdere klinische populaties inzetbaar is, zouden voor elk van die klinische populaties representatieve normgroepen per leeftijd ontwikkeld moeten worden. Een andere bemoeilijkende factor zijn de inclusie- en exclusiecriteria in deze; welke personen worden wel en niet meegenomen om mee te vergelijken. In het voorbeeld van ASS; wordt een persoon met Asperger wel of niet in deze klinische groep opgenomen en hoe zit dat met iemand met PDD-NOS. Men kan zich voorstellen dat dit het gemiddelde van de scores kan beïnvloeden en de bruikbaarheid van de vergelijkingsgroep verkleinen.

Een andere methode die min of meer hetzelfde resultaat oplevert is om naast een normgroep op basis van de totale bevolking waarmee standaardscores berekend kunnen worden, een klinische groep te testen. Van de individuen in deze groep kunnen op basis van de normen uit de normaalbevolking standaardscores berekend worden en de gemiddelde standaard-score en standaarddeviatie hiervan bepaald worden. Vervolgens kan de standaard-score van een individu vergeleken worden met de gemiddelde standaard-score van die klinische groep. Dit is minder intensief, omdat de normgroep dan de totale populatie is en de gemiddelden en standaardafwijkingen van specifieke klinische groepen op kleinere aantallen gebaseerd kunnen worden. Dit omdat de omzetting van ruwe score naar standaard-score op basis van bijvoorbeeld leeftijd dus al heeft plaatsgevonden.

De keuze of een cliënt alleen vergeleken wordt met de totale bevolking of daarnaast ook met een klinische groep hangt af van de vraagstelling van de psycholoog. Om bij het voorbeeld van een persoon met ASS te blijven; stel dat het gaat om het bepalen van de intelligentie om een schooladvies voor deze leerling te geven. Dan wil men weten hoe deze leerling functioneert ten opzichte van leeftijdsgenoten en dus is vergelijking met de totale bevolking geëigend. Stel dat de vraagstelling is of deze persoon geschikt is voor een bepaalde behandeling, dan wil men weten of de score ten opzichte van andere personen met ASS gemiddeld is of onder- of bovengemiddeld en of het advies voor plaatsing positief of negatief is. In deze situatie geeft vergelijking met een klinische groep juist meer informatie.

2.3 Uitbreiding van de norm

Bij het testen van klinische groepen bestaat soms de wens om de bestaande normen uit te breiden om aan de onderzijde of bovenzijde van de normaalverdeling meer te kunnen differentiëren. Zo hebben de meeste IQ-tests een bereik van 45-155, maar is er bijvoorbeeld de wens om ook onder de 45 te kunnen differentiëren. Dit om ook tussen de laagste scoorders nog een vergelijking te kunnen maken. Er kan dan een groep verzameld worden met een score van 2 standaardafwijkingen onder gemiddeld of lager om te kijken of de test binnen die groep ook nog kan differentiëren. Er moeten dan voldoende items in de test zitten met een lage moeilijkheidsgraad die door een deel van de groep goed en door een deel van de groep fout beantwoord worden. Hierbij wordt wel de eerder berekende norm gebruikt en alleen gecontroleerd of de getrokken lijn (die de samenhang tussen ruwe score en standaardscore beschrijft) ook op dezelfde wijze doorloopt bij lagere of hogere scores.

3 Criteriumgerichte interpretatie

Hoewel de meeste tests in de klinische setting (zoals eerder aangegeven) een normgerichte interpretatie hebben, zijn er tevens tests met een criteriumgerichte interpretatie. Dit zijn tests die een indeling maken in twee of meer categorieën met het doel om een onderscheid te maken tussen twee of meer groepen mensen. Deze type tests werken dan ook met een grensscore op basis waarvan een indeling gemaakt kan worden. De grensscore houdt geen rekening met een bepaalde mate van meetfout. Deze criteriumgerichte interpretatie kan bijvoorbeeld gebruikt worden bij een test voor de diagnosticering van depressie; de uitkomst van de test bepaald of iemand voldoet aan de diagnose of niet. Buiten de klinische setting kan hierbij ook gedacht worden aan een minimumscore om te kunnen voldoen aan de eis die gesteld is voor een test. Op basis van de score wordt een persoon dan wel of niet toegelaten tot een bepaalde opleiding. Er wordt bij de criteriumgerichte interpretatie verder niet gekeken naar de relatieve afstand tussen scores van individuen, het is daarom niet mogelijk om te zeggen dat iemand 2 keer zo goed is in Rekenen of net zo goed als 25 procent van de doelgroep. Men gebruikt de score alleen om een indeling in categorieën te maken.

