

Handleiding voor het opstellen en gebruiken van beslisregels

Bijlage AST-NIP, 2024



NIP. ψ

Nederlands
Instituut van
Psychologen

Inhoudsopgave

Voorwoord	4
Inleiding	5
Leeswijzer	7
Stappenplan	10
1. Kwalitatieve informatie kwantificeren	10
2. Bepalen hoe verzamelde informatie wordt gebruikt	10
3. Compensatorisch gebruik	11
3.1 Informatie uitgedrukt op verschillende schalen	11
3.1.1 Standardiseren o.b.v. beschikbare informatie	11
3.1.2 Omschalen naar een niet-gestandaardiseerd instrument	12
3.1.3 Omzetten naar gemeenschappelijke ordinale schaal	12
3.1.4 Voor selectievraagstukken: rangorde	13
3.2 Bepalen van gewichten	13
3.2.1 Meta-analytisch/empirisch	14
3.2.2 Direct statistisch	14
3.2.3 Gelijke gewichten	14
3.2.4 Policy capturing	15
3.2.5 Zelfbepaalde gewichten	15
4. Noncompensatorisch gebruik	16
4.1 Bestaande grenswaarden	16
4.2 Direct statistisch	16
4.3 Policy capturing	17
4.4 Expert oordeel	17
5. Beslisregels toepassen	18
Voorbeeld 1: Casus psychologisch rapport kind en jeugd	19
Voorbeeld 2: Casus personeelsselectie	26
Referenties	31

Voorwoord

Het voornaamste doel van het gebruik van tests in de praktijk is om zo goed mogelijke beslissingen te nemen. We hebben dit stuk geschreven omdat er nog altijd weinig aandacht is voor hoe men testresultaten moet wegen bij de besluitvorming en omdat hiervoor weinig tot geen concrete richtlijnen bestaan. Er is terecht veel aandacht voor de kwaliteit van instrumenten, tests en procedures, maar het correct gebruiken en combineren van informatie is even belangrijk voor de kwaliteit van onze beslissingen.

Gelukkig is er veel bekend over hoe tests het best gebruikt kunnen worden; een enorme hoeveelheid onderzoek laat overtuigend zien dat het gebruik van informatie volgens een vooraf opgestelde beslisregel tot betere conclusies en beslissingen leidt dan de gebruikelijke, holistische manier van besluitvorming (door erover na te denken, Aegisdottir et al. 2006; Grove et al., 2000; Hanson & Morton-Bourgon, 2009; Kuncel et al., 2013). Dat deze kennis in de praktijk weinig wordt toegepast heeft ons vaak verwonderd, maar is ook goed verklaarbaar. De meeste bestaande richtlijnen en handboeken zeggen hier namelijk niets over, zijn onjuist of onvolledig, of in ieder geval weinig concreet in richtlijnen rondom testgebruik en beslissen (Meijer et al., 2023). Deze handleiding is bedoeld als een stap om daar verandering in te brengen.

Het opstellen van beslisregels is eenvoudiger dan vaak wordt gedacht; ook zonder (ingewikkelde) statistische analyses van grote datasets zijn beslisregels op te stellen die tot betere uitkomsten leiden dan de holistische aanpak (Karelaia & Hogarth, 2008). Tegelijkertijd is het opstellen van beslisregels minder eenvoudig dan wij en andere onderzoekers op dit terrein aannamen ("*the whole trick is to know what variables to look at and then know how to add*", Dawes, 1979).

Deze handleiding is een eerste poging om de verschillende manieren van beslisregels opstellen te beschrijven en om twee voorbeelden (vanuit de arbeids- en organisatiepsychologie en de kind- en jeugdpsychologie) te schetsen in de vorm van casussen waarin beslisregels worden opgesteld. Het doel is om bij te dragen aan effectief testgebruik en besluitvorming. Er zijn vast aandachtspunten die we beter kunnen adresseren en de voorbeelden in de casussen zijn niet uitputtend. We nodigen daarom iedereen die graag wil bijdragen aan aanvullingen en uitbreidingen van harte uit om contact met ons op te nemen. Ten slotte bedanken we Helen Bakker, Nathan Kuncel en alle COTAN-leden voor de nuttige input bij het schrijven van deze handleiding.

November 2024

Susan Niessen, Petra Hurks, Marvin Neumann en Rob Meijer

Inleiding

Het belang van probleemdefinitie, het formuleren van hypothese(n) en het selecteren van betrouwbare en valide instrumenten om de hypothese(n) te toetsen wordt elders in de Algemene Standaard Testgebruik (AST-NIP, Nederlands Instituut van Psychologen, 2024) en in het reglement Basisaantekening Psychodiagnostiek (Nederlands Instituut van Psychologen, 2020) al uitgebreid beschreven. Op die onderwerpen gaan we hier daarom niet in.

In deze handleiding bieden we praktische richtlijnen en voorbeelden voor het gebruik van beslisregels om verzamelde informatie te combineren en zo tot een beslissing te komen. Het gebruik van beslisregels (ook wel bekend als 'algoritmes' of de mechanische-, actuariële- of statistische methode), is niet populair bij veel praktiserende psychologen. Onderzoek binnen diverse disciplines, waaronder de psychologie (Neumann et al., 2023) toont aan dat een meerderheid van professionals de holistische methode gebruikt. Hierbij combineert de professional alle verkregen informatie (inclusief kennis, ervaring, intuïtie) 'in het hoofd' om tot een verwachting of beslissing te komen. Ten onrechte. Beslisregels leiden vaker tot accurate beslissingen en zijn transparanter, waardoor het gemakkelijker is om na te gaan hoe een beslissing is genomen en hoe deze eventueel kan worden verbeterd.

Veel onderzoek heeft ondubbelzinnig aangetoond dat beslissingen op basis van beslisregels de voorkeur genieten (Grove et al., 2000), maar er is weinig aandacht besteed aan hoe deze beslisregels dan moeten worden opgesteld en gebruikt. Het opstellen van beslisregels is echter niet altijd makkelijk. Dit is mogelijk één belangrijke reden waarom beslisregels weinig worden gebruikt. Het doel van deze handleiding is om psychologen te helpen bij het

opstellen en gebruiken van deze beslisregels. De psycholoog moet hierbij expliciet een aantal keuzes maken, rekening houdend met de beschikbare informatie en eventuele praktische belemmeringen. We beginnen met het beschrijven van de verschillende keuzes die gemaakt moeten worden en verschillende aanpakken, met voor- en nadelen, in de vorm van een stappenplan.

De hoofdthema's zijn:

- Het gebruik van kwalitatieve informatie
- Bepalen hoe de verzamelde informatie wordt gebruikt
- De omgang met informatie uitgedrukt op verschillende schalen
- Het bepalen van gewichten
- Het bepalen van grensscores
- Het gebruik van beslisregels

In dit stappenplan zijn ook korte voorbeelden uitgewerkt. We hebben ervoor gekozen om deze allemaal in dezelfde context (personeelsselectie) te plaatsen, zodat de nadruk ligt op de verschillende manieren waarop beslisregels kunnen worden opgesteld en de stappen die daarbij komen kijken in vergelijkbare situaties. Alle hier beschreven principes zijn echter ook van toepassing in de andere toepassingsgebieden van de psychologie. Vervolgens presenteren we twee uitgewerkte casussen als voorbeeld; één op het terrein van kind en jeugd en één over personeelsselectie.

We willen verder vooraf nog drie zaken aanstippen. Ten eerste is het goed om te bedenken dat als dezelfde vragen worden beantwoord of dezelfde hypothesen worden getoetst zonder daarbij gebruik te maken van een expliciete, vooraf opgestelde en consistent toegepaste beslisregel, dezelfde afwegingen worden gemaakt als

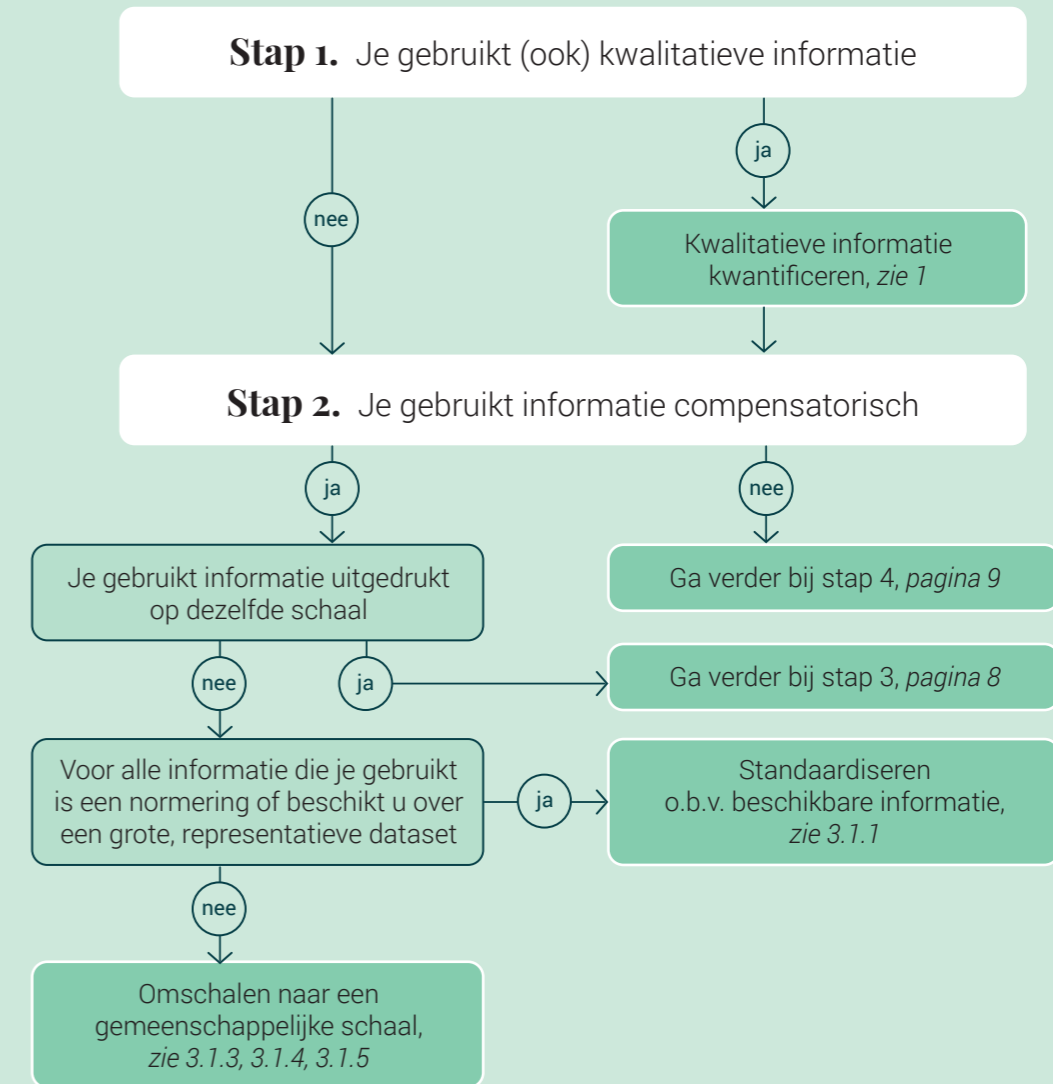
hieronder beschreven. Dat gebeurt dan echter impliciet, 'in het hoofd'. Het expliciteren van hoe we informatie gebruiken op basis van een beslisregel om tot beslissingen te komen, maakt alleen maar duidelijk welke uitdagingen daarbij komen kijken.

Ten tweede, hoewel we hieronder bespreken hoe beslisregels kunnen worden opgesteld voor het combineren van informatie, willen we uitdrukkelijk stellen dat in veel gevallen een beperkt aantal, goed gekozen informatiebronnen de voorkeur geniet. Vaak geldt 'less is more'. Zeer tegenintuïtief kan het gebruik van meer informatie het oordeel namelijk ook verslechteren. In het algemeen gebeurt dat als de 'extra' informatie een lagere validiteit heeft dan en ook nog sterk samenhangt met de bestaande informatie (Murphy, 2019). Wat betreft het gebruik van suboptimale informatie (zoals ongestructureerde interviews), gaat 'baat het niet, dan schaadt het niet', dus niet op (Meijer et al., 2020).

