

De wetenschap achter het voorspellen van prestaties op het werk bij werving

Auteurs:

Wyn Davies, wereldwijd productmanager, Pearson
TalentLens, Angus McDonald, Beëdigd psycholoog

Datum: mei 2018



MORE INSIGHT
MORE IMPACT™

Inleiding

Het werven van kandidaten is een risicovol vak waarbij de kansen om een kandidaat te identificeren die een effectieve medewerker zal blijken te zijn, worden afgewogen tegen de risico's van iemand te identificeren die niet zal presteren zoals verwacht. Recruiters gebruiken een breed scala aan tools en processen om de risico's van beslissingen bij het werven te beheren. Wanneer risico's effectief worden beheerd, is de kans groter dat recruiters kandidaten identificeren die goed presteren op het werk, maar zelfs onder optimale omstandigheden is dit proces verre van perfect.

Door de complexiteit van het begrijpen van kandidaten en het evalueren of ze echt over de capaciteiten, kenmerken en motivaties beschikken die nodig zijn voor een bepaalde functie, is het gemakkelijk om verkeerde beslissingen te nemen. Factoren zoals opzettelijke misleiding van de kandidaat enerzijds en de moeilijkheden voor recruiters anderzijds nemen alleen maar toe. Als recruiters het mis hebben lopen ze risico op kosten waaronder de noodzaak om medewerkers snel te vervangen en de potentiële verstoring die medewerkers die slecht bij de organisatie passen kan veroorzaken.

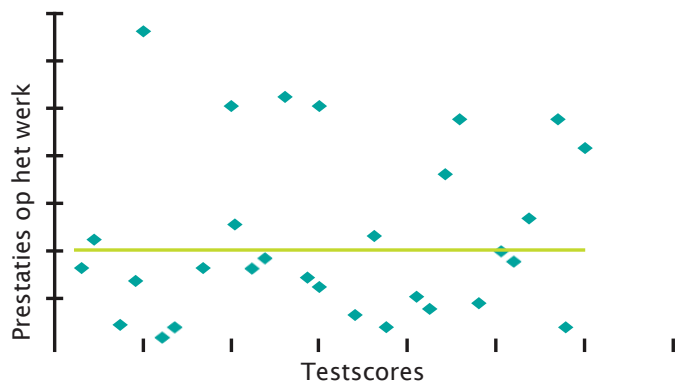
In antwoord op deze wervingsuitdaging is er nu meer dan 100 jaar wetenschappelijk bewijs voor de effectiviteit van verschillende tools die gebruikt kunnen worden in het wervingsproces. Dit bewijs is gebaseerd op het verband tussen de evaluaties van kandidaten tijdens het wervingsproces en hoe ze vervolgens presteren in de functie waarvoor ze zijn aangenomen. Op zijn beurt moet dit bewijs in aanmerking worden genomen bij het inwerkproces en op die manier recruiters helpen te bepalen welke tools het meest effectief zijn om het risico te beheren.

In dit document wordt het wetenschappelijk onderzoek naar de effectiviteit van verschillende wervingstools samengevat, de belangrijkste concepten uitgelegd en wordt aangegeven welke overwegingen in aanmerking moeten worden genomen bij het gebruik van wervingstools. Erkend wordt dat de processen en wervingstools waarbinnen recruiters werken verschillen per organisatie en regio. Bijgevolg moet de hier vermelde informatie worden toegepast op een wijze die in overeenstemming is met deze lokale kaders.

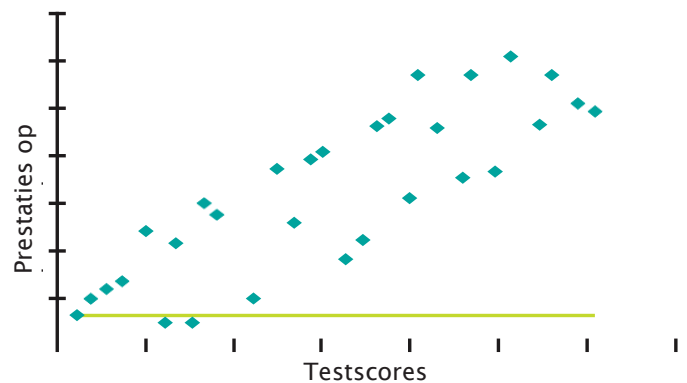


Correlatie

Er kunnen veel verschillende statistische technieken worden gebruikt om te bepalen hoe effectief een beoordelingstool is voor kandidaatselectie, maar het merendeel hiervan is gebaseerd op het idee van correlatie. Correlatie is een statistische techniek die kijkt hoe sterk de samenhang is tussen twee variabelen, meestal de score of scores verkregen op een beoordelingstool en een meting van prestaties op het werk. Hoe sterker de correlatie, hoe sterker de onderlinge samenhang is tussen de twee metingen, zoals geïllustreerd in de twee onderstaande diagrammen.



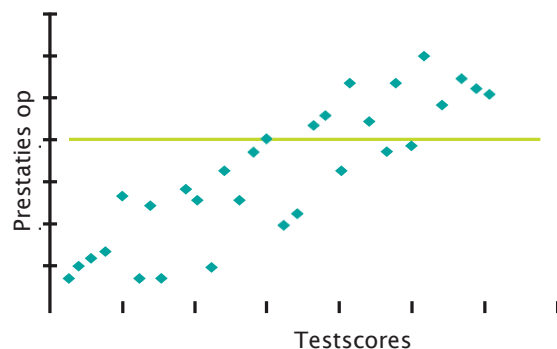
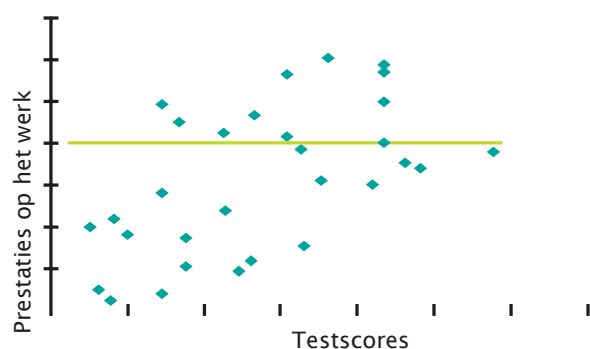
De diagrammen geven de verdeling aan van de scores in een groep van 30 mensen die een beoordeling hebben ingevuld. De grafiek links geeft aan dat er een minder sterke samenhang is tussen de scores uit de beoordelingstool en die van de prestaties op het werk, terwijl de grafiek rechts een veel sterkere samenhang aangeeft. Er is sprake van een sterke samenhang wanneer het grootste deel van datapunten op of dichtbij een 'lijn met de beste match' valt. Het belang van deze samenhang is dat we met meer nauwkeurigheid kandidaten kunnen identificeren die waarschijnlijk goed zullen presteren in de functie.