3.1 Klinische norm

Anders dan bij de normgerichte interpretatie, bestaat de normgroep bij een criteriumgerichte interpretatie wel uit een klinische groep. Dit is altijd een specifieke klinische groep, zoals patiënten met een klinische diagnose. Deze steekproef moet, net als bij de normgerichte interpretatie, een representatieve groep zijn. De steekproef 'depressieven' moet bijvoorbeeld een goede afspiegeling zijn van alle 'depressieven', wat betreft leeftijd, duur van de depressie, ernst etc. Deze groep moet ook afkomstig zijn uit alle verschillende regio's, verschillende opleidingsniveaus hebben en bestaan uit alle leeftijdsgroepen waarvoor het instrument ingezet gaat worden. Daarnaast wordt bij het bepalen van de grenswaarden gebruik gemaakt van een 'gezonde groep' om te bepalen welke waarde op de test het verschil maakt tussen een 'gezond' en 'klinisch' persoon. Een goede test zorgt ervoor dat alle gezonde mensen op basis van hun score op de test ook in de categorie gezond worden ingedeeld en alle 'klinische' personen ook in de categorie 'klinisch'. Dit zal echter nooit helemaal het geval zijn, omdat er bij een stoornis vaak sprake is van een geleidelijk verloop. Bijvoorbeeld bij dyscalculie is er een groep personen die veel moeite heeft met rekenen (op basis van de theoretische normaalverdeling haalt 15.9% van alle personen een score van 1 standaardafwijking of meer ondergemiddeld), maar wie van deze personen voldoet aan de diagnose dyscalculie? Het is nooit precies te bepalen waar de grens ligt dat er sprake is van een stoornis. Daarom moet er een grens worden gekozen waardoor er zo min mogelijk fout-positieven (onterecht door de test als 'klinisch' bestempeld) en fout-negatieven (onterecht door de test als 'gezond' bestempeld) zijn. En zoveel mogelijk terecht-positieven (sensitiviteit = het aantal echt positieven/het aantal echt positieven + het aantal fout negatieven) en terecht-negatieven (specificiteit = het aantal echt negatieven/het aantal echt negatieven + aantal fout positieven); personen die door de test juist ingedeeld zijn (Drenth & Sijsma, 2006).

4 Valideringsonderzoek

Een ander onderzoek wat erg belangrijk is om aan te tonen dat mensen uit een klinische setting op betrouwbare wijze getest kunnen worden met het instrument is een valideringsonderzoek. Een dergelijk onderzoek wordt opgezet om te onderzoeken of de test meet wat hij pretendeert te meten. Hierbij wordt onderzocht of de test bruikbaar is in een bepaalde klinische groep; is het niveau passend, zijn er geen beperkingen waardoor personen de opgaven niet kunnen maken etcetera. Zo zijn bijvoorbeeld visuele tests zijn niet af te nemen bij iemand die slechtziend is. Dit moet onderzocht worden om te voorkomen dat geconcludeerd wordt dat iemand ondergemiddelde performante capaciteiten heeft, terwijl blijkt dat diegene kleurenblind is en daardoor in de opgaven het verschil tussen groene en blauwe bollen niet kon zien.

Dit onderzoek is vooral op theoretische gronden gestoeld, waarbij de test wordt onderzocht op voorwaarden die nodig zijn om de test te kunnen maken. Hierbij valt bijvoorbeeld te denken aan voldoende gezichtsvermogen, voldoende hoorvermogen, voldoende motorische vaardigheden of voldoende taalvaardigheden. Ook worden de items om ze te onderzoeken op itembias (het bevoordelen of benadelen van bepaalde groepen) vaak voorgelegd aan experts uit de praktijk. Het afnemen van het instrument bij de klinische groep in de praktijk levert vaak al veel informatie op over het wel of niet afneembaar zijn van het instrument bij een klinische doelgroep.

Daarnaast wordt valideringsonderzoek gebruikt om aan te tonen dat een instrument bij afname bij mensen waar een afwijkende score verwacht wordt (klinische groep), deze afwijking ook laat zien. Dit gegeven onderbouwt de validiteit van het instrument. De gevonden afwijkende scores moeten in overeenstemming zijn met hetgeen verwacht wordt, op basis van de literatuur.