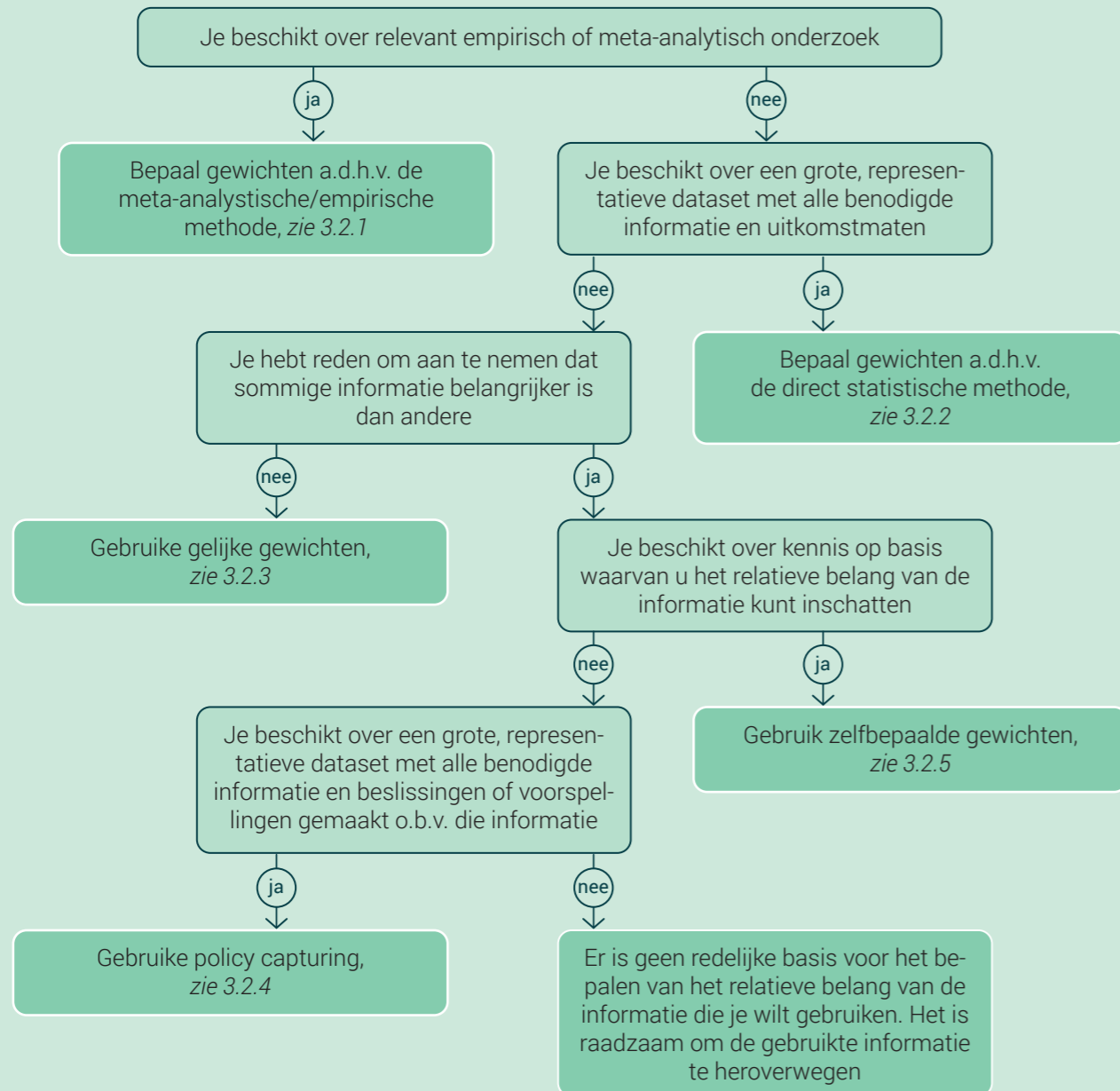
Ten derde maakt onderstaande duidelijk dat het gebruik van instrumenten waarvoor geen geschikte normgegevens beschikbaar zijn, het nodig maakt dat scores op suboptimale wijze worden omgeschaald. Het gebruik van dergelijke instrumenten heeft op die manier soms ook een negatieve invloed op de kwaliteit van scores gebaseerd op instrumenten waarvoor wel normgegevens bestaan, en daarmee de kwaliteit van de beslissingen die genomen worden. Indien mogelijk wordt het gebruik van dergelijke instrumenten daarom vermeden.

Leeswijzer

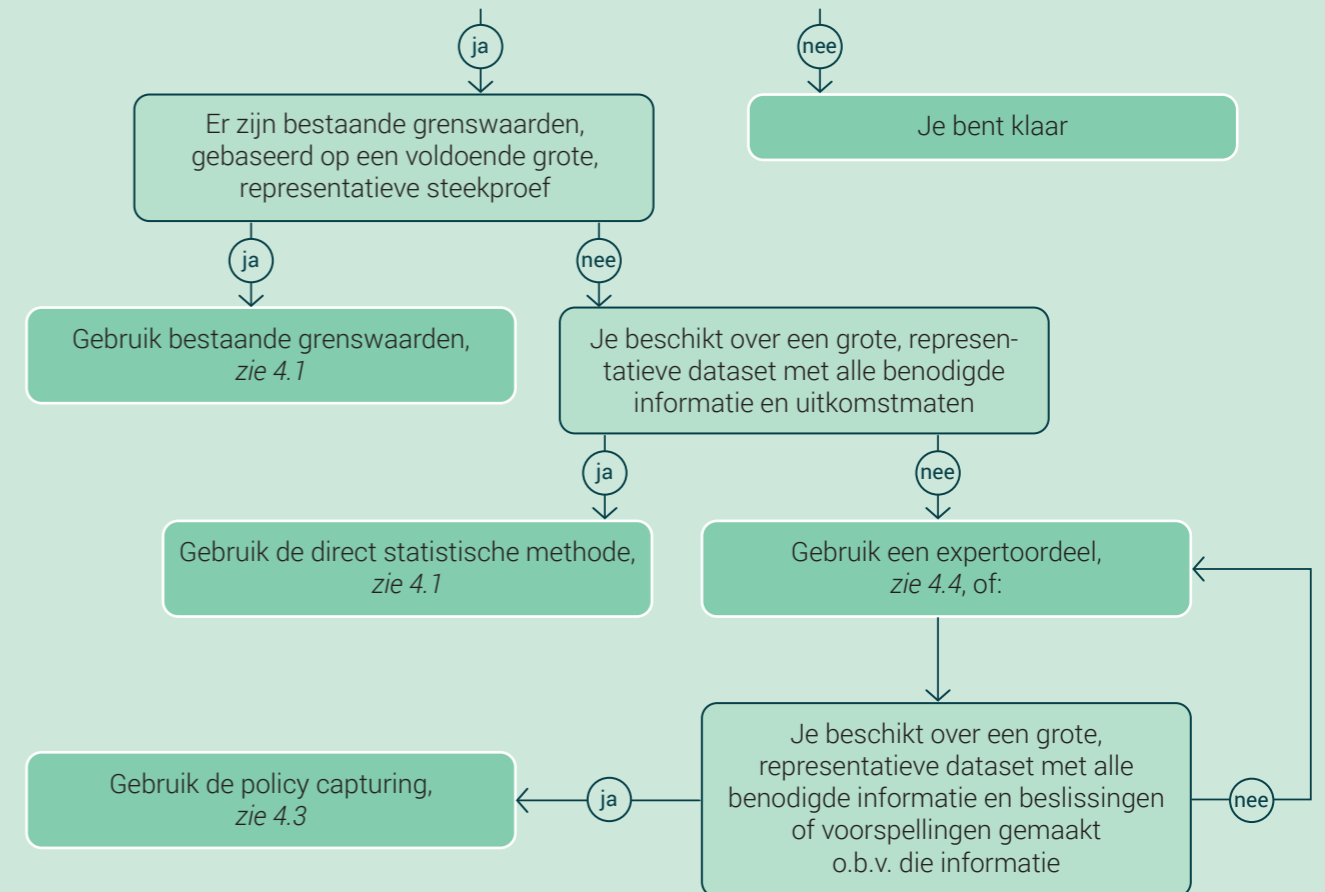
Niet alle onderdelen van het stappenplan zijn relevant voor iedere context of situatie. Deze leeswijzer helpt je bij het vinden van de onderdelen uit het stappenplan die voor jou relevant zijn. Waar mogelijk zijn methoden die minder inzet of middelen vergen geprioriteerd. Deze leeswijzer is niet bindend en dient als richtlijn.



Stap 3. Bepalen van gewichten



Stap 4. Je gebruikt informatie non-compensatorisch (zie 2)



Stappenplan

1. Kwalitatieve informatie kwantificeren

Wanneer er kwalitatieve informatie wordt gebruikt, zoals informatie op basis van observaties of gesprekken, dan kan deze worden meegenomen in een beslisregel. Deze informatie dient dan wel eerst te worden gekwantificeerd. Te denken valt aan: de aan- of afwezigheid van een minimaal benodigde mate van taalbeheersing, de mate of ernst van klachten die passen bij een depressie, of het aantal symptomen passend bij ADHD zoals beschreven in de DSM-5-TR. Afhankelijk van de aard van deze informatie kan dan een passende kwantificatie gebruikt worden. Voorbeelden zijn een binaire codering (0/1) voor aan- of afwezigheid van een eigenschap, klacht, of symptoom, een som van eigenschappen, klachten of symptomen wanneer het aantal ertoe doet, of een score op een Likertschaal wanneer de mate of ernst hiervan in kaart wordt gebracht.

2. Bepalen hoe verzamelde informatie wordt gebruikt

Vervolgens dient te worden nagedacht over hoe de verzamelde informatie wordt gebruikt bij het nemen van de beslissing. Hier beschrijven we drie mogelijkheden:

- Compensatorisch: alle informatie krijgt een bepaald gewicht.
- Noncompensatorisch: er moet een bepaalde grenswaarde bereikt worden.
- Een combinatie: voor sommige informatie moet een bepaalde grenswaarde worden bereikt, andere informatie wordt volgens een bepaalde weging gebruikt, of voor bepaalde informatie moet een bepaalde grenswaarde worden bereikt en wordt dan gewogen gecombineerd met andere informatie.

Indien informatie alleen noncompensatorisch wordt gebruikt, is het bepalen van de grenswaarde(n) de volgende stap (zie stap 4).

Indien informatie (ook) compensatorisch wordt gebruikt, moet worden nagedacht over hoe de informatie wordt gewogen. De volgende stappen hebben hier betrekking op.

3. Compensatorisch gebruik

3.1 Omgaan met informatie uitgedrukt op verschillende schalen

Voordat informatie compensatorisch gecombineerd kan worden volgens een beslisregel, is het noodzakelijk dat alle informatie, zoals testcores en scores op een interview, op eenzelfde schaal wordt geplaatst. Het is niet zinvol om, bijvoorbeeld, een gemiddelde te berekenen voor een capaciteitentestscore, uitgedrukt op een schaal met een gemiddelde van $M = 100$ en een standaarddeviatie van $SD = 15$, en een consciëntieusheidscore met een gemiddelde van $M = 5$ en een standaarddeviatie van $SD = 2$. Als we deze scores combineren zonder rekening te houden met de verschillende schalen, zou de score op de capaciteitentest een veel grotere invloed hebben op de totaalscore dan bedoeld.

Scores verkregen via verschillende instrumenten kunnen op verschillende manieren op één schaal worden geplaatst. We bespreken hieronder een aantal mogelijkheden, grofweg geordend met afnemende wenselijkheid.

3.1.1 Standaardiseren o.b.v. beschikbare informatie

Bij deze aanpak worden de scores op ieder instrument omgezet naar een standardscore, zoals een z-score. Als dat mogelijk is, heeft deze aanpak de voorkeur, omdat hierbij geen informatie verloren gaat. Deze aanpak is mogelijk in de volgende gevallen:

a. Alle informatie is verzameld op basis van instrumenten waarvan normgegevens (in ieder geval gemiddelde en spreiding) voor een relevante referentiegroep beschikbaar is. In dit geval kan alle informatie naar dezelfde schaal omgezet worden met de volgende formule:

$$Score_{nieuw} = (SD_{nieuw} / SD_{oud}) * (X - M_{oud}) + M_{nieuw}$$

Waarbij SD_{nieuw} de standaarddeviatie is van de nieuwe schaal, SD_{oud} de standaarddeviatie van de oorspronkelijke schaal, X de oorspronkelijke score van het individu op het

instrument, M_{oud} het oorspronkelijke gemiddelde op de schaal en M_{nieuw} het gemiddelde van de nieuwe schaal.

Voorbeeld: om een leerling te adviseren over een passend schooltype worden een intelligentietest, een vragenlijst omtrent studiegedrag en een diagnostisch interview gebruikt. Voor de intelligentietest en de vragenlijst omtrent studiegedrag zijn normgegevens voor een relevante referentiegroep beschikbaar (respectievelijk $M = 100$, $SD = 15$ en $M = 5$, $SD = 2$). Het interview resulteert in een score op een schaal van 1 tot 5, berekend als gemiddelde score (5-punts Likertschaal) van 8 onderwerpen die door de psycholoog met de leerling zijn besproken. Er zijn gegevens bijgehouden over eerdere scores op dit interview voor 40 leerlingen met een vergelijkbare vraagstelling ($M = 3.1$, $SD = 1.3$). Er wordt gekozen om de scores op de drie instrumenten te transformeren naar een schaal met $M = 50$ en $SD = 10$. De scores voor ieder instrument op deze schaal worden als volgt berekend:

$$\begin{aligned} \text{Intelligentietestscore} &= (10/15) * (X - 100) + 50 \\ \text{Studiegedrag score} &= (10/2) * (X - 5) + 50 \\ \text{Interviewscore} &= (10/1.3) * (X - 3.1) + 50 \end{aligned}$$

b. Alle informatie is verzameld onder een voldoende grote en voor de vraagstelling relevante, representatieve groep personen, zodat gestandaardiseerde scores, zoals z-scores, berekend kunnen worden.

In dit geval is het mogelijk om de scores op alle gebruikte instrumenten te standaardiseren, bijvoorbeeld door gebruik te maken van z-scores welke met de volgende formule berekend worden:

$$Z = \frac{X - \bar{M}}{SD}$$

waarbij X de ruwe score is van het individu op het instrument, M de gemiddelde score op het instrument en SD de standaarddeviatie van de scores op het instrument. Het is niet mogelijk om algemeen te stellen wat een voldoende grote, relevante en representatieve groep personen is voor een specifieke vraagstelling. Indien de groep

personen op basis waarvan de informatie is verzameld niet representatief is en/of klein is, zullen de schattingen van het gemiddelde en de spreiding minder nauwkeurig zijn. Als algemene richtlijn kan aangehouden worden dat er gegevens van ten minste 30 representatieve personen beschikbaar zijn.