Het concept van validiteit

In de context van werving betekent validiteit in hoeverre een specifieke beoordelingstool geschikt is voor een omschreven doel. Door tools met een hogere validiteit te gebruiken, kunnen recruiters met meer nauwkeurigheid de kandidaten identificeren die waarschijnlijk goed zullen presteren in de functie. Hierdoor helpen ze aanzienlijk bij het risicobeheer, wat een onvermijdelijk deel uitmaakt van elke wervingsbeslissing.

Een sterke validiteit betekent dat een beoordelingstool de prestaties op het werk met een hoge mate van nauwkeurigheid kan voorspellen. Als we dit weten is het mogelijk om beoordelingscores te gebruiken om die kandidaten te identificeren die hoogstwaarschijnlijk goed zullen presteren in de functie. Bekijk de onderstaande diagrammen waarin de scores van twee verschillende beoordelingen worden vergeleken met de prestaties op het werk van 30 kandidaten. De horizontale groene lijn geeft het ideale niveau van prestaties op het werk weer.



De belangrijkste boodschap luidt als volgt: met een sterke validiteit is het makkelijker het risico in verband met wervingsbeslissingen te beheren omdat de beoordelingsscores de prestatie nauwkeuriger voorspellen. Het is belangrijk om op te merken dat prestaties worden beïnvloed door een aantal factoren en dat er (voor zover wij weten) geen absolute meting of test bestaat die succes in de functie met 100% zekerheid voorspelt. Je kan echter een scala aan relevante en valide beoordelingen gebruiken zodat je de kans op het voorspellen van waarschijnlijke prestatie aanzienlijk vergroot en het risico dat je de verkeerde persoon aanneemt kleiner wordt.

Wat vertelt de wetenschap ons?

In de praktijk weten recruiters niet hoe medewerkers zullen presteren in de functie. Dit zorgt ervoor dat de keuze van valide beoordelingstools en de juiste toepassing van de resultaten ervan op het moment van de werving van cruciaal belang zijn.

We kennen nu de principes achter validiteit en waarom het belangrijk is, maar welke beoordelingstools zijn het meest valide om prestaties in de functie zelf te voorspellen? Er is uitgebreid onderzoek gedaan naar het gebruik van verschillende beoordelingstools en -technieken. Van dit onderzoek is recent een samenvatting gemaakt om de effectiviteit van verschillende methodes te kunnen vergelijken. Een van de belangrijkste manieren waarop dit gebeurde is via de zogenaamde meta-analysetechniek, een statistische techniek die de bevindingen bundelt. De resultaten worden gepresenteerd als correlatiecoëfficiënten. Een score van 1

geeft aan dat de beoordelingsmethode de prestaties perfect voorspelt. Hoe hoger de correlatiecoëfficiënt, hoe beter de beoordelingsmethode de prestatie voorspelt. Een recente publicatie van Frank Schmidt en collega's vat de effectiviteit samen van 31 verschillende beoordelingsmethodes om de prestaties op het werk te voorspellen en 16 methodes om de prestaties bij werkgerelateerde opleiding te voorspellen.

De belangrijkste bevinding uit dit onderzoek was dat algemene mentale vaardigheidentesten (general mental ability - GMA) de hoogste validiteit van alle selectiemethodes bleken te hebben. Ze werden ook gezien als de beste voorspeller van prestaties bij werkgerelateerde opleiding. Het uitgebreide onderzoek dat ten grondslag ligt aan het gebruik van GMA betekent dat deze bevindingen in hoge mate betrouwbaar zijn.

Wat is meta-analyse?

Organisaties en academische onderzoekers willen al jarenlang begrijpen wat het verband is tussen beoordelingen als onderdeel van het wervingsproces en de latere aspecten van prestaties op het werk. Er zijn al duizenden onderzoeken uitgevoerd die deze link hebben onderzocht, maar zonder consistente resultaten. Dit onderzoek dat is gebaseerd op echte wervingsprocessen kent vele uitdagingen zoals het kunnen volgen van mensen vanaf het moment van aanwerving tot aan het moment waarop prestaties op het werk zinvol gemeten kunnen worden, het verkrijgen van nauwkeurige metingen van prestaties op het werk en kleine referentiegroepen.

Voor het laatste punt is een probleem wanneer we de validiteit van wervingsprocessen beoordelen. Voor robuuste onderzoeksresultaten is het belangrijk dat er voldoende mensen bij betrokken zijn. Veel organisaties hebben maar een beperkt aantal mensen in elke functie zodat het tijd kost om voldoende referentiemateriaal samen te stellen, als dit al mogelijk is. Meta-analyse is een statistische techniek waarmee bevindingen uit veel verschillende afzonderlijke onderzoeken kunnen worden gebundeld. Daarbij wordt rekening gehouden met meetfouten en kleine referentiegroepen wat betekent dat er robuustere schattingen van verbanden tussen de verschillende beoordelingsmethodes en de prestaties op het werk vastgesteld kunnen worden.

Meta-analyses hebben consequent aangegeven dat sommige soorten beoordelingen veel predictiever voor prestaties op het werk zijn dan andere. Bij gebrek aan een goede lokale validatie voor een specifieke functie, verschaffen deze bevindingen de recruiters een solide basis om beoordelingen te vinden die waarschijnlijk het meest effectief zijn.

