

Vraagtekens bij kwaliteit assessment instrumenten

Astrid Hazelzet

Annemiek Mul

Een goede beoordeling van individuele arbeidsmogelijkheden is bij (re)integratie van arbeidsongeschikten, chronisch zieken en gehandicapten van grote betekenis. Het gebruik van goede instrumenten is een belangrijke voorwaarde voor het leveren van betrouwbare beoordelingen en bruikbare adviezen. Maar welke instrumenten worden in de praktijk toegepast? En hoe zit het met de kwaliteit?

In het artikel 'Assessment bij (re)integratie: de stand van zaken', dat in november in *Arbeidsomstandigheden* verscheen, is in hoofdlijnen beschreven welke werkwijzen organisaties hanteren bij assessment van individuele arbeidsmogelijkheden. Daarbij maken zij gebruik van een aantal instrumenten, die hieronder worden besproken.

Eind 1995 heeft TNO-PG een enquête uitgezet onder een groot aantal organisaties actief op het terrein van assessment. Doel van het door het Ministerie van SZW gefinancierde onderzoek: een globaal overzicht krijgen van de praktijk van assessments. Arbeidsonderzoekscentra, arbeidsexploratiecentra, (re)integratieprojecten, arbeidsrehabilitatieprojecten en expertisebureaus leverden gegevens aan over ruim 50 assessment instrumenten. Deze instrumenten zijn door de onderzoekers ingedeeld in zes categorieën. Met elkaar geven zij een representatief beeld van het instrumentgebruik in de niet-commerciële praktijk van assessments bij (re)integratie.

Hoe laten de verschillende typen instrumenten zich het beste (kort) omschrijven?



Kandidaat assessment moet zich herkennen in (re)integratie-advies.

foto: Chris Pennarts

Een *vragenlijst* is een verzameling vragen of opdrachten die achter elkaar worden aangeboden en waarop de respondent een antwoord/reactie geeft die hij/zij kiest uit een beperkt aantal mogelijkheden. Er zijn vragenlijsten die het maximale prestatieniveau bepalen (tests) en vragenlijsten die gedragswijzen vaststellen. De enquête leverde 17 reacties op die tot de categorie vragenlijst zijn gerekend.

Een *worksample* is een simulatie van functies die beoogt gedrag op te roepen dat zo veel mogelijk lijkt op het te voorspellen gedrag. In de regel wordt niet de gehele functie nabootst, maar wordt gebruik gemaakt van representatieve elementen uit een functie. Meestal worden fysieke handelingen gevraagd en beoordeeld. Twaalf beschrijvingen zijn in deze categorie ingedeeld.

Instrumenten voor assessment van *fysiek functioneren* zijn instrumenten die aan de hand van gestandaardiseerde onderzoeksprotocollen en met gebruik van (soms technisch hoogstaande) apparatuur aspecten van fysiek functioneren in kaart brengen. De testleider vervult een centrale rol bij de toepassing van deze instrumenten. Zeven instrumentbeschrijvingen passen in de categorie fysiek functioneren.

De *checklist* is een verzameling vragen of opdrachten die achter elkaar worden behandeld en waarop het individu of de testleider een ant-

woord geeft. Vier reacties zijn in de categorie checklist ingedeeld.

Het (*gestructureerd*) *gesprek* is de meest gangbare vorm van informatie-uitwisseling tussen mensen. Qua opzet en meetpretentie gaat het hierbij om een eenvoudig instrument. Drie instrumentbeschrijvingen zijn in de categorie ondergebracht.

Een *programma* is gedefinieerd als de gecombineerde toepassing van meerdere typen assessment instrumenten. Dertien reacties zijn in deze categorie ondergebracht.

KENMERKEN

Doel en inhoud

Tabel 1 laat zien welke doelstellingen volgens het onderzoek worden beoogd met het gebruik van de verschillende typen instrumenten en welk inhoudelijk domein zij bestrijken.

Instrumenten voor fysiek functioneren worden vooral ingezet om de huidige situatie te beoordelen. Voor de overige instrumenten geldt dat zij tevens worden gebruikt voor de voorspelling van toekomstige arbeidsmogelijkheden.

De tabel laat ook een duidelijke samenhang zien tussen instrumenttype en het soort adviezen dat hier op wordt gebaseerd.