Voorbeeld: om 10 nieuwe klantenservicemedewerkers te selecteren wordt een capaciteitentestscore, een consciëntieusheidscore en de score op een gestructureerd interview gebruikt. Er zijn 40 kandidaten voor de functie. Er zijn geen passende normgegevens beschikbaar voor de tests. De ruwe scores op ieder instrument worden omgezet naar z-sores ($M = 0$, $SD = 1$) waarbij de gemiddelden en standaarddeviaties worden gebruikt op basis van de 40 kandidaten.

c. Als er een combinatie van instrumenten met en zonder beschikbare relevante normgegevens gebruikt wordt in een voldoende grote groep, kan voor een instrument zonder normgegevens eerst standaardisatie worden toegepast (zie b). Indien nodig kunnen de scores op instrumenten waarvoor wel geschikte normgegevens beschikbaar zijn vervolgens naar dezelfde schaal worden omgezet (zie a).

3.1.2 Omschalen naar de schaal van een niet-gestandaardiseerd instrument

Indien er voor een instrument (of meerdere, waarbij dezelfde schaal is gebruikt) geen normgegevens beschikbaar zijn en geen informatie is verzameld onder een voldoende grote en relevante groep personen, kan ervoor worden gekozen om de scores om te schalen naar de schaal van het instrument waarvoor geen normgegevens bestaan. In dit geval worden de scores op alle instrumenten zo omgeschaald dat de onder- en bovengrenzen hetzelfde zijn.

De uiteinden van de schalen zijn dan wel hetzelfde, maar de spreiding zal verschillen. Een nadeel daarvan ten opzichte van het gebruik van gestandaardiseerde scores, zoals hierboven beschreven, is dat verschillen in variantie onbedoelde en mogelijk aanzienlijke verschillen veroorzaken in

het gewicht dat de informatie krijgt. Zelfs als een totaalscore wordt berekend op basis van een beslisregel die, bijvoorbeeld, een gelijk gewicht toekent aan alle informatie, zal informatie met meer spreiding meer invloed hebben op de totaalscore dan informatie met minder spreiding (Murphy, 2019).

$$X_{\text{nieuw}} = \frac{(\max X_{\text{nieuw}} - \min X_{\text{nieuw}})}{(\max X_{\text{oud}} - \min X_{\text{oud}})} \times (X_{\text{oud}} - \max X_{\text{oud}}) + \max X_{\text{nieuw}}$$

waarbij $\max X_{\text{nieuw}}$ de bovengrens is van de nieuwe schaal, $\min X_{\text{nieuw}}$, de ondergrens van de nieuwe schaal, $\max X_{\text{oud}}$ de bovengrens van de oorspronkelijke schaal, $\min X_{\text{oud}}$ de ondergrens van de oorspronkelijke schaal en X_{oud} de oorspronkelijk score op het instrument.

Voorbeeld: om nieuwe klantenservicemedewerkers te selecteren wordt een capaciteitentestscore, een consciëntieusheidscore en de score op een gestructureerd interview gebruikt. Voor de capaciteitentest en de consciëntieusheidschaal zijn normgegevens voor een relevante referentiegroep beschikbaar. Het interview resulteert in een score op een schaal van 1 tot 5, berekend als gemiddelde score (5-punts Likertschaal) op vijf competenties, beoordeeld door twee beoordelaars. Er zijn 4 kandidaten voor de functie, waardoor het berekenen van gestandaardiseerde scores, zoals z-scores, voor het interview niet zinvol is. Daarom worden de scores op de capaciteitentest en de consciëntieusheidschaal op basis van de minimale en maximale scores op deze instrumenten (zoals gerapporteerd in de testhandleidingen) omgezet naar scores tussen 1 en 5, met één decimaal. Merk op dat hier informatie op basis van de capaciteitentest en de consciëntieusheidschaal verloren gaat.

3.1.3 Omzetten naar gemeenschappelijke ordinale schaal

Een andere mogelijkheid is het omzetten van alle verzamelde scores naar een gemeenschappelijke, ordinale schaal, zoals een Likertschaal. Scores binnen een bepaalde range krijgen dan dezelfde ordinale score toebedeeld.

Deze aanpak lijkt eenvoudig, maar een nadeel is dat er informatie verloren gaat door aan verschillende scores binnen een bepaalde bandbreedte dezelfde waarde toe te kennen. Ook kan het lastig zijn om de grenswaarden van iedere categorie te bepalen. Dat zal in de meeste gevallen moeten gebeuren op basis van een inschatting van de te verwachten verdeling en range van scores en een interpretatie van welke scores in de context in kwestie als laag, gemiddeld en hoog kunnen worden gezien.

Voorbeeld: om een nieuwe klantenservicemedewerker te selecteren wordt een score op een capaciteitentest, een score op consciëntieusheid o.b.v. een persoonlijkheidsvragenlijst, en de score op een gestructureerd interview gebruikt. Voor de capaciteitentest en de consciëntieusheidschaal zijn normgegevens voor een relevante referentiegroep beschikbaar. Het interview resulteert in een score op een 5-punts Likertschaal, gegeven door de afdelingsmanager. Er zijn 4 kandidaten voor de functie, waardoor het berekenen van gestandaardiseerde scores, zoals z-scores, voor het interview niet zinvol is. Daarom worden de scores op de capaciteitentest en de consciëntieusheidschaal ook omgezet naar een 5-punts Likertschaal. Dit kan bijvoorbeeld op de volgende wijze:

Cognitieve capaciteiten:

< 85 = 1, 85-90 = 2, 91-110 = 3, 111-120 = 4, > 120 = 5

Consciëntieusheid:

< 36 = 1, 36-45 = 2, 46-55 = 3, 56-60 = 4, > 60 = 5

3.1.4 Voor selectievraagstukken: rangorde

Een andere aanpak kan toegepast worden indien het rangschikken van personen binnen een specifieke groep volstaat. Dat is vaak het geval bij selectievraagstukken waarbij de beste x kandidaten uit een groep van X sollicitanten gekozen worden. In dat geval is het mogelijk om op basis van ieder instrument een rangordening van personen te maken. De rang is dan de gemeenschappelijke schaal. Ook hier is een nadeel dat er informatie verloren gaat; de omvang van het verschil in scores tussen personen wordt niet langer meegenomen.

Voorbeeld: we kiezen weer de selectie van nieuwe klantenservicemedewerkers op basis van cognitieve capaciteiten, consciëntieusheid en de score op een gestructureerd interview. Voor de capaciteitentest en de consciëntieusheidschaal zijn normgegevens voor een relevante referentiegroep beschikbaar. Het interview resulteert in een score op een 5-punts Likertschaal, gegeven door de afdelingsmanager. Er zijn weer 4 kandidaten voor de functie, waardoor het berekenen van gemiddelde en spreiding – vanwege de geringe steekproefomvang – voor de interviewscore niet zinvol is. Daarom wordt voor elk instrument een rangordening gemaakt op basis van de behaalde scores. In het geval van dezelfde score op hetzelfde instrument krijgen die kandidaten dezelfde gemiddelde rang (zie de laatste kolom in onderstaande tabel).

Kandidaat	Score			Rang		
	Cogn. cap.	Consc.	Interview	Cogn. cap.	Consc.	Interview
1	111	51	3	1	2	3.5
2	108	49	4	2	3	2
3	98	55	3	4	1	3.5
4	105	45	5	3	4	1

3.2 Bepalen van gewichten

Voordat informatie gecombineerd kan worden volgens een beslisregel, is het noodzakelijk om te bepalen hoe de verschillende informatie (zoals scores op verschillende tests en een interview) gewogen wordt. Gewichten kunnen op verschillende manieren bepaald worden en geven het relatieve belang van de informatie aan. De schaal waarop gewichten worden weergegeven is vrij te kiezen. Dat kunnen, indien mogelijk, bijvoorbeeld regressiecoëfficiënten zijn, maar ook percentages (verdeel 100% over alle informatie), of een ander aantal punten dat verdeeld wordt. Bij de keuze voor de schaal waarop gewichten worden weergegeven, kan ook het gemak van begrip en communicatie met de cliënt of opdrachtgever meegenomen worden. Bij het gebruik van rangorde kan de (gewogen) gemiddelde rang berekend worden. We bespreken verschillende opties om gewichten te bepalen.

3.2.1 Meta-analytisch/empirisch

Indien beschikbaar kunnen resultaten uit relevante meta-analyses of empirische studies als bron dienen om de gewichten te bepalen. Dat kan bijvoorbeeld door de gerapporteerde correlaties tussen de verschillende typen gebruikte informatie en de criteria in te voeren in een tool zoals die van Failenschmid et al. (2021). Deze tool berekent de optimale gewichten op basis van deze correlaties, of geeft de resulterende validiteit met behulp van deze correlaties en bepaalde zelfgekozen gewichten (zie Murphy, 2019 voor de gebruikte formules).

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt als volgt gewogen:

$$\text{Totaalscore geschiktheid} = (\text{Cognitieve capaciteiten} * 0.60) + (\text{Consciëntieusheid} * 0.25) + (\text{interview} * 0.15).$$

De gewichten zijn hierboven weergegeven als percentagegewichten en zijn tot stand gekomen op basis van een meta-analyse over het verband met baanprestaties en de verbanden tussen de informatie onderling (Cortina et al., 2000). Deze correlaties zijn ingevoerd in een tool om gewichten te bepalen (Failenschmid et al., 2021). De optimale gewichten zijn afgerond om het gebruik en de communicatie te vergemakkelijken. Met behulp van de tool is het effect van afronden gecontroleerd, maar in dit specifieke geval leidde dit niet tot een verandering in predictieve validiteit.

3.2.2 Direct statistisch

Indien een voldoende grote en representatieve dataset met de gebruikte informatie en criteriumgegevens beschikbaar is, kunnen gewichten bepaald worden op basis van statistische analyse, zoals een lineaire regressieanalyse. Wanneer een dataset groot genoeg is, is in het algemeen niet te zeggen, maar bij $n < 200$ is het voordeel van de direct statistische aanpak boven de meta-analytische aanpak of het gebruik van gelijke gewichten (zie onder) in de meeste gevallen verwaarloosbaar (Einhorn & Hogarth, 1975), omdat de gewichten dan niet met voldoende nauwkeurigheid geschat kunnen worden.

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt gebruik gemaakt van een bestaande dataset met daarin de scores op de capaciteitentest, consciëntieusheid, het interview en de beoordeling van het functioneren van medewerkers na één jaar, zoals beoordeeld door de leidinggevende, van 296 medewerkers die in de afgelopen vier jaar zijn aangenomen. Een lineaire regressieanalyse van de dataset resulteert in de volgende gewichten:

$$\text{Totaalscore geschiktheid} = (\text{Cognitieve capaciteiten} * 0.66) + (\text{Consciëntieusheid} * 0.22) + (\text{interview} * 0.12).$$

Om het gebruik en de communicatie met kandidaten te vergemakkelijken zijn de gewichten weergegeven als percentagegewichten (gewichten optellend tot 100, gebaseerd op relatieve gewichtenanalyse, zie Stadler et al., 2017) in plaats van gestandaardiseerde regressiegewichten en zijn de gewichten afgerond op hele percentages. Afronden zorgde niet voor een verandering in predictieve validiteit.

3.2.3 Gelijke gewichten

Indien er geen passende data of onderzoeksresultaten beschikbaar zijn om gewichten te bepalen, kan ervoor gekozen worden om alle informatie hetzelfde gewicht te geven (unit weights). Indien alle informatie in grofweg vergelijkbare mate relevant is en onderling ook positief samenhangt, is de validiteit van gelijke gewichten vaak sterk vergelijkbaar met het gebruik van gewichten geschat op basis van een statistische analyse van empirische data. Bovendien wordt de validiteit vaak overschat o.b.v. een dergelijke analyse; wanneer de beslisregel met deze gewichten op nieuwe gevallen wordt toegepast, blijkt de validiteit vaak lager uit te vallen (overfitting). Deze overschatting is vaak groter bij het gebruik van kleinere datasets om de gewichten te bepalen. Dit probleem doet zich niet voor bij het gebruik van gelijke gewichten. Daarom wordt het gebruik van gelijke gewichten in dit geval aanbevolen (Murphy, 2019).

Voorbeeld: stel dat er voor het selecteren van de klantenservicemedewerkers een cognitieve capaciteitentest en een assessment center wordt gebruikt. Beide instrumenten

zijn ongeveer even relevant (Sackett et al., 2022). De scores op beide instrumenten krijgen daarom hetzelfde gewicht:

$$\text{Totaalscore geschiktheid} = \text{Score cognitieve capaciteiten} + \text{Score assessment center}$$

3.2.4 Policy capturing

Indien er wel een voldoende grote en representatieve dataset beschikbaar is voor de relevante informatie, maar enkel de criteriumgegevens ontbreken, kan er ook voor worden gekozen om een beslisregel op te stellen middels *policy capturing* (ook wel *judgmental bootstrapping of models of man* genoemd, Armstrong, 2001).