Algemene mentale vaardigheidentesten (GMA)

Zoals Schmidt en collega's opmerken, selecteert een werkgever die GMA gebruikt om medewerkers te selecteren die sterk zullen presteren in de functie ook diegenen die het meest zullen opsteken van werkopleidingsprogramma's en die sneller vakkennis zullen opdoen door ervaring op het werk. **Dit laatste punt heeft betrekking op de hoofdreden waarom kandidaten die hoger scoren op GMA, beter presteren in de functie. Ze verwerven sneller en grondiger vakkennis dan degenen die laag scoren op GMA.**

Er zijn veel tests ontworpen om de aspecten van mentaal vermogen te beoordelen en er zijn tal van voorbeelden waar deze in de eerste plaats zijn ontworpen voor gebruik in wervingscontexten. Voorbeelden hiervan zijn de Watson Glaser Critical Thinking Appraisal en Raven's Progressive Matrices. De versies van deze tests kunnen online en zonder toezicht worden afgenomen of aan een groep kandidaten worden voorgelegd die de test dan allemaal tegelijk invullen zodat ze in hoge mate

bruikbaar zijn voor wervingscontexten. Ze zijn relatief makkelijk te gebruiken en te interpreteren, het duurt niet lang om ze af te leggen en ze zijn vrij kosteneffectief.

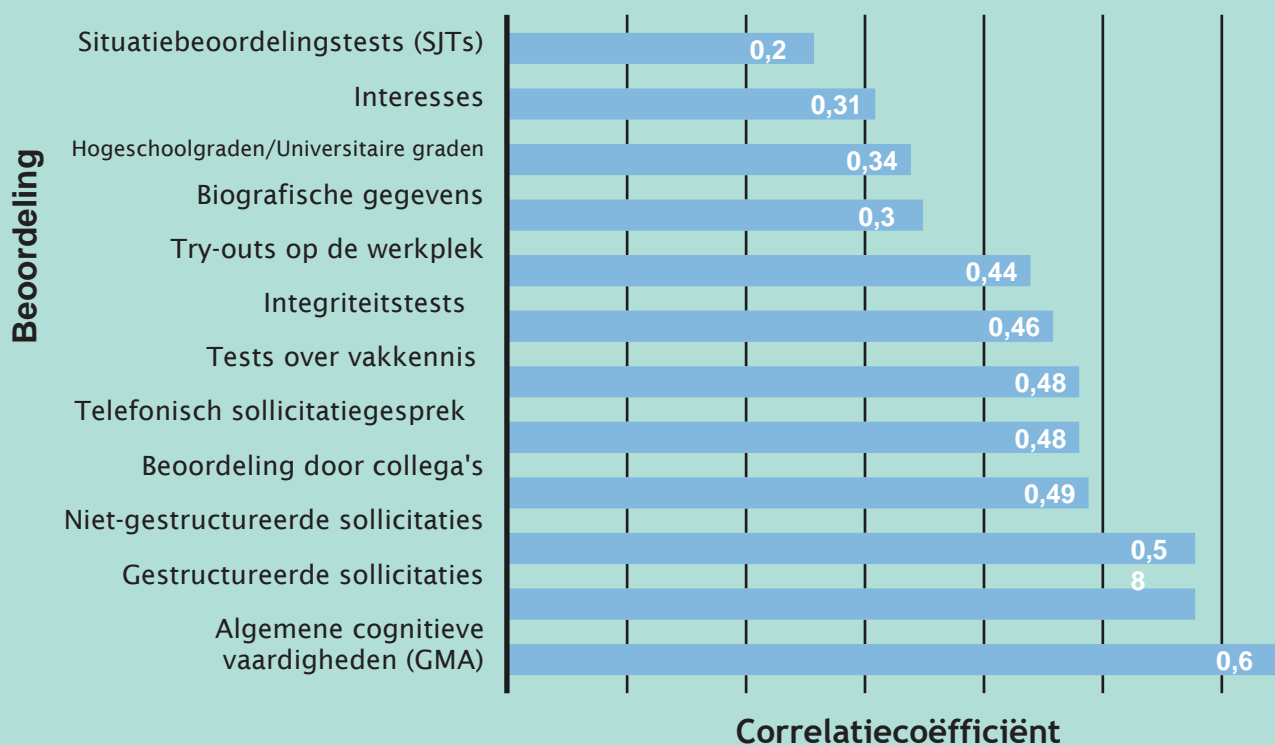
Een ander vermeldenswaardige bevinding met betrekking tot GMA is dat de tests valide zijn voor een breed scala aan functies. Deze tests worden meestal gebruikt voor werving van afgestudeerden waarbij men ervan uitgaat dat de functies aanzienlijke cognitieve eisen stellen aan de kandidaten. GMA-tests staan erom bekend dat ze effectief de prestaties voorspellen voor functies op het niveau van net afgestudeerden, maar ze hebben ook een aanzienlijke predictieve validiteit voor nagenoeg elke functie, ook voor ongeschoolde functies. Dit vermogen van GMA om prestaties te voorspellen neemt meestal toe naarmate de cognitieve vereisten van de functie toenemen. De hoogste validiteit zien we bij professionele en leidinggevende functies waar GMA gemiddeld meer dan 50 procent van de variabiliteit in prestaties op het werk kan verklaren.

Dieper inzicht in validiteit

Behalve GMA zijn er nog vele wervingstools en de meeste selectieprocessen gebruiken een combinatie hiervan.

Dus hoe verhouden andere metingen zich tot de effectiviteit van GMA?

Correlatiecoëfficiënten die prestaties op het werk voorspellen



Het zou uitzonderlijk zijn om een functie aangeboden te krijgen zonder enige vorm van sollicitatiegesprek. Onderzoek ondersteunt de voordelen van sollicitatiegesprekken omdat ze na GMA een van de meest effectieve selectiemethodes zijn. Interessant is dat er weinig bewijs is om te suggereren dat gestructureerde sollicitatiegesprekken effectiever zijn dan niet-gestructureerde sollicitatiegesprekken hoewel gestructureerde telefonische sollicitatiegesprekken meestal iets minder valide zijn.

Andere methodes waarvan in het onderzoek van Schmidt en collega's blijkt dat ze op z'n minst bescheiden predictieve validiteit hebben, zijn onder meer integriteitstests, biografische gegevens, assessment centers, gewogen gemiddelden (d.w.z. academische resultaten op masterniveau), beoordelingen door collega's, testen met echte werkvoorbeelden, try-outs op de werkplek, gedragsconsistentiemethode en tests over de vakkennis. Niet alle deze methodes

kunnen echter direct worden toegepast op alle wervingscontexten. Try-outs op de werkplek bijvoorbeeld, kunnen niet makkelijk worden gebruikt in hoogrisicofuncties of functies waar een aanzienlijke opleiding voor nodig is en beoordelingen door collega's zijn ook alleen maar bruikbaar voor interne kandidaten.