Tot slot is het voor de klinische praktijk ook belangrijk om te weten in welke mate instrumenten binnen eenzelfde meetdomein met elkaar overeenkomen. Als men bijvoorbeeld een algemene intelligentietest vergelijkt met een non-verbale intelligentietest, zullen de scores van de normale populatie elkaar weinig ontlopen. Maar als men beide instrumenten in de klinische setting afneemt, zal mogelijk zichtbaar worden dat mensen met verbale problemen op de non-verbale test veel hoger scoren dan op de algemene intelligentietest. Er worden met beide tests dus niet op dezelfde wijze de algemene cognitieve vaardigheden in kaart gebracht. Iets waar vooral in de klinische praktijk rekening mee gehouden moet worden. Om te onderzoeken in hoeverre instrumenten binnen eenzelfde meetdomein hetzelfde meten kan een neventestonderzoek worden gedaan. Daarmee is te zien welke overlap er is en in hoeverre de gemiddelde scores van de normgroep verschillen tussen de meetinstrumenten. In hoeverre dit echter de scores van de klinische groep beïnvloedt, wordt bijna nooit onderzocht.

5 Conclusie

In dit artikel is getracht uiteen te zetten op welke verschillende manieren omgegaan kan worden met normen en betrouwbaarheidsgegevens in klinische groepen. Wij hebben hiermee een kader geschetst voor het gebruik van psychologische meetinstrumenten bij klinische groepen. Bruikbaarheid hangt vooral af van het doel waarmee de test wordt ingezet.

Indien het doel is om de score af te zetten tegen een norm- of referentiegroep, is een instrument vereist met een normgerichte interpretatie. Deze normgroep kan bestaan uit een representatieve steekproef uit de totale bevolking of een steekproef uit de klinische populatie. Of men vergelijkt met de totale bevolking of met de klinische populatie hangt af van het doel van de testafname. Op basis van beide normgroepen kan aangegeven worden in hoeverre het individu afwijkt van de 'norm' en of dit gemiddeld is of onder- of bovengemiddeld.

Indien het doel van afname is om de score te vergelijken met een absolute norm, is een instrument met een criteriumgerichte interpretatie vereist. Hierbij kan voor een individu enkel aangegeven worden tot welke groep hij/zij behoort en kan dat niet in termen van onder- of bovengemiddeld beschreven worden.

Bij het inzetten van tests bij een cliënt uit een klinische groep zijn er vele factoren om rekening mee te houden. Een startpunt kan in elk geval zijn om na te gaan of de test die u wilt afnemen aansluit bij de vraagstelling en de gewenste normgroep of vergelijkingsgroepen bevat.

Over de auteurs

J.E. Dek MSc en M.S.L. Braak MSc zijn als Product Developers werkzaam bij Pearson Assessment and Information BV op de afdeling Research and Development. Drs. A.P. Kooij is uitgever bij Pearson Assessment and Information BV, Radarweg 60-A1, 1043 NT Amsterdam. Voor correspondentie over dit artikel: joelle.dek@pearson.com.

Referenties

- American Educational Association. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Drenth, P.J.D., Sijtsma, K. (2006). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen*. Houten: Bohn Stafleu van Loghum.

Summary

Interpreting the results of clients from clinical groups.

For every psychological test the usability on clinical groups is of interest for diagnostic purposes. This article suggests it's important to make a distinction between norm referenced and criterium referenced norms.

If you want to compare your scores to a reference group; intra-individual comparison, you need norm referenced. In a norm referenced test the normgroup could be composed of a representative random group of the total population or a normgroup based on a representative clinical group. When describing the scaled score of an individual it gives information on how the score is related to that specific population.

If you want to compare with an absolute norm; only individual comparison, you need criterium referenced. The score only classifies the individual in a category, no information about the whether the result is average or below or above average is received.

When choosing a test for use in clinical setting these considerations should be kept in mind.



Pearson Assessment and Information BV

Radarweg 60-A1, 1043 NT Amsterdam

Postbus 78, 1000 AB Amsterdam

T: +31 (0)20 581 5500

E: info-nl@pearson.com

www.pearsonclinical.nl

www.pearsonclinical.be

Twitter: [@PearsonNL](https://twitter.com/PearsonNL)