In het kort is de aanpak als volgt: Er wordt gebruik gemaakt van een dataset met daarin de verschillende typen informatie die worden meegenomen en de beslissing (oordeel) die, op basis van diezelfde informatie, holistisch is gemaakt. Deze gegevens worden vervolgens gebruikt om een model te maken dat weergeeft hoe de informatie in het algemeen wordt gewogen, bijvoorbeeld via (logistische) regressieanalyse, aangevuld met relatieve gewichten analyse of dominantie analyse (Stadler et al., 2017). De gewichten die voortkomen uit dit model vormen de basis voor de beslisregel. Uit onderzoek blijkt dat dergelijke modellen tot een betere validiteit leiden dan de holistische oordelen waarop ze gebaseerd zijn, omdat de inconsistentie (ruis) in de manier waarop de informatie wordt gebruikt uit de beslisprocedure verwijderd wordt (Dawes, 1979; Karelai & Hogarth, 2008). Een meer gedetailleerde beschrijving is te vinden in Kuncel (2018) en Yu & Kuncel (2022).

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt gebruik gemaakt van scores op de capaciteitentest, consciëntieusheid en scores op een gestructureerd interview. Deze scores en het oordeel over de mate van geschiktheid aan het einde van de sollicitatieprocedure (beoordeeld op een vijf-puntsschaal door verschillende managers), zijn bijgehouden voor 312 medewerkers die de afgelopen vier jaar zijn aangenomen. Om in kaart te brengen hoe de managers de informatie wegen om tot een

oordeel over geschiktheid te komen, wordt een policy capturing analyse uitgevoerd. Daaruit blijkt dat de managers de informatie in het algemeen als volgt wegen:

$$\text{Totaalscore geschiktheid} = (\text{Cognitieve capaciteiten} * 0.46) + (\text{Consciëntieusheid} * 0.31) + (\text{interview} * 0.23).$$

In het vervolg wordt de geschiktheidsscore daarom volgens bovenstaande formule berekend.

3.2.5 Zelfbepaalde gewichten

Als geen van bovenstaande mogelijk is en het gebruik van gelijke gewichten niet redelijk is omdat sommige informatie toch duidelijk veel relevanter is dan andere, kunnen gewichten op basis van eigen kennis over de betrouwbaarheid en validiteit van de gebruikte instrumenten gebruikt worden.

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt gebruik gemaakt van scores op een capaciteitentest, consciëntieusheid en een nieuw ontwikkeld, game-based assessment center. Er zijn geen data beschikbaar over deze scores en voor het game-based assessment center zijn (nog) geen studies of meta-analyses beschikbaar die een goede indicatie van de validiteit of de correlatie met de andere instrumenten geven. Daarom kan de meta-analytische/empirische methode niet gebruikt worden om de gewichten te bepalen. Er is echter reden om aan te nemen dat niet alle informatie even relevant is, want cognitieve capaciteitentests hebben een aanzienlijk hogere predictieve validiteit dan consciëntieusheid (Sackett et al., 2022). Daarom worden de gewichten bepaald op basis van de kennis die een psycholoog heeft over de verschillende instrumenten. Van het game-based assessment center is wel bekend dat de test-hertestbetrouwbaarheid gelijk is aan .85 en dat de scores samenhangen met scores op enkele andere relevante variabelen, gemeten op basis van zelfrapportage. Omdat de score op capaciteitentest een hogere validiteit heeft dan consciëntieusheid en er nog niet voldoende bekend is over de validiteit van het assessment center om een groot gewicht aan die score te rechtvaardigen, kiest de psycholoog voor de volgende gewichten:

$Totaalscore\ geschiktheid = (Cognitieve\ capaciteiten * 0.50) + (Consciëntieusheid * 0.25) + (Game-based\ assessment * 0.25).$

4. Noncompensatorisch gebruik

Zojuist bespraken we hoe informatie compensatorisch gecombineerd kan worden volgens een beslisregel. Hier bespreken we dit voor het noncompensatorisch gebruik van informatie. Het bepalen van een grenswaarde, bijvoorbeeld een minimale score op een test, is relevant als:

a. individuele informatie noncompensatorisch wordt gebruikt.

Individuele informatie wordt noncompensatorisch gebruikt als de beslissing direct afhangt van het bereiken van een bepaalde grenswaarde, of als de beslissing afhangt van de aan- of afwezigheid van een eigenschap, klacht, of symptoom.

b. een eindbeslissing wordt genomen op basis van informatie die eerst compensatorisch is gebruikt.

Ook indien informatie compensatorisch wordt gecombineerd volgens een beslisregel (zie boven), moet uiteindelijk nog besloten worden bij welke scores resulterend uit de beslisregel een bepaalde beslissing wordt genomen. Soms is dat eenvoudig, bijvoorbeeld in het geval van een selectievraagstuk waarin een x aantal personen uit een X aantal kandidaten moet worden gekozen. De grenswaarde voor het al dan niet selecteren (de beslissing) wordt dan bepaald op basis van het x aantal personen dat geselecteerd moet worden. In veel gevallen is er echter geen dergelijke 'natuurlijke' grenswaarde. Het bepalen van grenswaarden kan op verschillende manieren die hieronder uiteengezet worden.

4.1 Bestaande grenswaarden

Voor sommige instrumenten worden grenswaarden gespecificeerd op basis van eerder onderzoek of inhoudelijke argumenten. Wanneer er van bestaande grenswaarden gebruik wordt gemaakt, is het van belang om te reflecteren

op de relevantie en representativiteit van de steekproef op basis waarvan de grenswaarde tot stand is gekomen en op de kwaliteit van de procedure waarmee de grenswaarde is bepaald.

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt op basis van de beschrijving van opleidingsniveaus vanuit de Rijksoverheid besloten dat kandidaten ten minste een mbo-opleiding op niveau 3 afgerond moeten hebben, omdat vanaf dat niveau wordt opgeleid tot zelfstandig uitvoerend werk.

4.2 Direct statistisch

Indien een voldoende grote en representatieve dataset met de informatie en criteriumgegevens beschikbaar is, kan een grensscore bepaald worden op basis van statistische analyse. Dat kan op veel verschillende manieren die hier niet allemaal uiteengezet kunnen worden. Voorbeelden zijn de contrasting groups methode, forward of reverse regressie, en expectancy charts (zie Cascio et al., 1988; Mueller et al., 2007).

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt gebruik gemaakt van een bestaande dataset met daarin de scores op de capaciteitentest, consciëntieusheid, en een gestructureerd interview. Omdat het afnemen van interviews tijdrovend is en de capaciteitentest een hoge predictieve validiteit heeft, wordt in een eerste fase van de selectieprocedure alleen de capaciteitentest afgenomen. Alleen kandidaten die boven een grenswaarde op deze test scoren, gaan door naar de laatste fase van de procedure.

De grenswaarde is bepaald op basis van een dataset met daarin de scores op de capaciteitentest en de beoordeling van het functioneren van medewerkers na één jaar, zoals beoordeeld door de leidinggevende, van 296 medewerkers die in de afgelopen vier jaar zijn aangenomen. Medewerkers met een beoordeling ≥ 3 op functioneren (op een schaal van 1 = onvoldoende, 2 = beneden verwachting, 3 = naar verwachting, 4 = boven verwachting, 5 = excellent)

worden als naar behoren functionerend geclassificeerd. Vervolgens is gekozen voor de reverse regressiemethode, omdat die methode lagere grenswaarden oplevert dan de forward methode, als het gemiddelde functioneren van alle medewerkers boven het niveau van 'naar behoren functioneren' ligt, zoals hier het geval was (Mueller et al., 2007). Een lagere grenswaarde wordt als wenselijker gezien, omdat de organisatie in deze fase van de procedure meer belang hecht aan het voorkomen van vals-negatieve beslissingen dan van vals-positieve beslissingen. Uit deze analyse volgt een grenswaarde van 90. Kandidaten die op de capaciteitentest een score behalen van ≥ 90 gaan daarom door naar de volgende fase van de selectie.

4.3 Policy capturing

Indien er wel een voldoende grote en representatieve dataset beschikbaar is voor de relevante informatie, maar enkel de criteriumgegevens ontbreken, kan er ook voor worden gekozen om een grensscore te bepalen middels policy capturing (ook wel judgmental bootstrapping of models of man genoemd, Armstrong, 2001). Bij deze aanpak wordt gebruik gemaakt van een dataset met daarin de informatie die wordt meegenomen en het oordeel dat op basis van die informatie holistisch is gegeven (Jaeger, 1995). Middels een regressieanalyse, zoals de borderline regression method of de contrasting groups methode (Van der Vleuten & Cohen-Schotanus, 2009), kan dan een grenswaarde worden afgeleid zoals die in het algemeen wordt gebruikt wanneer beslissingen holistisch worden genomen. Die grenswaarde wordt vervolgens consistent toegepast, zodat inconsistentie (ruis) in de manier waarop de informatie wordt gebruikt uit de beslisprocedure verwijderd wordt (Dawes, 1979; Karelaia & Hogarth, 2008).

Voorbeeld: voor het selecteren van de klantenservice-medewerkers wordt gebruik gemaakt van z-scores op de capaciteitentest, consciëntieusheid en scores op een gestructureerd interview. Kandidaten die op een gewogen combinatie van deze instrumenten boven een bepaalde grenswaarde scoren gaan door naar de laatste fase van het selectieproces; een gestructureerd interview met de

afdelingsmanager. De scores op de instrumenten zijn gecombineerd volgens een weging op basis van meta-analytische gegevens (zie 3.2.1). Ook hebben managers op basis van de individuele informatie en de gewogen totaalscore van iedere kandidaat beoordeeld of deze naar hun mening wel of niet moest worden aangenomen. Al deze gegevens zijn bijgehouden voor 312 medewerkers die de afgelopen vier jaar zijn aangenomen.

Om een grenswaarde voor geschiktheid te bepalen is gekozen voor de forward regressiemethode, omdat die methode hogere grenswaarden oplevert dan de reverse methode als de meeste kandidaten als geschikt worden beoordeeld, zoals hier het geval was (Mueller et al., 2007). Een hogere grenswaarde wordt hier als wenselijker gezien, omdat dat voorkomt dat er in de laatste fase van de selectieprocedure veel meer kandidaten geïnterviewd worden door de afdelingsmanager dan er vacatures zijn. Uit deze analyse volgt een grenswaarde van 1.5. Kandidaten met een gewogen totaalscore van ≥ 1.5 gaan daarom door naar de laatste fase van de selectie.

4.4 Expert oordeel

Indien er geen geschikte data of passend eerder onderzoek beschikbaar is, kan gebruik worden gemaakt van het oordeel van experts. Bij deze aanpak wordt vaak het oordeel van verschillende experts in kaart gebracht en samengevoegd om tot een gemeenschappelijke grensscore te komen. Verschillende manieren van aanpak zijn onder andere de *modified of extended Angoff-* en *Nedelskymethode* en de direct consensus methode (Hambleton & Plake, 1995; Mueller et al., 2007). Als meerdere experts niet beschikbaar zijn of een formele procedure voor het vaststellen van grenswaarden niet praktisch realiseerbaar is, kunnen grenswaarden ook op basis van een eigen inschatting bepaald worden (zogenoemde naïeve methoden). Daar zitten veel nadelen aan (Mueller et al., 2007) en die methode wordt daarom niet aangeraden.

Voorbeeld: voor het selecteren van de klantenservicemedewerkers wordt gebruik gemaakt van scores op de capaci-

teitentest, consciëntieusheid en scores op een gestructureerd interview. Voor elk van die instrumenten wordt een grensscore bepaald waaraan kandidaten moeten voldoen om geschikt geacht te worden. Voor de capaciteitentest wordt een minimumscore van $z = -1$ gehanteerd, uitgaande van een normgroep van personen met een mbo-opleiding als hoogst behaald opleidingsniveau. Voor consciëntieusheid en het interview wordt gebruik gemaakt van de extended Angoff methode, omdat die methode geschikt is voor instrumenten die uit polytome items bestaan of op basis van observatieschalen beoordeeld worden. Voor consciëntieusheid vullen de 10 medewerkers en twee managers voor ieder item het antwoord in dat een minimaal gekwalificeerde medewerker zou geven en voor het interview geven ze voor iedere vraag aan onder welke categorie op het beoordelingsformulier het antwoord ten minste zou moeten vallen. Deze scores worden eerst per item gemiddeld. De grenswaarde wordt berekend als de som van de gemiddelden op de items.

5. Beslisregels toepassen

Het volgen van alle bovenstaande stappen leidt tot een beslissing zoals die volgens de opgestelde beslisregel zou moeten zijn. In sommige gevallen is de psycholoog of iemand anders wellicht van mening dat de beslissing zoals die volgt uit de beslisregel toch niet juist is. Uit onderzoek blijkt echter dat wanneer we een beslissing nemen die afwijkt van het resultaat van de beslisregel, die beslissing vaker onjuist is dan wanneer het resultaat van de beslisregel was gevolgd (Ægisdóttir et al., 2006; Dietvorst et al., 2018; Garb & Wood, 2019). Met andere woorden, het afwijken van een beslisregel zorgt in de regel voor minder juiste beslissingen. Achteraf besluiten om toch van de opgestelde procedure af te wijken wordt daarom afgeraden.