Zelfs wanneer andere beoordelingsmethodes gebruikt kunnen worden betekent dit niet noodzakelijk dat men hierdoor in staat zal zijn om het risico van wervingsbeslissingen beter te beheren. Om dit te bepalen moeten we begrijpen hoe aanvullende metingen zorgen voor incrementele validiteit. Hoewel metingen op zich hun nut kunnen hebben, kan hun impact veel minder zijn nadat er rekening is gehouden met de effecten van GMA. Schmidt en collega's beschouwen GMA zodanig fundamenteel voor de voorspelling van prestaties op het werk dat ze de effectiviteit van alle andere methodes pas evalueren na rekening te hebben gehouden met GMA.

Incrementele validiteit

GMA-tests zijn de allerbeste voorspellers van prestaties op het werk, maar ook vele andere methodes kunnen effectief zijn. Op basis hiervan zou het verleidelijk kunnen zijn om een aantal beoordelingsmethodes van hoge validiteit te gebruiken om het wervingsrisico te verminderen. Helaas is de situatie niet zo eenvoudig. De uitdaging bestaat erin om te begrijpen hoe elke methode op zich bijdraagt aan ons inzicht in de prestaties op het werk. Dit wordt incrementele validiteit genoemd.

Laten we ervan uitgaan dat de grote blauwe cirkels hieronder de prestaties op het werk laten zien. Beoordelingen worden gebruikt om inzicht te krijgen in waarschijnlijke prestaties op het werk en we weten dat GMA goed is voor ongeveer 50 procent hiervan onder optimale omstandigheden. Dit wordt geïllustreerd door de bruine 'GMA'-cirkel die de prestaties op het werk overlapt. Laten we nu eens kijken wat er gebeurt als we een extra meting uitvoeren zoals aangeduid met de groene cirkel.



Deze groene meting heeft een vergelijkbaar verband met prestaties op het werk zoals GMA maar in het geval links voegt het weinig toe aan het algemene inzicht in de prestaties op het werk.

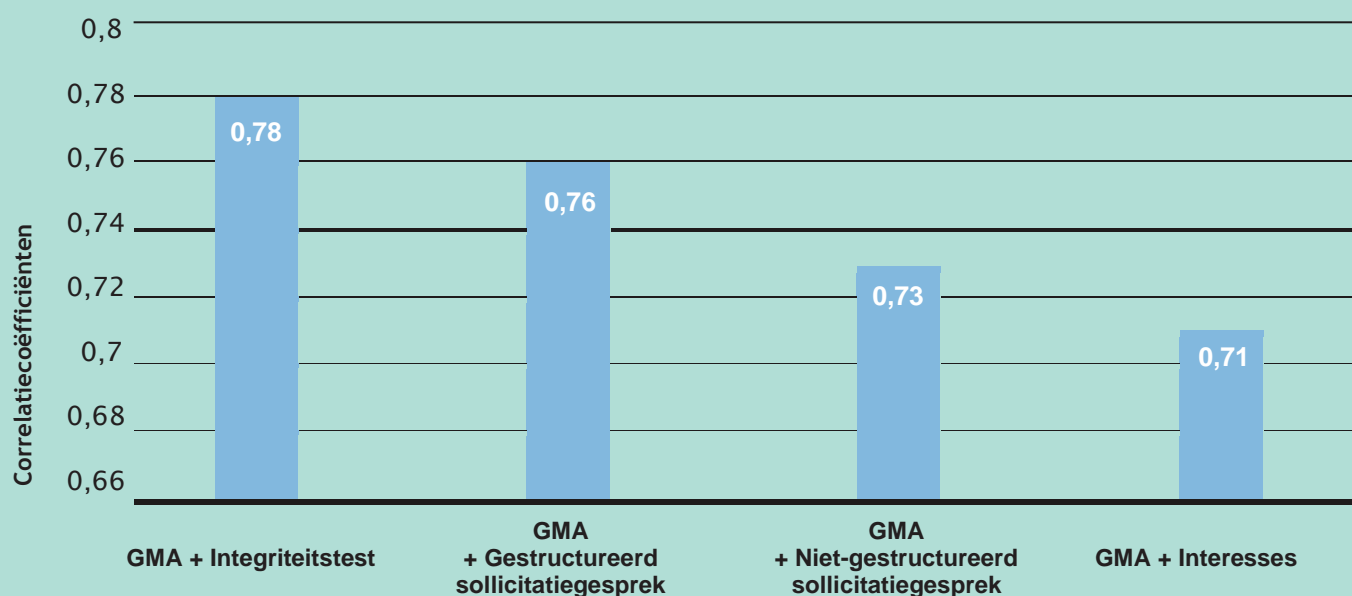
Dit komt omdat de overlap, of correlatie, tussen GMA en de extra meting aanzienlijk is. In het geval rechts is de overlap tussen de extra meting en GMA veel minder. In dit geval geeft de extra meting naast GMA ons wel aanzienlijk meer algemeen inzicht in de prestaties op het werk.

De mate waarin het gebruik van een extra meting aanvullende informatie verschaft over de prestaties op het werk, wordt 'incrementele validiteit' genoemd. Vanwege het grote effect van GMA, is het zinvol om na te gaan wat de incrementele is het zinvol

om na te gaan wat de incrementele validiteit is van andere metingen zodra rekening is gehouden met GMA.

Zodra rekening is gehouden met het predictieve vermogen van GMA, verandert het resultatenpatroon enigszins. De hoogste gecombineerde predictieve validiteit zien we bij het gebruik van GMA en een integriteitstest. Dit wordt op de voet gevolgd door een combinatie van GMA en een gestructureerd sollicitatiegesprek en vervolgens door de combinatie van GMA en een niet-gestructureerd sollicitatiegesprek. Het gebruik van integriteitstests voegt ongeveer 20% toe aan het predictieve vermogen van GMA alleen terwijl gestructureerde en niet-gestructureerde sollicitatiegesprekken respectievelijk 18% en 13% toevoegen. De enige andere meting die 10% of meer toevoegt aan het predictieve vermogen van GMA zijn beoordelingen van carriëredoelen.