Ook ten aanzien van het inhoudelijk terrein verschillen de instrumenttypen van elkaar. De vragenlijsten en de instrumenten voor het fy-

Doel	Domein
Vragenlijst	- zowel huidige toestand als toekomstverwachting - scholings- en begeleidingsadviezen
Worksampl	- zowel huidige toestand als toekomstverwachting - adviezen t.b.v. concrete arbeidsplaats en training
Fysiek functioneren	- huidige toestand - adviezen t.b.v. concrete arbeidsplaats
Checklist	- zowel huidige toestand als toekomstverwachting - in algemeen geen advisering
Gestructureerd gesprek	- zowel huidige toestand als toekomstverwachting - arbeidsmogelijkheden - advisering op een breed terrein
Programma	- zowel huidige toestand als toekomstverwachting - trainings- en begeleidingsadviezen

tabel 1

siek functioneren beperken zich tot een specifiek domein; worksampl, checklist, (gestructureerd) gesprek en programma kunnen een breed inhoudelijk terrein bestrijken.

Doelpopulaties

Alle instrumenten worden gebruikt voor een brede populatie wat betreft leeftijd en werkervaring. De doelgroepen kunnen echter wel verschillen in opleidingsniveau: worksamples worden bijvoorbeeld in hoofdzaak toegepast bij kandidaten tot en met MBO-niveau. Een deel van de vragenlijsten, checklists en programma's wordt afgenomen bij kandidaten waarvan de opleidingsachtergrond varieert van LBO tot universitair onderwijs.

Aard, duur en kosten van afname

Tabel 2 geeft per instrument de aard van de afname aan. De vragenlijst, checklist en programma's hebben met elkaar gemeen dat de kandidaten getest worden met schriftelijke lijsten. Bij één checklist levert lichamenlijk onderzoek additionele meetgegevens op, bij programma's dragen observaties en gestructureerde gesprekken bij aan het meetresultaat. De afname van worksampl en fysiek functioneren vindt plaats in een testopstelling, observatie-gegevens vormen bij deze instrumenten een belangrijk onderdeel van de testresultaten.

Alle instrumenten worden individueel en vrijwel altijd onder begeleiding afgenomen.

Tabel 3 presenteert voor elk type instrumenten de duur en kosten van afname. Het vragenlijstonderzoek is, gelet op de afnameduur, relatief kostbaar. Dit hangt waarschijnlijk sa-

men met het hoge kwalificatieniveau van de gebruikers/beoordelaars van vragenlijsten. Blijkbaar zijn de (lange) afnameduur en de (hoge) aanschafkosten van een instrument minder van invloed op de prijsbepaling.

Beoordeling van de meetresultaten

Tabel 4 toont de wijze waarop de beoordeling en de interpretatie van de individuele meetresultaten tot stand komt en wie deze in de regel beoordeelt.

Normen voor de beoordeling van de meetresultaten vormen een integraal onderdeel van alle instrumenten voor fysiek functioneren en van de meeste vragenlijsten. Bij worksamples komen bestaande normen en zelf ontwikkelde normen ongeveer even vaak voor. Bij dit type instrument wordt daarnaast regelmatig beoordeeld naar eigen inzicht, dus zonder dat daarbij gebruik wordt gemaakt van standaardnormen.

Ook bij checklists worden scores naar eigen inzicht geïnterpreteerd, maar wordt tevens gebruik gemaakt van normen. Bij programma's wordt in het algemeen geen standaardnormering gehanteerd. Hier oordeelt de beoordelaar naar eigen inzicht, terwijl bij het (gestructureerde) gesprek de kandidaat zelf mede beoordeelt.

Vrijwel alle gerapporteerde normen zijn gebaseerd op gezonde normpopulaties.

Bij alle instrumenten worden de meetresultaten (ook) in combinatie met andere bevindingen geïnterpreteerd.

Ten slotte toont tabel 4 dat elk type assessment instrument voorbehouden lijkt te zijn voor een bepaalde discipline: arbeidskundigen passen worksamples toe, psychologen gebruiken vragenlijsten, ergo- en fysiotherapeuten beoordelen het fysieke functioneren, etc.

KWALITEIT

Wat valt nu te zeggen over de wetenschappelijke kwaliteit van de assessment instrumenten? Daartoe zijn de volgende aspecten geanalyseerd: theoretische achtergrond, standaardisering, normering en systematische registratie van de meetresultaten.