Samenvattend, psychologen gebruiken in de praktijk vaak meerdere instrumenten om daarmee zo goed mogelijke beslissingen te nemen. Ondanks de aandacht voor de kwaliteit van instrumenten, tests en procedures, blijft het wegen van testresultaten bij besluitvorming vaak onderbelicht. Onderzoek toont daarbij herhaaldelijk aan dat het gebruik van beslisregels leidt tot betere conclusies en beslissingen dan de gebruikelijke holistische benadering. In het eerste hoofdstuk van deze handreiking gaven we concrete richtlijnen voor het correct gebruiken en combineren van informatie, op een transparante wijze. In de volgende hoofdstukken illustreren we het opstellen van beslisregels – en de keuzes die een psycholoog daarbij kan maken – aan de hand van twee fictieve casussen: één vanuit de kind- en jeugdpsychologie en één vanuit de arbeids- en organisatiepsychologie. Hoewel de voorbeelden in de casussen niet allesomvattend zijn – en we herhaaldelijk benadrukken dat de expertise van de diagnosticus steeds nodig is om expliciete keuzes te maken rondom beslisregels passend bij de casus – hopen we zo een bijdrage te leveren aan een beter begrip en gebruik van beslisregels in de praktijk.

Voorbeeld 1: Casus psychologisch rapport kind en jeugd

Onderwerp: De neuropsychologische evaluatie van een 10-jarig kind met vermoedens van ADHD.

Reden van aanmelding

Dit rapport beschrijft een fictieve casus van een 10-jarige jongen met een vermoeden van ADHD, waarbij ouders de diagnosticus vragen om een psychodiagnostisch onderzoek uit te voeren. Het doel van het onderzoek is om inzicht te krijgen in het gedrag en de cognitieve functies, met name de aandachtregulatie, van het kind.

Ontwikkelingsanamnese

Het kind in kwestie is een 10-jarige jongen die momenteel regulier onderwijs geniet. De ouders hebben zich tot de kliniek gewend vanwege vermoedens van ADHD.

Gedrag: het kind heeft volgens de ouders altijd al moeite gehad met het voltooien van taken en om langdurig geconcentreerd te werken. Hij heeft moeite met stilzitten en zichzelf vermaken, behalve als hij met de Playstation of de iPad mag spelen (“dan kan hij zich gemakkelijk uren vermaken”). Ook merken ouders op dat het kind vaak impulsief reageert en moeite heeft met het volgen van regels en instructies. Daarnaast is het voor hem lastig om zelf een plan van aanpak te bedenken bij een taak. De ouders geven aan dat deze problemen de laatste jaren zijn verergerd en dat het kind daardoor soms onhandelbaar kan zijn. Ook de school herkent het gedrag dat de ouders beschrijven. Daarbij geeft de leerkracht aan dat het kind moeite heeft om aandachtig te luisteren tijdens lessen en dat het kind vaak onrustig is en veel beweegt in de klas.

De ouders en leerkrachten hebben verschillende strategieën geprobeerd om de “uitdagingen qua gedrag” aan te pakken, waaronder het aanbieden van extra begeleiding, structuur en duidelijke regels. Het kind lijkt gevoelig voor het consistent aanbieden van directe beloningen en straffen, zowel thuis als op school. Ondanks dat, blijft het kind qua gedrag moeilijk te sturen en zijn er zorgen over de toekomst van het kind.

Schoolloopbaan: tot op heden regulier (groep 1-2-3-4-5-6). Op school hebben leerkrachten vaker opgemerkt dat het kind moeite heeft met het aanleren en automatiseren van nieuwe taken. Hij lijkt volgens de ouders moeite te hebben om zichzelf hierbij te motiveren. Zowel de school als de ouders denken dat “er meer inzit dan dat het kind nu vaak laat zien”. Qua lezen en rekenen valt op dat het kind te snel lijkt te willen gaan, waardoor er fouten optreden in het schoolwerk. Het kind geeft aan een hekel te hebben aan lezen en zal niet snel uit zichzelf gaan lezen of in een boek kijken. Het handschrift is moeilijk leesbaar, ondanks extra oefeningen aangeboden door school op het gebied van schrijven en lezen. Het kind heeft moeite met het organiseren van zijn huiswerk; ouders moeten hier iedere dag bij helpen.

Ondanks deze uitdagingen presteert het kind over het algemeen op het gemiddelde niveau op de schoolse toetsen. Echter, het vergt aanzienlijke inspanning voor het kind om zich te concentreren tijdens deze toetsen. Wanneer mogelijk staat de leerkracht toe dat het kind de toetsen aflegt in een afzonderlijke, prikkelarme ruimte, omdat dit lijkt te helpen met de concentratie, zoals de leerkracht heeft opgemerkt. Het kind kan van streek raken als de opzet van de toets anders is dan verwacht. In zulke gevallen heeft het kind vaker moeite met het maken van de toets, tenzij de leerkracht het kind geruststelt en “even op weg helpt”.

Sociaal functioneren: de ouders geven aan dat het kind sociaal actief is en vrienden heeft, maar dat hij soms moeite heeft om te communiceren met leeftijdsgenoten (“andere kinderen begrijpen niet altijd wat hij zegt; hij is vaak warrig in zijn verhaal”). Hij wil net zo goed presteren als zijn leeftijdsgenoten; als hij neigt te falen, stagneert hij vaak en wordt hij “buitenproportioneel boos”.

Familiale voorgeschiedenis: de ouders geven aan dat er geen gedrag- en leerproblemen voorkomen in de familie en dat het kind geen lichamelijke problemen heeft. Ouders hebben beiden een universitaire opleiding genoten.

Hypotheses

Hieronder presenteren we, ter illustratie, twee hypothesen en bijbehorende criteria voor de toetsing ervan. We zijn ons ervan bewust dat in de beschreven context een diagnosticus waarschijnlijk meer uitgebreide hypothesen en toetsingscriteria zou hebben opgesteld als het geen fictieve casus betrof. Bovendien pretenderen we niet dat de geselecteerde instrumenten het meest geschikt zijn om de hypothesen te beantwoorden. We gebruiken deze fictieve casus hoofdzakelijk om het concept van besluitvorming op basis van beslisseregels (de mechanische besluitvorming) en de implicaties van implementatie in de (klinische) praktijk te verduidelijken.

Hypothese 1:

Het gedrag van het kind is passend bij de DSM-5-TR classificatie ‘Aandachttekortstoornis met Hyperactiviteit’ (ADHD).

Onderbouwing: Ouders en leerkrachten benoemen aandachtproblemen, moeite met stilzitten en impulsregulatieproblemen bij het kind. Dit gedrag wordt al langere tijd door hen waargenomen. Deze gedragsbeschrijving is passend bij de classificatie ‘Aandachttekortstoornis met Hyperactiviteit’ (ADHD) zoals beschreven in de DSM-5-TR (American Psychiatric Association, 2022). De impact van dit gedrag op het dagelijks leven van het kind is niet volledig duidelijk, maar ouders geven wel aan dat het kind “buitenproportioneel boos” kan worden.

Hypothese 2:

De ADHD-problemen komen voort uit beneden gemiddelde executieve functies, en meer specifiek een beneden gemiddelde inhibitie.

Onderbouwing: De term “executieve functies” verwijst naar een set van gerelateerde cognitieve functies die betrokken zijn bij het reguleren van het denken en gedrag (Diamond, 2013). Executieve functies zijn cruciaal voor het dagelijks functioneren en problemen met deze executieve functies worden in wetenschappelijk onderzoek vaak geassocieerd met onder andere lagere academische prestaties, problemen in sociale interacties, en een verminderd algemeen welzijn van kinderen (Meuwissen & Zelazo, 2014). Problemen met executieve functies kunnen (maar dat hoeft niet) zich manifesteren bij ADHD (Kofler et al. 2019). Onder andere Biederman en collega’s (2004) vonden dat het gecombineerd hebben van zowel ADHD als executieve functieproblemen samenhangt met een verhoogd risico op aanzienlijke beperkingen in bijvoorbeeld academisch functioneren (zoals meer kans op doubleren en lagere academische prestaties) in vergelijking met het hebben van “alleen ADHD”.

Inhibitie is een belangrijke executieve functie. Hoewel er veel meer unieke executieve functies zijn die mogelijk ook relevant zijn in deze casus, dient inhibitie in dit geval als voorbeeld. Inhibitie omvat het vermogen om impulsen, automatische reacties of ongewenst gedrag te onderdrukken of te remmen. Het houdt in dat men weerstand kan bieden aan verleidingen, afleidingen en onmiddellijke behoeften, en in plaats daarvan kan handelen op basis van doelgericht gedrag of sociale normen.

Toetsingscriteria per hypothese

Om de hypothesen te testen, maakt de diagnosticus gebruik van verschillende tests en vragenlijsten om het gedrag en de cognitieve functies van het kind te beoordelen. Hier volgt een korte beschrijving van deze tests en de toetsingscriteria voor elke hypothese. De tests die in deze casus worden genoemd, zijn slechts gekozen als voorbeeld

van hoe beslissingen in deze context genomen kunnen worden. Het is, zoals eerder gezegd, belangrijk op te merken dat we niet beweren dat deze tests de enige mogelijke of geschikte tests zijn voor het toetsen van de hypothesen die hier genoemd worden.

Hypothese 1:

Om hypothese 1 te toetsen, maakt de diagnosticus gebruik van twee instrumenten, namelijk de ADHD-module van de SCID-5 Junior en de leerkrachtenvragenlijst van de Strength and Difficulties Questionnaire.

De *ADHD-module* van de *SCID-5 Junior* (Wante et al., 2021) is een semigestructureerd interview, dat wordt afgenomen bij de ouders. Dit interview verzamelt onder andere informatie over de kenmerkende eigenschappen van ADHD, om te bepalen of het kind voldoet aan de DSM-5 beschrijving van ADHD-symptomen, de ontwikkeling van de symptomen en de impact van deze symptomen op het dagelijks leven. Het interview is gebaseerd op de DSM-5 (American Psychiatric Association, 2013). Hoewel de SCID-5 Junior volgens Wante en collega’s (2021) betrouwbaar en valide is om DSM-5-diagnoses bij kinderen en adolescenten te classificeren, is deze niet beoordeeld door de COTAN.

Zorgstandaarden voor ADHD, zoals die aangeboden door Richtlijnen Jeugdhulp en Jeugdbescherming (2018) en de GGZ Standaarden (2019), bevelen aan om bij de ADHD diagnostiek gesprekken te voeren met de ouders, de leerkracht, en indien mogelijk met het kind zelf. Deze casus wijkt hierin af, omdat alleen het interview bij de ouders wordt afgenomen.

De *Strength and Difficulties Questionnaire leerkrachtenvragenlijst* (SDQ; Diepenmaat et al., 2014) is een gedragsbeoordelvragenlijst die door de leerkracht wordt ingevuld. De vragenlijst verschaft informatie over de aanwezigheid en de ernst van verschillende gedragsaspecten, waaronder ADHD-gerelateerd gedrag. Hoewel o.a. de Richtlijnen Jeugdhulp en Jeugdbescherming (2018) aangeven dat dergelijke vragenlijsten niet noodzakelijk zijn voor de

diagnostiek van ADHD –zij adviseren een semigestructureerd interview- stellen ze dat dergelijke vragenlijsten “een aanvullende rol spelen op het oordeel van de professional”. Ze kunnen namelijk het gedrag van jeugdigen “breed in kaart brengen” en specifieke ADHD-symptomen meten.

In deze casus wordt de SDQ afgenomen om informatie te verzamelen over de zienswijze van de leerkracht (aangezien er geen semigestructureerd interview met de leerkracht is gehouden). De vragenlijst omvat verschillende subschalen, waaronder Emotionele problemen, Gedragsproblemen, Hyperactiviteit en aandachtsproblemen en Problemen met leeftijdsgenoten. In het kader van deze onderzoeksvraag concentreert de diagnosticus zich voornamelijk op de “*Hyperactiviteit en aandachtsproblemen*”-schaal. De COTAN heeft de betrouwbaarheid en validiteit van deze leerkrachtenvragenlijst als voldoende tot goed beoordeeld, met uitzondering van de criteriumvaliditeit. Volgens Goodman (2001) worden scores boven een bepaalde drempelwaarde (> 90e percentiel) als ‘hoog/verhoogd’ (klinisch) beschouwd.

Gedragsobservaties: De diagnosticus overwoog observaties van het kind tijdens het onderzoek als onderdeel van Hypothese 1. Het idee was om waardevolle theoretische informatie te verkrijgen, zoals taakgerichtheid of de mogelijkheid om stil te zitten tijdens de testafname. Het ideaal zou zijn dat de diagnosticus een gestandaardiseerd observatieformulier gebruikt, met normering en uitgebreide betrouwbaarheids- en validiteitsonderzoeken. Zo’n formulier biedt een gestructureerde methode om relevante observaties vast te leggen en te kwantificeren.

Echter, in deze casus is er geen standaardformulier beschikbaar voor de diagnosticus om observaties tijdens de uitgevoerde tests systematisch te registreren. De diagnosticus heeft ervoor gekozen om primair te vertrouwen op gestandaardiseerde psychodiagnostische tests zoals de SCID-Junior en SDQ, in lijn met de zorgstandaarden en wetenschappelijk onderzoek. Dit komt doordat niet-systematische observaties aanzienlijke psychometrische

beperkingen hebben in vergelijking met gestandaardiseerde tests (Meyer et al., 2001). Het meewegen van deze minder betrouwbare niet-systematische observaties zou de kans op een juiste conclusie kunnen verminderen, wat bekend staat als *het dilution effect* (Niessen et al., 2022). Het meewegen van die suboptimale informatie zou immers ten koste gaan van de weging/invloed van informatie van hogere kwaliteit (de scores op de instrumenten). Hier geldt daarom het principe *less is more*.