Voorspelbaarheid van gecombineerde beoordelingsmethodes



Integriteitstests

Aangezien integriteitstests de meeste incrementele validiteit toevoegen aan GMA, is het de moeite waard deze nader te bekijken. We kunnen twee soorten integriteitstests onderscheiden: de tests met openlijke vraagstelling over oneerlijkheid ('openlijke' of 'directe' beoordelingen), en de tests die meer gebaseerd zijn op persoonlijkheid ('indirecte' beoordelingen). In directe integriteitstests kunnen vragen worden gesteld over de houding ten opzichte van oneerlijkheid, overtuigingen over de prevalentie van oneerlijkheid en in sommige gevallen wordt deelnemers gevraagd naar oneerlijk gedrag in het verleden. Indirecte beoordelingen daarentegen proberen aspecten van iemands karakter te identificeren die aan de basis kunnen liggen van oneerlijk

gedrag.

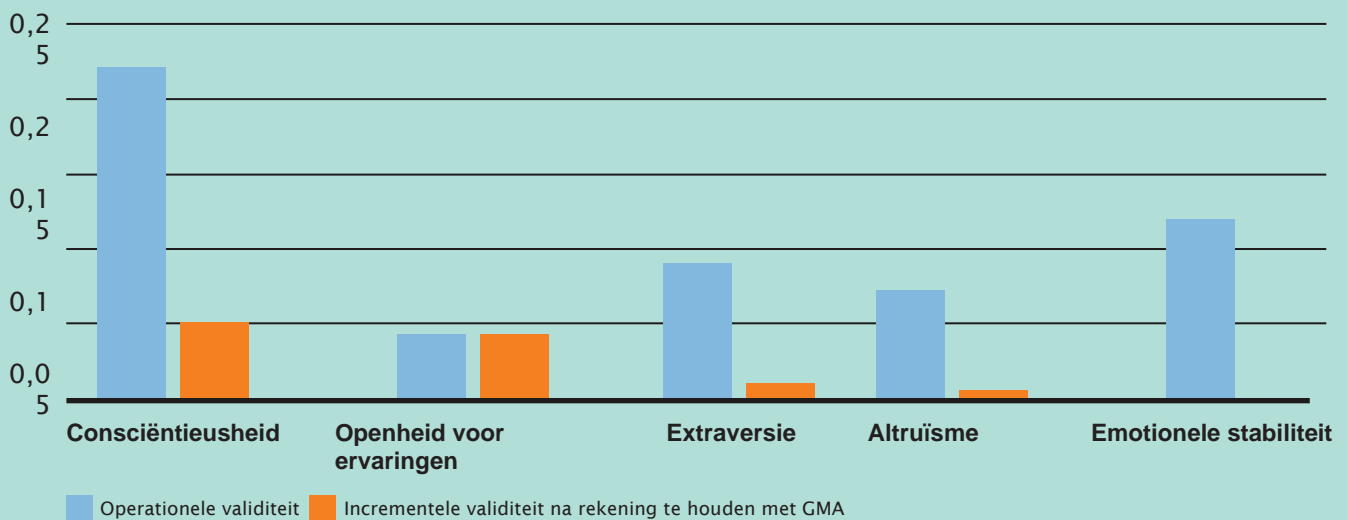
Hoewel beoordelingen op basis van persoonlijkheid verder kunnen afstaan van het feitelijke gedrag, hebben ze het voordeel minder transparant te zijn en bijgevolg minder gevoelig voor opzettelijke vervalsing van de antwoorden. Beide soorten worden geassocieerd met gedrag op de werkplek maar ondanks hun schijnbare gevoeligheid voor vervalsing blijken directe tests iets meer validiteit te hebben dan indirecte metingen².

Persoonlijheidsbeoordelingen

Omdat persoonlijkheidsbeoordelingen veel worden gebruikt bij wervingen, en al werden genoemd in verband met integriteit, is het de moeite waard om stil te staan bij de validiteit van dergelijke instrumenten in verband met prestaties op het werk. Hoewel de Big Five-factoren van Consciëntieusheid en Emotionele stabiliteit redelijk predictief zijn, zien we in de onderstaande grafiek dat

de effecten van de meeste persoonlijkheidsfactoren nagenoeg nul zijn zodra rekening is gehouden met GMA. De uitzonderingen hierop zijn Consciëntieusheid, dat zoals eerder gezegd een belangrijk onderdeel vormt van veel integriteitsbeoordelingen, en in iets mindere mate Openheid voor ervaringen.

Validiteit van de Big Five-persoonlijkheidsfactoren



De rol van functieanalyse

Onderzoek naar predictieve validiteit is een goede richtlijn om te weten hoe nuttig verschillende soorten van metingen kunnen zijn in het wervingsproces. De exacte beslissing om een bepaalde combinatie van beoordelingen te gebruiken moet echter gebaseerd zijn op een duidelijk inzicht in de functievereisten. Zonder dit lopen recruiters het risico om blindelings beoordelingen te doen op basis van gegevens zonder breder te overwegen hoe deze verband houden met de specifieke functievereisten.

De functieanalyse geeft details over de vereisten en het takenpakket van een specifieke functie en vormt de basis voor de functiebeschrijving en het profiel van de kandidaat. Een duidelijke beschrijving van het profiel van de kandidaat identificeert welke aspecten bij de kandidaten beoordeeld moeten worden en geeft ook structuur aan het beoordelingsproces. Een gestructureerd wervingsproces bestaat uit duidelijk gedefinieerde fases en elke fase zorgt voor meer informatie over de kandidaten. Op basis van deze informatie kunnen kandidaten worden afgewezen of doorgestuurd naar de volgende fase. Het is dus aan de recruiters om de juiste

beoordelingsmethodes te vinden en te bepalen waar deze het best passen in het wervingsproces. Bijvoorbeeld, wanneer een functie bepaalde kwalificaties vereist of indien de sollicitant een licentie nodig heeft om het beroep uit te oefenen of iets dergelijks, moet dit in een zeer vroeg stadium gecontroleerd worden. Het heeft geen zin andere beoordelingen te doen om later in het wervingsproces vast te stellen dat een kandidaat niet over de vereiste kwalificaties beschikt.