Op basis van de verstrekte gegevens over de theoretische achtergrond is de meerderheid van de beschreven instrumenten ingedeeld in één van de categorieën vragenlijst, worksampl of instrument ter beoordeling van fysiek functioneren. Deze drie typen zijn gedefinieerd als gestandaardiseerde en genormeerde instrumenten die (pretenderen) objectieve, betrouwbare en valide meetresultaten (te) geven. Ook programma's kunnen instrumenten bevatten die voldoen aan deze kwaliteitseisen. Deze resultaten geven aanleiding om te veronderstellen

Aard van afname

Vragenlijst	- schriftelijk
Worksample	- testopstelling in (experimentele) werkomgeving, observatiegegevens, in sommige gevallen ook (gestructureerd) gesprek
Fysiek functioneren	- gevarieerd: testopstelling niet arbeidsspecifiek en testopstelling in (experimentele) werkomgeving, observatie. In sommige gevallen ook interactief computerprogramma en lichamelijk onderzoek.
Checklist	- schriftelijke vragenlijst, lichamelijk onderzoek. In een enkel geval ook testopstelling in (experimentele) werkomgeving, observatie en gesprek
(Gestructureerd) gesprek	- gestructureerd gesprek. In een enkel geval ook observatie.
Programma	- gevarieerd: schriftelijke vragenlijsten, observaties, gestructureerde gesprekken, fysiologisch onderzoek, psychodiagnostisch onderzoek

tabel 2

dat ongeveer driekwart van de assessment instrumenten de toets der wetenschappelijk kritiek kunnen doorstaan.

Enige relativerende opmerkingen over de vermeende (hoge) kwaliteit van assessment instrumenten, toegepast in de reïntegratie praktijk, zijn echter op zijn plaats.

Een instrument is gestandaardiseerd wanneer instructie, afname en uitwerking voor

iedere kandidaat hetzelfde is. Niet alle instrumenten in de categorieën 'vragenlijst', 'worksample' en 'fysiek functioneren' blijken hieraan te voldoen. De instrumenten uit deze categorieën worden vrijwel allemaal voor verschillende doeleinden gebruikt. Zowel voor de beoordeling van de huidige als van de toekomstige situatie, maar ook komt het voor dat op basis van de testresultaten adviezen worden

gegeven die op meer dan één terrein betrekking hebben. De wetenschappelijke literatuur maakt doorgaans duidelijk onderscheid in het soort informatie dat assessment instrumenten (dienen te) leveren. Elk type informatie vraagt om een instrument dat op een daarbij passende manier geconstrueerd is. Wie bijvoorbeeld scholingsadviezen geeft dient op een andere manier relevante aspecten te onderzoeken dan wanneer men adviezen over de werkplek geeft.

ONGEWENST

Genormeerde instrumenten bevatten beoordelingsnormen gebaseerd op één of meer welomschreven populaties. Bij bijna de helft van de worksamples en een kwart van de vragenlijsten ontbreken richtlijnen en/of normen voor de beoordeling van de resultaten. Dit is een onverwacht resultaat, die het optimisme over de kwaliteit van de gebruikte instrumenten temperd. Bij de programma's wordt in de helft van de gevallen geen gebruik gemaakt van gedocumenteerde richtlijnen; bij deze instrumenten oordeelt de beoordelaar vooral naar eigen inzicht. Onverwacht is bovendien dat bij de programma's, op één uitzondering na, de beoordeling plaatsvindt door één persoon en niet door een team van meerdere deskundigen. Gezien het feit dat het hier om meerdere typen instrumenten gaat, lijkt ook dit vanuit het oogpunt van kwaliteit een ongewenste uitkomst.

Een ander punt is of de gebruikte instrumenten geschikt zijn voor de specifieke doelpopulatie waar het in dit onderzoek om gaat; de beroepsbevolking met gezondheidsproblemen. Zoals gezegd dienen genormeerde instrumenten beoordelingsnormen te bevatten die gebaseerd zijn op één of meer welomschreven populaties. Slechts een enkele keer blijken de beschreven instrumenten normen te bevatten die gebaseerd zijn op dezelfde populatie als de populatie waarbij ze in het kader van reïntegratie-advisering worden toegepast. Dit plaatst vraagtekens bij de toepasbaarheid van het instrumentarium.

Deze uitspraak verdient een nuancering. Bij een aantal instrumenten, bijvoorbeeld bij de worksamples, worden de testresultaten expliciet afgezet tegen de prestaties van gezonde reguliere werknemers. Op deze wijze wordt duidelijk hoe de capaciteiten van de geteste persoon zich verhouden tot een bepaald prestatiecriterium en welke maatregelen genomen moeten worden om een kandidaat te (re)integreren. Dit is een valide werkwijze.