In situaties waarin een systematisch observatieformulier niet beschikbaar is, kan de diagnosticus ervoor kiezen om zijn eigen systematische observatieschema te creëren voor de casus. Echter, dit schema is niet getoetst op psychometrische kwaliteiten en mist normgegevens voor vergelijking met gedrag van andere kinderen tijdens een vergelijkbare testafname. Het is dan van cruciaal belang dat de diagnosticus transparant is over de specifieke aspecten die worden geobserveerd, hoe deze worden gekwantificeerd en hoe deze observaties worden meegewogen in het diagnostisch proces. Omdat het opstellen van een kwalitatief goed observatieprotocol om deze redenen niet haalbaar was, heeft de diagnosticus in deze casus ervoor gekozen om geen eigen observatieschema op te stellen en te gebruiken.

Beslisregel hypothese 1

In deze casus overweegt de diagnosticus het belang van een vals positieve en een vals negatieve conclusie. De vraag die de diagnosticus zich daarbij stelt is: wat is erger, onterecht een classificatie geven of onterecht geen classificatie geven? Als toetsingscriterium voor het eerste scenario wordt beoordeeld of het gedrag van het kind overeenkomt met de DSM-5 classificatie 'ADHD', conform de richtlijnen van de SCID-5 Junior, én of de score van de door de leerkracht ingevulde SDQ-schaal "*Hyperactiviteit en Aandachtsproblemen*" boven de drempelwaarde ligt zoals vastgesteld door Goodman (2001).

Voor het tweede scenario zou een regel kunnen zijn dat aan één van beide toetsingscriteria voldaan moet zijn, en

niet per se beide. In dit voorbeeld wordt verondersteld dat terughoudendheid met diagnoses gewenst is. Daarom wordt de eerste hypothese aangenomen als aan beide criteria is voldaan.

Hypothese 2

Het testen van de tweede hypothese met betrekking tot de executieve functies, specifiek gericht op de inhibitie van het kind, is uitdagend omdat er weinig instrumenten beschikbaar zijn die voldoen aan alle algemene standaarden van betrouwbaarheid, validiteit en normen van goede kwaliteit om inhibitie te meten bij Nederlandstalige kinderen. Er rijst dan de vraag of de beschikbare instrumenten voldoende kwaliteit bieden om een uitspraak te doen over deze specifieke executieve functie. Zo niet, dan kunnen we de tweede hypothese niet toetsen.

In deze casus is besloten om twee instrumenten te gebruiken om hypothese 2 te testen, namelijk de D-KEFS en de BRIEF-2. Hoewel beide instrumenten sterke punten hebben, kennen ze ook beperkingen. Bijvoorbeeld, de D-KEFS is niet genormeerd voor het Nederlandstalige gebied en de BRIEF-2 is een vragenlijst in plaats van een prestatietaken. Desondanks concludeert de diagnosticus dat het testen op basis van deze instrumenten de voorkeur verdient boven het niet onderzoeken van de hypothese.

In deze casus maakt de diagnosticus gebruik van zowel prestatietaken (in dit geval de D-KEFS) als ouderbeoordelingen (de BRIEF-2) om specifieke executieve functies te meten. Onderzoek toont consistente, relatief lage correlaties tussen ouderbeoordelingen en prestatietests, wat meestal wordt geïnterpreteerd als het meten van verschillende aspecten van cognitief functioneren (Friedman & Banich, 2019). De prestatietaken worden afgenomen door de diagnosticus in een gecontroleerde, prikkelarme setting. De diagnosticus kiest ervoor om ouderbeoordelingen te gebruiken, omdat ouders de cognitieve sterktes en zwaktes van hun kind in het dagelijkse leven kunnen beoordelen. Dit biedt waardevolle inzichten in hoe de cognitieve functies

van het kind van invloed zijn op zijn algeheel functioneren, zowel thuis als op school. Hoewel leerkrachtbeoordelingen en zelfbeoordelingen door het kind relevante informatie hadden kunnen verschaffen over hoe het kind en de leerkracht naar de cognitieve functies kijken, zijn deze in dit geval niet meegenomen.

Beschrijving van de instrumenten

Drie subtests van de D-KEFS (Delis-Kaplan Executive Function System) en de BRIEF-2 (Behavior Rating Inventory of Executive function) worden afgenomen om het inhibitie vermogen van het kind in kaart te brengen. De keuze voor deze specifieke tests dient louter ter illustratie van de procedure voor mechanische besluitvorming.

D-KEFS

De D-KEFS is ontworpen om verschillende executieve functies te evalueren (Delis et al., 2001) en bestaat uit diverse subtests, allemaal prestatietaken. De volgende D-KEFS subtests worden gebruikt:

1. Color-Word Interference Test: deze subtest evalueert onder andere inhibitie door te meten in hoeverre het kind in staat is automatische reacties te onderdrukken.

2. Tower Test: hierbij worden planning en probleemoplossende vaardigheden gemeten. Deelnemers moeten gekleurde schijven herschikken om een doelconfiguratie te bereiken, waarbij specifieke regels gevolgd moeten worden. Inhibitie kan een rol spelen doordat deelnemers de neiging moeten onderdrukken om acties uit te voeren die niet overeenkomstig zijn met de regels (zoals meer dan 1 schijf per keer in je hand houden).

3. Design Fluency Test: deze test meet het vermogen van het kind om unieke non-verbale ontwerpen te genereren en cognitieve flexibiliteit door deelnemers te vragen zoveel mogelijk unieke en niet-repetitieve ontwerpen binnen een tijdslimiet te maken. Hierbij kan inhibitie van belang zijn om herhaling van ontwerpen te voorkomen

en in plaats daarvan nieuwe, verschillende ontwerpen te genereren.

De ruwe scores van de subtests worden omgezet naar standardscores (range: 1-19, M = 10, SD = 3), conform de handleiding. Hoewel inhibitie vooral wordt gemeten door de Color-Word Interference Test, beïnvloedt het ook de prestaties op de Tower en de Design Fluency tests, zij het in mindere mate. Bij de hypothese met betrekking tot inhibitie weegt de diagnosticus daarom de prestaties van het kind op de Color-Word Interference Test, met name conditie 3 (inhibitie) en conditie 4 (inhibitie/switching) het zwaarst. De D-KEFS subtests zijn niet door de COTAN beoordeeld. Uit de Nederlandstalige handleiding blijkt dat de normen verzameld zijn in de Verenigde Staten en niet in Nederland. Dat geldt ook voor het onderzoek naar de betrouwbaarheid. Voor 10-jarigen is de interne consistentie op deze subtests voldoende. Echter, de test-hertest betrouwbaarheid is niet specifiek voor 10-jarigen onderzocht, maar voor de groep van 8-19-jarigen (Color-Word Interference Test: variërend van 0.77-0.90 afhankelijk van de uitkomstmaat; Tower: 0.51; Design Fluency: variërend van 0.13-0.66 afhankelijk van de uitkomstmaat). Over het algemeen presteren kinderen beter bij de hertest. De begripsvaliditeit is slechts beperkt onderzocht, en niet specifiek per normgroep uitgevoerd. De criteriumvaliditeit is ook niet onderzocht.

BRIEF-2

De BRIEF-2 ouderversie (Behavior Rating Inventory of Executive Function versie 2, Huizinga & Smidts, 2020) is een vragenlijst voor ouders, gebruikt om de executieve functies van kinderen van 5 tot 18 jaar te beoordelen. De BRIEF-2 ouderversie bestaat uit verschillende subschalen die specifieke executieve functies meten. De 8 subschalen van de BRIEF-2 zijn: 1) inhibitie, 2) zicht op eigen gedrag, 3) flexibiliteit, 4) emotieregulatie, 5) uit zichzelf beginnen, 6) werkgeheugen, 7) plannen en ordenen en 8) taakgedrag. Daarnaast kunnen vier indexscores worden berekend op basis van de 8 subschalen: de gedragsregulatie (GRI)-, emotieregulatie (ERI)-, en cognitieve-regulatie (CRI)-index en een totaalscore (GEC). Elke subschaal van de BRIEF-2

wordt beoordeeld aan de hand van meerdere items door ouders. Om inzicht te krijgen in de executieve functies van een kind, worden de ruwe scores van de BRIEF-2 omgezet naar T-scores ($M = 50$, $SD = 10$). In deze casus wordt enkel de GRI (als maat voor inhibitie) gebruikt om de hypothese van de diagnosticus te toetsen. We gebruiken deze hogere orde schaal omdat de twee subschalen die hieraan ten grondslag liggen beide een beroep doen op inhibitie.

De BRIEF-2 is niet beoordeeld door de COTAN. De interne consistentie en test-hertest betrouwbaarheid lijken voor Nederland onvoldoende voor de GRI (interne consistentie: ca. 0.61; test-hertest betrouwbaarheid: 0.67) bij 9-10-jarigen. Hierbij is de test-hertest betrouwbaarheid berekend over de gehele leeftijdsgroep van 5-18 jaar. Verder is het onderzoek naar de begripsvaliditeit in Nederland niet uitgevoerd per specifieke normgroep en niet onderzocht met andere soorten instrumenten dan vragenlijsten. Ook is de criteriumvaliditeit niet onderzocht. De meeste resultaten in het uitgevoerde validiteitsonderzoek in het buitenland waren echter wel conform verwachtingen.

Overwegingen

De uitdaging is nu hoe om te gaan met het onderzoek van een hypothese gebaseerd op scores op instrumenten waarvan de betrouwbaarheid, validiteit en/of normgegevens niet overtuigend zijn, of waarvan onvoldoende bekend is over de psychometrische kwaliteit. Meestal wordt bij de beperkingen in bijv. de normen en validiteit aangeraden om minder nadruk te leggen op (of gewicht toe te kennen aan) scores die problematisch zijn en meer op scores met sterke(re) evidentie voor validiteit en/of betere normgegevens. In dit geval zijn echter beide instrumenten, naar onze inschatting, en in vergelijkbare mate beperkt.

Verder wordt vaak geadviseerd om rekening te houden met de precisie van scores. Bijvoorbeeld, wanneer er sprake is van geringe betrouwbaarheid, kan de breedte van het betrouwbaarheidsinterval worden meegenomen in de interpretatie (Zie de AST-NIP, Nederlands Instituut van Psychologen, 2024). In het geval van hypothesetoetsing, diagnostiek en besluitvorming is dit echter minder bruikbaar; een hypothese wordt aangenomen of niet, waarna in zekere mate aandacht wordt besteed aan problemen rondom specifieke executieve functies van het kind in het vervolgtraject. Door de keuze om de hypothese te testen met behulp van meerdere instrumenten, moeten we ons houden aan de scores die daaruit voorkomen. Er is simpelweg geen andere relevante informatie beschikbaar waar meer gewicht (waarde) aan kan worden gehecht. Bij het testen van de hypothese met betrekking tot de inhibitie van het kind is de diagnosticus wederom terughoudend en prioriteert hij het voorkomen van een vals-positieve beslissing boven het voorkomen van een vals-negatieve beslissing.

Beslisregel hypothese 2

Voor hypothese 2 (het kind heeft inhibitieproblemen) kan de diagnosticus verschillende benaderingen toepassen. We geven hier twee voorbeelden, die voornamelijk dienen als illustratie van compensatorisch en noncompensatorisch gebruik van de verzamelde informatie bij mechanische besluitvorming, en niet als voorbeelden van 'best practices'. Het is uiteindelijk aan de diagnosticus om op een transparante wijze te rechtvaardigen hoe hij gegevens combineert om tot een beslissing te komen.

Illustratie van een noncompensatorische beslisregel

In dit scenario moet een bepaalde grenswaarde worden bereikt op één of meerdere tests. Een voorbeeld hiervan is dat de diagnosticus in deze casus de hypothese aanneemt als alle scores op de drie subtests van de D-KEFS in het potentieel klinisch gebied vallen (< 5.5), en daarnaast de gecombineerde GRI op de ouderversie van de BRIEF-2 ten minste in het potentieel klinische gebied valt (> 75).

Illustratie van een compensatorische beslisregel

Hierbij krijgt alle verzamelde informatie een bepaald gewicht. Er zijn geen data of onderzoeksgegevens beschikbaar om de gewichten te bepalen. Daarom zijn de wegingen bepaald op basis van het oordeel van de diagnosticus. Gezien de relatief lage betrouwbaarheid van de GRI van de BRIEF-2 in vergelijking tot de betrouwbaarheden gevonden voor de D-KEFS, kan de diagnosticus overwegen om de D-KEFS-scores – vooral Stroop conditie 3 en conditie 4 (omdat deze meer inhibitie meten) – zwaarder te laten wegen dan de GRI van de BRIEF-2, zoals hieronder wordt beschreven:

*Totaalscore = BRIEF-2 ouder GRI * .10 + D-KEFS Color-Word Interference Test conditie 3 * .30 + D-KEFS Color-Word Interference Test conditie 4 * .30 + D-KEFS Tower * .15 + D-KEFS Design fluency * .15.*

Om een samengestelde score te berekenen, moeten de scores op de verschillende instrumenten eerst op dezelfde schaal worden gebracht. Dat wordt gedaan door de gestandaardiseerde scores op beide instrumenten om te zetten naar z-scores, met $M = 0$ en $SD = 1$ (zie stap 3). Verder duiden hoge scores op de BRIEF-2 juist op problemen met executieve functies zoals inhibitie, terwijl dat voor de D-KEFS lage scores zijn. Om dat te uniformiseren, wordt het teken voor de D-KEFS-scores omgekeerd (positief wordt negatief en vice versa). De omgeschaalde scores worden vervolgens gebruikt om bovenstaande gecombineerde score te berekenen.