Kwalificaties zijn een voorbeeld van de essentiële vereisten met uitzondering van de situaties waarin organisaties bereid zijn kandidaten te ondersteunen om hier naartoe te werken. Vuistregel is om eerst deze essentiële vereisten te beoordelen voordat de vereisten die slechts wenselijk zijn worden beoordeeld. Dergelijke essentiële vereisten kunnen bestaan uit kwalificaties of geschiktheid voor het werk maar kunnen ook betrekking hebben op beoordelingen in de vorm van GMA waarin de minimumvereisten voor een bepaalde functie zijn opgesteld. Aangezien vele GMA-tests online beschikbaar zijn voor afname

zonder toezicht, kunnen ze zeer vroeg in het selectieproces worden gebruikt.

Dankzij de technologie is het mogelijk om steeds meer beoordelingen van kandidaten op afstand te doen, met inbegrip van GMA-tests, persoonlijkheidsbeoordelingen, testen met echte werkvoorbeelden en zelfs sollicitatiegesprekken. Enkele van de meest valide beoordelingsmethodes kunnen daardoor worden gebruikt zonder direct contact met de kandidaten. Hoewel dit efficiënt kan zijn voor het verwerken van sollicitaties, zullen organisaties die het aantrekken van kandidaten belangrijk vinden, wellicht niet kiezen voor deze aanpak. Ze hebben liever vroeg persoonlijk contact met kandidaten als een manier om ze bij de organisatie te betrekken. Toch worden GMA-tests doorgaans goed ontvangen door kandidaten³,

ondanks het feit dat ontwikkelaars van meer op een spel lijkende beoordelingen het tegenovergestelde beweren.

Kortom, er bestaat niet één beste manier om een wervingsproces te ontwerpen. Elk proces moet worden aangepast aan de specifieke behoeften van de organisaties maar de belangrijkste overwegingen zijn de volgende: 1) wat is er volgens de functieanalyse vereist voor succesvolle prestaties, 2) hoe beoordelen we de essentiële en wenselijke kenmerken voor de functie op de meest valide manier, 3) hoe gaan we structuur brengen in het wervingsproces en de verzamelde informatie gebruiken om de kandidaten geleidelijk te filteren en 4) hoe en wanneer willen we kandidaten via het wervingsproces betrekken bij de organisatie.

Groepsverschillen en vooroordelen

We gaan ervan uit dat alle beoordelingen voor alle groepen gelijkwaardig moeten zijn. Als er groepsverschillen zijn moeten we dit dus zien als een aanwijzing dat de beoordelingstool op een of andere manier een vertekend beeld geeft of geen eerlijke beoordeling doet voor alle groepen.

Anders dan met fysieke kenmerken zoals lengte of gewicht die rechtstreeks kunnen worden gemeten, hebben we geen directe toegang tot de constructen die worden gemeten door psychometrische beoordelingen. In de plaats hiervan worden kandidaten in deze beoordelingen gevraagd om problemen op te lossen, dingen te vertellen over hun gedrag, oordelen te geven of iets dergelijks, en op basis van de verstrekte informatie trekken wij conclusies over de mate waarin de kandidaten beschikken over de verschillende constructen. In dit proces is het goed mogelijk dat de beoordelingen op de een of andere manier een vertekend beeld geven. In het verleden zijn daar zeker tal van voorbeelden van te vinden.

Moderne testontwikkelingstechnieken hebben echter veel gedaan om deze bron van vooroordelen onder controle te houden.

Een andere interpretatie van gegevens over groepsverschillen is dat zij wijzen op reële verschillen tussen de beoordeelde groepen. We kunnen geneigd zijn om aan te nemen dat

alle mensen gelijkwaardig zijn en dat verschillen wellicht te wijten zijn aan een vertekende beoordeling. In de fysieke wereld bestaan er echter verschillen, dus waarom zouden er geen zijn in de psychologische wereld? Gewoon omdat een meetlint aangeeft dat mannen gemiddeld groter zijn dan vrouwen wil bijvoorbeeld nog niet zeggen dat het meetlint een vertekend beeld geeft. Dit verschil wordt aanvaard omdat het zichtbaar is en we kunnen de lengte direct meten, in tegenstelling tot vele psychologische kenmerken die afgeleid moeten worden uit andere waarnemingen.

Sommige groepsverschillen blijven bestaan ondanks de beste praktijken in testontwikkeling. Op dit gebied werd een toonaangevend onderzoek uitgevoerd onder leiding van de American Psychological Association⁴. De conclusie van dit onderzoek luidde dat sekseverschillen in prestaties op GMA-tests over het algemeen klein bleken te zijn maar dat dit niet het geval was voor etnische groepsverschillen. Hoewel er verschillende redenen werden aangehaald om deze verschillen te verklaren, kon er geen afdoende reden hiervoor worden geïdentificeerd.

Diversiteit en negatieve impact

Diversiteit is een belangrijke kwestie voor veel organisaties omdat ze ernaar streven de ongelijke vertegenwoordiging van verschillende groepen in hun personeelsbestand aan te pakken. In debatten over diversiteit zijn al meermaals veel aspecten van het wervingsproces ter sprake gekomen. Zo kreeg onder andere het oordeel van recruiters veel aandacht waarbij onderzoek wees op talrijke vormen van vooroordelen die invloed kunnen hebben op wervingsbeslissingen en die zijn gebaseerd op informatie uit sollicitatiebrieven, cv's, sollicitatiegesprekken of vergelijkbare technieken.

Psychometrische beoordelingen worden ook al geruime tijd genoemd in discussies over diversiteit. De aard van de gegevens die dergelijke beoordelingen opleveren maakt dat het vrij makkelijk is om te bestuderen in welke mate zij een objectieve beoordeling geven. Net zoals in validiteitsonderzoek, worden er veel onderzoeken uitgevoerd naar groepsverschillen in psychometrische beoordelingsscores en de synthese van deze resultaten is niet altijd eenvoudig. Hoewel er algemene overeenstemming is over het feit dat cognitieve vaardigheidstesten enige verschillen vertonen, vooral wanneer de resultaten worden vergeleken

tussen verschillende etnische groepen, voorspellen ze ook in gelijke mate de prestaties op het werk voor verschillende groepen⁵. Dit wil zeggen dat twee mensen die dezelfde score halen op een GMA-test, ongeveer hetzelfde niveau van prestaties op het werk zullen laten zien ongeacht de groep waartoe ze behoren⁶.