De diagnosticus zal vervolgens een drempelwaarde moeten vaststellen om te bepalen wanneer de hypothese wordt aangenomen. Dit kan bijvoorbeeld zijn als de totaalscore ten minste 1.4 standaarddeviaties onder het gemiddelde ligt, wat op beide instrumenten als '(potentieel) klinisch gebied' wordt aangeduid.

Voorbeeld 2:

Casus personeelsselectie

Onderwerp: Vacature klantenservicemedewerker

Opdracht

Een groot telecombedrijf heeft regelmatig vacatures voor klantenservicemedewerkers, waarop naar verwachting vaak meerdere kandidaten solliciteren. De organisatie vraagt hulp van een arbeids- en organisatiepsycholoog bij het selecteren van de meest geschikte kandidaat.

Intake

Er zijn in deze situatie twee vragen die beantwoord moeten worden.

1) Welke kandidaten zijn in voldoende mate geschikt voor de functie?

De onderzoeksvraag die daarbij hoort is, voor iedere kandidaat: Zal deze kandidaat naar voldoende tevredenheid presteren in de functie van klantenservicemedewerker?

Indien het antwoord op deze vraag voor geen enkele kandidaat 'ja' is, wordt de vacature opnieuw opengesteld. Indien slechts één kandidaat voldoende geschikt wordt geacht, krijgt die kandidaat de functie aangeboden. Indien meerdere kandidaten in beginsel geschikt worden geacht, moet de volgende vraag beantwoord worden:

2) Welke kandidaat is het meest geschikt en krijgt de functie aangeboden?

De onderzoeksvraag die daarbij hoort is: Welke kandidaat zal het beste presteren in de functie van klantenservicemedewerker?

Om de eerste vraag te beantwoorden wordt voor iedere kandidaat de hypothese getoetst dat de kandidaat voldoende geschikt is voor de functie. Om de tweede vraag te beantwoorden worden alle kandidaten die in beginsel geschikt worden geacht, geordend van meest tot minst geschikt. Om de eerste hypothese te toetsen moeten minimumvereisten gespecificeerd worden voor voldoende geschiktheid. Om de tweede vraag te beantwoorden moeten relevante eigenschappen die de mate van geschiktheid bepalen gespecificeerd worden en moet besloten worden hoeveel gewicht er aan die verschillende eigenschappen wordt toegekend.

Onderzoeksoptzet

De eerste onderzoeksvraag, "Welke kandidaten zijn geschikt voor de functie?", is een onderkennende of onderscheidende vraag, die is bedoeld om vast te stellen of een kandidaat wel of niet geschikt is. Het antwoord op deze onderzoeksvraag biedt echter beperkte inzichten in waarom iemand mogelijk ongeschikt is of wie van alle kandidaten de meest geschikte is. De tweede onderzoeksvraag (Welke kandidaat zal het beste presteren in de functie van klantenservicemedewerker?) is een selecterende vraag.

Om zowel de minimumeisen voor geschiktheid als eigenschappen die de mate van geschiktheid bepalen te specificeren, wordt een functie-analyse uitgevoerd (Brannick et al., 2017). Drie leidinggevenden en 10 collega's zijn gevraagd op een lijst met taken en eigenschappen (gebaseerd op O*NET, *National Center for O*NET Development*, 2022), aan te geven hoe belangrijk ze zijn voor de functie.

Na samenvoeging van de reacties wordt vastgesteld dat de volgende eigenschappen en vaardigheden het belangrijkste zijn:

- Analytisch vermogen;
- Nauwkeurig en zelfstandig werken;
- Servicegerichtheid;
- Goede mondelinge en schriftelijke beheersing van de Nederlandse taal.

In samenspraak met de organisatie wordt besloten dat de minimumvereisten noncompensatorisch zijn. Daarom dienen er grenswaarden bepaald te worden die gebruikt worden om voor iedere kandidaat te bepalen of die in beginsel geschikt wordt geacht (zie 4.1). Om de grenswaarden te bepalen is gebruik gemaakt van de *extended Angoff methode*, omdat deze methode goed past als een functie-analyse, ook gebruikt kan worden voor instrumenten die polytoom gescoord worden en tot betrouwbare en valide grenswaarden leidt (Hambleton & Plake, 1995; Mueller et al., 2007). Deze methode is toegepast onder dezelfde drie leidinggevenden en 10 collega's.

Op basis van deze functie-analyse is besloten dat er tenminste sprake moet zijn van een mbo-opleiding op niveau 3- zoals gedefinieerd door de rijksoverheid, omdat vanaf dit niveau opgeleid wordt tot zelfstandig werken. Kandidaten die daar niet aan voldoen, worden niet in overweging genomen voor de functie (zoals ook vermeld in de vacaturetekst).

De instrumenten die gebruikt worden om de criteria te beoordelen zijn:

Opleidingsniveau

Het opleidingsniveau wordt beoordeeld aan de hand van het cv van de kandidaat. Indien de functie wordt aangeboden zal om een diploma of getuigschrift worden gevraagd ter verificatie.

Analytische vaardigheden

Voor een beoordeling van de analytische vaardigheden van kandidaten wordt de *Q1000 Capaciteiten Midden* (volledige, digitale versie, Eelloo, 2016) gebruikt. Cognitieve capaciteitentests worden veel gebruikt om analytische vaardigheden te meten en de scores op deze tests zijn sterk gerelateerd aan functioneren in alle typen banen (Sackett et al., 2022). Cognitieve capaciteiten zijn daarnaast ook voorspellend in homogene groepen, zoals sollicitanten die zijn voorgeselecteerd op opleidingsniveau (Kuncel & Hezlett, 2010). Op deze vacature kunnen bovendien kandidaten solliciteren met verschillende onderwijsachtergronden, van mbo tot wo. Daarom is een relevante mate van spreiding te verwachten op deze test. De *Q1000 Capaciteiten Midden* is een prestatietest die algemene cognitieve capaciteiten in kaart brengt (verbaal, cijfermatig en figuraal).

De variant *Midden* wordt gekozen omdat deze geschikt is voor afname bij mensen met een mbo niveau 3 opleiding. Bovendien heeft de COTAN deze test op alle zeven kwaliteitscriteria ten minste als voldoende beoordeeld en is dit de enige cognitieve capaciteitentest die past bij deze doelgroep en deze context, waarvoor ook de criteriumvaliditeit als voldoende is beoordeeld (Egberink et al., 2009-2024), wat in deze context erg relevant is. Op basis van de Angoff-methode is bepaald dat een kandidaat voldoende geschikt wordt geacht bij een z-score ≥ 0 , op basis van de normgroep mbo-niveau 3.

Nauwkeurig en zelfstandig werken

Voor de beoordeling van het vermogen tot nauwkeurig en zelfstandig werken wordt de persoonlijkheidstrek consciëntieusheid gemeten, gebruikmakend van de digitale zelfrapportage lijst NEO-PI-3 (Hoekstra & De Fruyt, 2014). Om logistieke redenen wordt de gehele vragenlijst afgenomen, maar enkel de score op consciëntieusheid wordt meegenomen in de beoordeling van geschiktheid, omdat de andere vier persoonlijkheidsfactoren niet inhoudelijk gerelateerd zijn aan de functie-eisen zoals gespecificeerd op basis van de functie-analyse (Bartram, 2005). Bovendien is consciëntieusheid van alle Big Five persoonlijkheidskenmerken de

beste voorspeller voor professioneel functioneren en heeft deze persoonlijkheidseigenschap incrementele validiteit bovenop cognitieve capaciteiten (Sackett et al., 2022; Schmidt & Hunter, 1998).

De NEO-PI-3 wordt door de COTAN tenminste voldoende beoordeeld op zes van de zeven kwaliteitscriteria. Alleen op criteriumvaliditeit is de beoordeling onvoldoende, wegens gebrek aan onderzoek (Egberink et al., 2009-2024). Hoewel de criteriumvaliditeit één van de belangrijkste kwaliteitscriteria is in deze context, is er geen alternatief instrument beschikbaar dat wel ten minste voldoende wordt beoordeeld op criteriumvaliditeit. Wel vermeldt internationaal meta-analytisch onderzoek een klein, maar praktisch relevant verband tussen scores op consciëntieusheid zoals gemeten met vergelijkbare vragenlijsten en latere prestaties in banen (Sackett et al., 2022). Daarbij moet worden opgemerkt dat het meeste onderzoek niet in 'high-stakes' situaties (zoals sollicitatieprocedures) is uitgevoerd. Mogelijk zijn de criteriumvaliditeit en incrementele validiteit in dit soort situaties lager (Niessen et al., 2017; Schilling et al., 2021), als gevolg van sociaal-wenselijk antwoorden. Er is echter geen sluitende oplossing voor dit probleem. Op basis van de extended Angoff-methode is bepaald dat een kandidaat geschikt wordt geacht bij een T-score ≥ 50 , op basis van de algemene bevolkingsnormen.

Mondelinge- en schriftelijke taalvaardigheid en servicegerichtheid

Deze vaardigheden worden gezamenlijk beoordeeld op basis van een assessment center dat bestaat uit gestandaardiseerde praktijkopdrachten gebaseerd op deze drie competenties. Uiteindelijk volgt hieruit een gezamenlijke totaalscore. Daarvoor is gekozen omdat beoordelingen op individuele componenten van assessment centers vaak niet voldoende betrouwbaar zijn en empirisch ook niet van elkaar zijn te onderscheiden (Wirz et al., 2020). Wel hebben assessment center scores in het algemeen een goede predictieve validiteit voor het functioneren in banen (Sackett et al., 2022) en hebben assessment centers gericht op communicatieve vaardigheden incrementele

validiteit bovenop scores op cognitieve capaciteitentests en scores op vragenlijsten over persoonlijkheidseigenschappen, zoals consciëntieusheid (Dilchert & Ones, 2009).

Om de schriftelijke taalvaardigheid te beoordelen wordt aan kandidaten gevraagd om drie e-mails van fictieve klanten te beantwoorden. De e-mails worden onafhankelijk door twee leden van de sollicitatiecommissie beoordeeld op de volgende componenten: correcte spelling en grammatica, passende toon, duidelijke structuur van de tekst en correcte opmaak (aanhef, afsluiting). Deze componenten worden voor iedere e-mail gescoord volgens een rubric, met voor iedere component drie niveaus (1 = onvoldoende, 2 = voldoende, 3 = goed). De niveaus zijn gedefinieerd middels een beschrijving (bijv. spelling en grammatica: goed = minder dan twee fouten in spelling, werkwoordsvormen en interpunctie, voldoende: tussen twee en vier fouten in spelling, werkwoordsvormen en interpunctie, onvoldoende: meer dan vier fouten in spelling, werkwoordsvormen en interpunctie). De gemiddelde score van de twee beoordelaars wordt gebruikt als de totaalscore op dit onderdeel.

Om de mondelinge taalvaardigheid te beoordelen wordt aan kandidaten gevraagd om drie telefonische vragen van drie fictieve klanten te beantwoorden. De antwoorden worden onafhankelijk door twee leden van de sollicitatiecommissie beoordeeld op de volgende componenten: correct taalgebruik (grammatica/vervoegingen), taalgebruik passend bij de ontvanger, inhoud afgestemd op de ontvanger, passende toon, passend stemgebruik en passende non-verbale communicatie. Deze componenten worden voor ieder antwoord gescoord volgens een rubric, met voor iedere component drie niveaus (1 = onvoldoende, 2 = voldoende, 3 = goed). De niveaus zijn gedefinieerd middels een beschrijving (bijv. correct taalgebruik: goed = minder dan twee fouten in vervoegingen, verwijzwoorden, verbuigingen, voldoende: tussen twee en vier fouten in vervoegingen, verwijzwoorden, verbuigingen, onvoldoende: meer dan vier fouten in vervoegingen, verwijzwoorden, verbuigingen). De gemiddelde score van de twee beoordelaars wordt gebruikt als de totaalscore op dit onderdeel.

Servicegerichtheid wordt beoordeeld aan de hand van de e-mails en mondelinge antwoorden op de vragen uit de assessments voor mondelinge- en schriftelijke taalvaardigheid. Daarbij worden de volgende componenten beoordeeld:

Mondelinge opdrachten: het begrip van de vraag wordt getoetst, de vraag is inhoudelijk beantwoord, controle of de vraag beantwoord is, het antwoord wordt uitgelegd of onderbouwd.

Schriftelijke opdrachten: verwijzen naar/herhalen van (delen van) de vraag, de vraag is inhoudelijk beantwoord, aanbieden om vervolgvragen te beantwoorden, verwijzen naar aanvullende informatiebronnen/benoemen dat men ergens op terugkomt na navraag, het antwoord wordt uitgelegd of onderbouwd.