We spreken van negatieve impact wanneer een aanzienlijk kleiner deel van de mensen uit een minderheidsgroep of 'beschermd' groep succesvol is in vergelijking met het deel uit de meerderheidsgroep dat succesvol is. In de VS omvatten beschermde klassen ras, geslacht, mensen van 40 jaar en ouder, religie en handicap om er enkele te noemen. De 'vier/vijfde'

of '80%'regel wordt algemeen gebruikt om te bepalen of er sprake is geweest van negatieve impact. Negatieve impact is niet beperkt tot werving: het kan zich ook voordoen in andere activiteiten zoals opleiding en ontwikkeling, ontslagen en zelfs functioneringsgesprekken. De vier/vijfde-regel heeft een wettelijke status in de Verenigde Staten maar niet in veel andere landen, hoewel het in elk land een nuttig uitgangspunt is om selectieactiviteiten of andere beslissingen met betrekking tot medewerkers te bewaken (bijvoorbeeld in geval van promotie, toewijzing van ontwikkelingskansen enz.).

Aanpak van negatieve impact

Het vermogen van GMA-tests om prestaties op het werk te voorspellen betekent dat hun gebruik vanuit een juridisch standpunt verdedigbaar is in de meeste situaties, maar de resultaten in verschillende groepen kunnen tot gevolg hebben dat de diversiteit in het personeelsbestand afneemt. Dit is duidelijk niet wenselijk omdat organisaties ervoor willen zorgen dat hun personeelsbestand rijk en divers is qua samenstelling en ervaringen.

Een praktische impact van groepsverschillen is dat negatieve impact vaak groter wordt naarmate de slaagscores hoger worden ingesteld. Voor GMA-tests in het bijzonder kan het instellen van een lage slaagscore, op bijvoorbeeld het 25^e of 30^e percentiel, helpen om de negatieve impact tot een minimum te beperken.

Negatieve impact kan een complex gebied zijn en we adviseren organisaties om psychologen en andere deskundigen op dit gebied te raadplegen om meer te weten te komen over manieren om dit te matigen en tot een minimum te beperken. Uitgevers van psychometrische tests adviseren ten stelligste om zich niet te beperken tot GMA-tests als enige bron bij de beoordeling van gegevens waarop een wervingsbeslissing

wordt gebaseerd. Wanneer ze in combinatie met andere relevante beoordelingen worden gebruikt kan hun sterke vermogen om prestaties op het werk te voorspellen echter van grote waarde zijn voor recruiters.

Bij andere vormen van beoordelingen zoals sollicitatiegesprekken, integriteitstests en situatiebeoordelingstests zijn er doorgaans veel minder aanwijzingen voor groepsverschillen. Ze zijn dan misschien minder voorspellend dan GMA-tests, werkgevers vinden ze steeds wenselijker omdat ze de diversiteit in de kandidatenpool in stand houden. Dergelijke beoordelingen kunnen worden gebruikt als een alternatief voor GMA of in combinatie ermee. Wanneer ze in combinatie worden gebruikt, kan de GMA tegelijkertijd of op een ander moment dan de andere beoordelingen worden toegepast, waarna de resultaten van alle beoordelingen gecombineerd kunnen worden voor een gewogen resultaat met minder negatieve impact. Dergelijke strategieën kunnen een bescheiden voordeel geven om de impact van GMA op diversiteit te verminderen maar ze bieden geen totaaloplossing⁷.

Meer voorspellen dan alleen prestaties

Voor een groot deel van het onderzoek naar selectiemethodes is de belangrijkste uitkomst altijd gericht geweest op prestaties op het werk of productiviteit. Meer recentelijk hebben werkgevers belangstelling getoond voor extra gebieden die niet altijd goed begrepen zijn.

Medewerkersbetrokkenheid is voor veel werkgevers een belangrijk aandachtsgebied geworden. Geëngageerde medewerkers krijgen energie van hun werk en zetten zich in voor hun werk en de organisatie. Betrokkenheid houdt ook verband met een scala aan positieve gedragingen op de werkvloer en minder verloop. Gallup heeft medewerkersbetrokkenheid gevolgd in de VS,

en vastgesteld dat doorgaans maar één derde van het personeel aangeeft betrokken te zijn bij hun werk. Wereldwijd is dit cijfer nog veel lager⁸. Niet-betrokken medewerkers presteren misschien niet slechter maar ze kunnen een negatief effect hebben op collega's wat betekent dat problemen met betrekking tot productiviteit zich verder dan de individuele medewerker kunnen verspreiden.

Organisaties proberen steeds meer inzicht te krijgen in de potentiële match tussen kandidaten en de functie, een belangrijke bepalende factor voor betrokkenheid. In deze context gaat match niet over het hebben van kwalificaties, vaardigheden of kennis om het werk te doen, of zelfs niet over de specifieke

gedragingen die iemand doorgaans vertoont, maar of de waarden, houdingen of behoeften van de persoon in kwestie overeenstemmen met die van de organisatie. Meestal wordt dit beoordeeld op organisatorisch niveau (bijvoorbeeld "deelt deze persoon de waarden van de bredere organisatie?"), maar het kan ook meer gesitueerd zijn op een macroniveau zoals de werkeenheid of een specifiek team.

De focus op het vinden van een match betekent een verschuiving van hoofdzakelijk te kijken van 'kunnen doen' naar 'zullen doen'. Het werk van Schmidt en collega's komt tot de conclusie dat de beoordelingen van een match tussen personen en organisaties in het algemeen een lage predictieve validiteit hebben maar wat 'match' precies inhoudt en welke uitkomsten het beïnvloedt, is niet duidelijk. Hoewel er een zwak verband is tussen match en prestatie, toont het wel sterkere verbanden met de intenties om de functie te verlaten en positieve houdingen op de werkvloer⁹. Of er een match is, wordt vaak beoordeeld door te kijken naar de match tussen de waarden van iemand en die van de organisatie, maar de vraag blijft hoe we de waarden op organisatorisch niveau het best begrijpen en hoe deze in verband moeten worden gebracht met de individuele behoeften en verwachtingen van de kandidaat.