Deze componenten worden door twee leden van de sollicitatiecommissie (andere leden dan die taalvaardigheid hebben beoordeeld) gescoord volgens een rubric, met voor ieder component drie niveaus (1 = onvoldoende, 2 = voldoende, 3 = goed). De niveaus zijn gedefinieerd middels een beschrijving (bijv. het begrip van de vraag wordt getoetst: goed = de vraag wordt samengevat en juist begrip wordt gecontroleerd, voldoende: de vraag wordt samengevat, onvoldoende: de vraag wordt niet samengevat en er wordt geen begrip getoetst). De gemiddelde score van de twee beoordelaars wordt gebruikt als de totaalscore op dit onderdeel.

Om tot een totaalscore op het assessment center te komen wordt een gemiddelde score over de drie componenten (mondelinge- schriftelijk taalvaardigheid en servicegerichtheid) en de twee beoordelaars berekend. Op basis van de *extended Angoff-methode* is bepaald dat een kandidaat geschikt wordt geacht als er ten minste een totaalscore van 2 wordt behaald.

Procedure en beslisregel

Kandidaten die niet het vereiste opleidingsniveau hebben, krijgen bericht dat zij niet aan de vereisten voor de functie voldoen en daarom niet uitgenodigd worden voor verdere deelname aan de selectieprocedure. Kandidaten die wel het vereiste opleidingsniveau hebben worden uitgenodigd om de cognitieve capaciteitentest (Q1000 Capaciteiten Midden) en de persoonlijkheidsvragenlijst (NEO-PI-3) te maken. Kandidaten die op beide instrumenten een score halen waarmee ze in beginsel geschikt bevonden worden (zie boven), worden uitgenodigd voor deelname aan het assessment center. Voor deze getrapte aanpak wordt gekozen omdat het afnemen van een assessment center relatief tijdrovend is.

Indien na afname van alle instrumenten meerdere kandidaten geschikt worden bevonden, wordt voor iedere kandidaat een totaalscore bepaald. Dat gebeurt op compensatorische wijze. Om tot een totaalscore te komen moeten de scores op de cognitieve capaciteitentest, consciëntieusheid en het assessment center eerst op dezelfde schaal worden geplaatst (zie stap 3). Omdat er niet voor alle instrumenten normgegevens beschikbaar zijn, er geen scores zijn bijgehouden van eerdere sollicitanten voor vergelijkbare functies, verwacht wordt dat slechts een klein aantal personen (≤ 6) zullen overblijven in de laatste fase van het selectieproces, en deze procedure relatief eenvoudig uit te voeren is, wordt ervoor gekozen om voor ieder instrument een rangordening te maken.

De totale 'geschiktheidsscore' voor iedere kandidaat wordt vervolgens berekend als de gewogen gemiddelde rang (zie stap 3), waarbij de rang op basis van het assessment center twee keer zo zwaar weegt als de rang op basis van cognitieve capaciteiten en consciëntieusheid. Hier wordt voor gekozen omdat op cognitieve capaciteiten en consciëntieusheid al een voorselectie heeft plaatsgevonden, waardoor er restrictie in range optreedt en het belang van deze twee voorspellers in het maken van onderscheid in geschiktheid tussen kandidaten daardoor is afgenomen.

Op het assessment center is nog geen voorselectie geweest en zal dus het beste onderscheid kunnen maken in geschiktheid tussen kandidaten die de laatste selectiefase hebben bereikt.

De kandidaat met de hoogste totale geschiktheidsscore (hoogste rang) krijgt de functie aangeboden. Mocht deze kandidaat het aanbod niet accepteren, dan wordt de functie aangeboden aan de kandidaat met de één na hoogste geschiktheidsscore.

Referenties

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382.
<https://doi.org/10.1177/0011000005285875>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders – Fifth edition* (DSM-5).
<https://doi.org/10.1176/appi.books.9780890425596>
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders – Text revision* (DSM-5-TR).
<https://doi.org/10.1176/appi.books.9780890425787>
- Armstrong, J. S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting. In J. S. Armstrong (Ed.), *Principles of Forecasting* (Vol. 30, pp. 171–192). Springer US.
https://doi.org/10.1007/978-0-306-47630-3_9
- Bartram, D. (2005). The Great Eight Competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*, 1185–1203.
<https://doi.org/10.1037/0021-9010.90.6.1185>
- Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., Morgan, C. L., & Faraone, S. V. (2004). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *Journal of consulting and clinical psychology, 72*, 757–766.
<https://doi.org/10.1037/0022-006X.72.5.757>
- Brannick, M. T., Pearlman, K., & Sanchez, J. I. (2017). Work Analysis. In *Handbook of Employee Selection* (pp. 134–161). Routledge.
<https://doi.org/10.4324/9781315690193-6>
- Cascio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology, 41*, 1–24.
<https://doi.org/10.1111/j.1744-6570.1988.tb00629.x>
- Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325–351.
<https://doi.org/10.1111/j.1744-6570.2000.tb00204.x>
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582.
<https://doi.org/10.1037/0003-066X.34.7.571>
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan executive function system*. Psychological Corporation.
- Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135–168.
<https://doi.org/10.1146/annurev-psych-113011-143750>
- Diepenmaat, A., Van Eijsden, M., Janssens, J., Loomans, E., & Stone, L. (2014). *Verantwoording SDQ Leerkrachtvragenlijst*. GGD Amsterdam/Radboud Universiteit Nijmegen.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*, 1155–1170.
<https://doi.org/10.1287/mnsc.2016.2643>

- Dilchert, S., & Ones, D. S. (2009). Assessment Center Dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment, 17*, 254–270. <https://doi.org/10.1111/j.1468-2389.2009.00468.x>
- Eelloo. (2016). Q1000 *Capaciteiten Midden*. Meurs HRM.
- Egberink, I. J. L., De Leng, W. E., & Vermeulen, C. S. M. (2009-2024). *COTAN Documentatie*. Boom Uitgevers. <https://www.cotandocumentatie.nl>
- Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171–192. [https://doi.org/10.1016/0030-5073\(75\)90044-6](https://doi.org/10.1016/0030-5073(75)90044-6)
- Failenschmid, J., Neumann, M., Meijer, R. R., & Niessen, A. S. M. (2021). *The predictive validity of weighted combinations of predictors and criteria*. [R Shiny app]. <https://failenschmid-neumann-meijer-niessen-2021.shinyapps.io/publication/>
- Friedman, N. P., & Banich, M. T. (2019). Questionnaires and task-based measures assess different aspects of self-regulation: Both are needed. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 116*, 24396–24397. <https://doi.org/10.1073/pnas.1915315116>
- Garb, H. N., & Wood, J. M. (2019). Methodological advances in statistical prediction. *Psychological Assessment, 31*, 1456–1466. <https://doi.org/10.1037/pas0000673>
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 1337–1345.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30. <https://doi.org/10.1037/1040-3590.12.1.19>
- GGZ Standaarden. (2019). <https://www.ggzstandaarden.nl/zorgstandaarden/adhd/info/status>
- Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff Procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41–55. https://doi.org/10.1207/s15324818ame0801_4
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1–21. <https://doi.org/10.1037/a0014421>
- Hoekstra, H., & De Fruyt, F. (2014). *NEO-PI-3 en NEO-FFI-3 persoonlijkheidsvragenlijsten*. Hogrefe.
- Huizinga, M., & Smidts, D. P. (2020). *BRIEF-2 Vragenlijst executieve functies voor 5- tot en met 18-jarigen*. Hogrefe.
- Jaeger, R. M. (1995). Setting performance standards through two-stage judgmental policy capturing. *Applied Measurement in Education, 8*, 15–40. https://doi.org/10.1207/s15324818ame0801_3
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404–426. <https://doi.org/10.1037/0033-2909.134.3.404>
- Kofler, M. J., Irwin, L. N., Soto, E. F., Groves, N. B., Harmon, S. L., & Sarver, D. E. (2019). Executive functioning heterogeneity in pediatric ADHD. *Journal of abnormal child psychology, 47*, 273–286. <https://doi.org/10.1007/s10802-018-0438-2>
- Kuncel, N. R. (2018). Judgment and decision making in staffing research and practice. In D. S. Ones, N. Anderson, C. Viswesvaran & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work & organizational psychology: Personnel psychology and employee performance* (2nd ed., pp. 474–488). Sage Reference.
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*, 339–345. <https://doi.org/10.1177/0963721410389459>
- Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology, 98*, 1060–1072. <https://doi.org/10.1037/a0034156>
- Meuwissen, A. S., & Zelazo, P. D. (2014). Hot and cool executive function: Foundations for learning and healthy development. *Zero to Three, 35*, 18–23.
- Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A tutorial on mechanical decision-making for personnel and educational selection. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.03002>
- Meijer, R. R., Niessen, S., & Neumann, M. (2023). Psychological and educational testing and decision making: The lack of knowledge dissemination in textbooks and test guidelines. In L. A. Van der Ark, W. H. M. Emons, & R. R. Meijer (Eds.), *Essays on contemporary psychometrics* (pp. 47–67). (Methodology of educational measurement and assessment). Springer. https://doi.org/10.1007/978-3-031-10370-4_3
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E.J., Kubiszyn, T.W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128–165.
- Mueller, L., Norris, D., & Oppler, S. (2007). Implementation based on alternate validation procedures: Ranking, cut scores, banding, and compensatory models. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 349–405). John Wiley & Sons.
- Murphy, K. R. (2019). Understanding how and why adding valid predictors can decrease the validity of selection composites: A generalization of Sackett, Dahlke, Shewach, and Kuncel (2017). *International Journal of Selection and Assessment, 27*, 249–255. <https://doi.org/10.1111/ijsa.12253>
- National Center for O*NET Development. (2022). Geraadpleegd op 4 oktober 2022, van <https://www.onetonline.org/>
- Nederlands Instituut van Psychologen. (2020). Reglement Basisaantekening Psychodiagnostiek (BAPD NIP). <https://psynip.nl>
- Nederlands Instituut van Psychologen. (2023). Algemene Standaard Testgebruik (AST-NIP). <https://psynip.nl>
- Neumann, M., Niessen, A. S. M., Hurks, P. P. M., & Meijer, R. R. (2023). Holistic and mechanical combination in psychological assessment: Why algorithms are underutilized and what is needed to increase their use. *International Journal of Selection and Assessment, 31*, 267–285. <https://doi.org/10.1111/ijsa.12416>
- Niessen, A. S. M., Kausel, E. E., & Neumann, M. (2022). Using narratives and numbers in performance prediction: Attitudes, confidence and validity. *International Journal of Selection and Assessment, 30*, 216–229. <https://doi.org/10.1111/ijsa.12364>
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences, 106*, 183–189. <https://doi.org/10.1016/j.paid.2016.11.014>
- Richtlijnen jeugdhulp en jeugdbescherming. (2018). <https://richtlijnenjeugdhulp.nl/adhd/>

adhd-signalering-screening-en-diagnostiek/
adhd-diagnostiek/

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *The Journal of applied psychology, 107*, 2040–2068. <https://doi.org/10.1037/apl0000994>

Schilling, M., Becker, N., Grabenhorst, M. M., & König, C. J. (2021).

The relationship between cognitive ability and personality scores in selection situations: A meta-analysis. *International Journal of Selection and Assessment, 29*, 1–18. <https://doi.org/10.1111/ijsa.12314>

Schmidt, F. L., & Hunter, J. E. (1998).

The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>

Stadler, M., Cooper-Thomas, H. D., & Greiff, S. (2017).

A primer on relative importance analysis: Illustrations of its utility for psychological research. *Psychological Test and Assessment Modeling, 59*, 381–403.

Van der Vleuten, C. P. M., & Cohen-Schotanus, J. (2009).

Standard setting. In N. Patil & L. K. Chan (Eds.), *Assessment in medical and health sciences education* (pp. 62–71). Institute of Medical and Health Sciences Education, Li Ka Shing Faculty of Medicine, the University of Hong Kong.

Wante, L., Braet, C., Bögels, S., & Roelofs, J. (2021).

SCID 5 Junior: Gestructureerd klinisch interview voor DSM-5 stoornissen bij kinderen en adolescenten. Boom uitgevers.

Wirz, A., Melchers, K. G., Kleinmann, M., Lievens, F., Annen, H., Blum, U., & Ingold, P. V. (2020).

Do overall dimension ratings from assessment centres show external construct-related validity? *European Journal of Work and Organizational*

Psychology, 29, 405–420.

<https://doi.org/10.1080/1359432X.2020.1714593>

Yu, M. C., & Kuncel, N. R. (2022).

Testing the value of expert insight: Comparing local versus general expert judgment models. *International Journal of Selection and Assessment, 30*, 202–215. <https://doi.org/10.1111/ijsa.12356>

Het Nederlands Instituut van Psychologen

Het Nederlands Instituut van Psychologen (NIP) is met ruim 17.000 leden al meer dan 85 jaar de grootste vereniging van psychologen in Nederland.

Het NIP behartigt de belangen van academisch opgeleide psychologen, bevordert de inzet van psychologische kennis, versterkt de positie van psychologen van alle beroepsniveaus en in alle werkvelden en ondersteunt haar leden in hun beroepsuitoefening.

Het NIP is de beroepsvereniging van alle wetenschappelijk opgeleide psychologen in Nederland, vanaf hun studie, tijdens hun hele loopbaan en in ieder werkveld. Het NIP verenigt en ondersteunt psychologen, bepaalt de professionele norm van het vak en positioneert de psychologen en de psychologie.

Lees meer over het NIP op de website: psynip.nl

© NIP / november 2024

Niets van deze uitgave mag worden verveelvoudigd of gekopieerd zonder toestemming van het NIP.