Beperkingen van wat we weten

Veel van dit document is gebaseerd op de recente publicatie van Schmidt en collega's over de validiteit van verschillende beoordelingsmethodes. Dit onderzoek is waarschijnlijk het meest uitgebreid in zijn soort, maar het is niet zonder grenzen. Het is belangrijk dat gebruikers van beoordelingen zich bewust zijn van deze beperkingen als ze beoordelingsprocessen willen ontwerpen met de nodige aandacht voor de grenzen van onze wetenschappelijke kennis.

Het feit dat de details van specifieke onderzoeken zoals de organisatie waar ze werden uitgevoerd, de beoordeelde functies enz., vaak verloren gaan, vormt een aanzienlijke beperking van al het meta-analytisch onderzoek. Hoewel de beste meta-analytische onderzoeken de kwaliteit van het onderzoek dat ze gebruiken kunnen evalueren en resultaten overeenkomstig afwegen, is dit niet altijd het geval. Hoewel lokale validatie waarbij een link tussen beoordelingen en prestaties in een specifieke organisatie is vastgesteld, altijd de voorkeur geniet, zijn de kansen hiervoor beperkt tot organisaties die een relatief groot aantal mensen tewerkstellen en aanwerven in vergelijkbare functies. Het vereist ook dat medewerkers voldoende tijd krijgen om hun draai te vinden in de functie, wat betekent dat de beoordeling van de prestaties misschien pas zo'n 12 maanden na de aanwerving kan plaatsvinden.

Verder moeten we ook rekening houden met hoe de beoordelingsresultaten werden toegepast en op welk moment in het selectieproces. Er is al opgemerkt dat werving een proces is dat bestaat uit diverse fases waarin informatie wordt verzameld en beslissingen worden genomen. Er is geen goed model voor dit proces

Waarden zijn een ander gebied dat steeds belangrijker wordt voor organisaties. Naast hun aandacht voor het vinden van een match proberen organisaties meer 'waardebewust' te zijn en naar de individuele sterke punten te kijken. Benaderingen op basis van opvallend sterke punten focussen zich op waarden¹⁰, met de bewering dat individuele sterke punten het gevolg zijn van de identificatie en toepassing van waarden. In wervingscontexten zijn benaderingen op basis van sterke punten geneigd zich te focussen op het unieke patroon van sterke punten van elke individuele kandidaat en hoe dit algemeen gebruikt kan worden binnen de functie en de organisatie. Deze vervangen de meer gevestigde selectiemethodes niet, maar worden als aanvulling daarop gebruikt. Organisaties halen tal van redenen aan om benaderingen op basis van sterke punten te gebruiken waaronder het vinden van een betere match, het creëren van organisatorisch voordeel en toename van diversiteit. Er dient vermeld te worden dat waarden en benaderingen op basis van sterke punten niet werden opgenomen in het onderzoek van Schmidt en collega's, vermoedelijk omdat onderzoek op dit gebied nog in ontwikkeling is.

in onderzoek dat een meer tweedimensionale benadering van validiteit gebruikt; een snapshot van het verband tussen scores op een beoordelingsinstrument en prestaties op het werk. Gegevens kunnen al vroeg in het wervingsproces worden verzameld of later, misschien nadat sommige kandidaten al zijn afgewezen. In nog andere gevallen kunnen gegevens worden verzameld van mensen die al aan de slag zijn in een functie, wat wil zeggen dat er geen gegevens beschikbaar zijn van kandidaten die al zijn afgewezen. Beslissingen over wanneer we verschillende beoordelingsmethodes moeten gebruiken zijn van cruciaal belang, maar onderzoek slaagt er niet altijd in om voldoende gedetailleerde informatie te geven over deze beslissingen.

Tot slot, validiteit wordt afgezet tegen 'prestaties op het werk'. Wat effectieve prestaties op het werk inhouden varieert aanzienlijk tussen functies en organisatie en het is een uitdaging om prestatie zodanig te definiëren dat het gemakkelijk en nauwkeurig kan worden gemeten. Prestaties op het werk zijn vaak gewoon afhankelijk van beoordelingen van de manager, die beïnvloed kunnen worden door meerdere bronnen van vooroordelen. Zoals we al aangaven, moeten andere uitkomsten die belangrijk zijn voor werkgevers nog even grondig aan onderzoek worden onderworpen als 'prestaties op het werk'.

Hoewel het belangrijk is om rekening te houden met deze beperkingen, geeft onderzoek consistent aan dat door een gevoelige toepassing van valide metingen de risico's waar werving mee gepaard gaat aanzienlijk verminderen en de productiviteit van het personeel sterk toeneemt. Recruiters kunnen voor het merendeel van de functies vertrouwen hebben in de algemene resultaten uit validiteitsonderzoek maar moeten deze met de nodige gevoeligheid en aandacht voor de specifieke situatie toepassen voor een maximaal resultaat.

Referenties

- 1 Schmidt, F. L. (2016) The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical implications of 100 Years. Op 16 oktober 2017 gevonden op: www.researchgate.net/publication/309203898
- 2 Van Iddekinge, C. H., Roth, P. L., Raymark, P. H. and Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97, 499-530.
- 3 Anderson, N., Salgado, J. F. and Hülshager, U. R. (2010). Applicant reactions in selection: Comprehensive meta- analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291-304.
- 4 Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R.; Sternberg, R. J. and Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, 51, 77-101.
- 5 Schmidt, F. L. and Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.
- 6 Rothstein, H. R and McDaniel, M.A. (1992). Differential validity by sex in employment settings. *Journal of Business and Psychology*, 7, 45-62.
- 7 Sackett, P. R., Schmitt, N., Ellingson, J. E. and Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56, 302-318.
- 8 Gallup (2013). State of the Global Workplace: Employee Engagement Insights for Business Leaders Worldwide. Washington DC, Gallup.
- 9 Arthur, W. Jr., Bell, S. T., Villado, A. J. and Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology*, 91, 786-801.
- 10 Peterson, C. and Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