Een laatste bedenking ten aanzien van de kwaliteit van het gebruikte instrumentarium is het lage aantal instrumenten waarbij een data-

Duur

Kosten

Vragenlijst	van minder dan een uur tot een hele dag	variërend van max. f 500 of f 1000 tot f 1500
Worksample	variërend van enkele dagen tot een week	variërend van max. f 500 of van f 500 tot f 1000
Fysiek functioneren	een hele dag	van f 1000 tot f 1500
Checklist	variërend van minder dan een uur tot enkele uren	f 150
Gestructureerd gesprek	variërend van minder dan een uur tot één dagdeel	ongeveer f 500
Programma	variërend van enkele uren tot langer dan een week	doorgaans f 1850 of meer

tabel 3

	Beoordeling individuele meetresultaten	Interpretatie meetresultaten	Beoordelaar
Vragenlijst	doorgaans op basis van normen/referenties gezonde populatie	zelfstandig en in combinatie met andere bevindingen	psychiater of psycholoog
Worksample	normen/referenties als onderdeel van instrument: zelfontwikkelde normen gezonde populatie: beoordeling naar inzicht beoordelaar	zelfstandig en in combinatie met andere bevindingen	arbeidskundige
Fysiek functioneren	uitsluitend op basis van normen/referenties gezonde populatie	zelfstandig en in combinatie met andere bevindingen	ergotherapeut of fysiotherapeut
Checklist	op basis van normpopulaties (eventueel zelf ontwikkeld)	meestal zelfstandig	arbeidskundige of revalidatie-arts
Gestructureerd gesprek	beoordelingen altijd mede op basis van eigen oordeel kandidaat	variërend: alleen zelfstandig, alleen in combinatie of zelfstandig én in combinatie	paramedicus of sociaal geschoolede
Programma	meestal geen standaardnormering: beoordeling op basis eigen inzicht beoordelaar en kandidaat	meestal in combinatie met andere bevindingen	psycholoog/ psychiater of arbeidskundige of arbeids-/ fysio-/ ergotherapeut

tabel 4

bestand van de meetresultaten wordt aangelegd. Bij vragenlijsten, worksamples en programma's wordt meestal geen databestand bijgehouden. Alleen bij de instrumenten die het fysiek functioneren bepalen wordt dit meestal wel gedaan. De voordelen van het systematisch registreren van de testresultaten zijn legio. Met de gegevens kunnen op termijn, voor de doelgroep passende, normen ontwikkeld worden, kan nagegaan worden of het instrument valide is, kan het belang van het instrument voor het uiteindelijke (re)integratie-advies vastgesteld worden en kunnen verschillen tussen groepen en trends in de tijd zichtbaar worden.

CONCLUSIE

Op het eerste gezicht lijkt ongeveer driekwart van de instrumenten die in het onderzoek zijn beschreven te voldoen aan hoge wetenschappelijke kwaliteitseisen. Enkele nuance-

rende opmerkingen zijn in dit verband echter op zijn plaats. Het onderzoek maakt duidelijk dat de meeste instrumenten worden ingezet voor veel uiteenlopende doelstellingen. Bovendien ontbreken bij een aantal instrumenten richtlijnen en/of normen voor de beoordeling van de meetresultaten. In veel gevallen worden de meetresultaten (onterecht) afgezet tegen een gezonde normpopulatie.

De betrouwbaarheid (de nauwkeurigheid) en de validiteit (de geldigheid) van assessments kunnen op verschillende manieren op een aanvaardbaar niveau worden gebracht. Voorbeelden van maatregelen zijn: het gebruik van goede instrumenten, de inzet van deskundig personeel, het protocolleren van het proces van oordeelsvorming, goede voorlichting aan kandidaten, heldere rapportages van de resultaten, etc. Al deze maatregelen beogen het verzamelen van gegevens te standaardiseren en de

beoordeling ervan te normeren. De toepassing van goede instrumenten is dus een noodzakelijke, maar nog lang geen voldoende voorwaarde voor een kwalitatief goede assessment.

De wetenschappelijke, professionele kwaliteit van assessments is niet het enige belangrijke thema binnen het onderzoek naar assessment. Vanuit het perspectief van de kandidaat, degene die beoordeeld wordt, en vanuit het perspectief van de opdrachtgever/financier zijn eveneens thema's te onderscheiden die wezenlijk zijn voor goede assessments. Voor kandidaten is bijvoorbeeld van groot belang dat zij actief participeren in het assessment programma, opdat zij zich kunnen herkennen in de resultaten. Immers, indien het advies niet aansluit bij de eigen wensen en opvattingen is de kans gering dat (re)integratie mogelijkheden gerealiseerd zullen worden.

Onder meer deze thema's worden uitgewerkt in het meerjarig onderzoeksprogramma 'Assessment en arbeidstoeleiding' van TNO. Doel van dit programma is het bevorderen van de kwaliteit van oordeelsvorming over arbeid en gezondheid van individuele werknemers.

Astrid Hazelzet en Annemiek Mul - NIA
TNO, Amsterdam.

Literatuur

Mul, C.A.M., Hazelzet, A.M., Wevers, C.W.J. Assessment bij (re)integratie deel I, deel II, deel III. Leiden, TNO Preventie en Gezondheid, publ.nr. 96.057. De rapporten kunnen tegen kostprijs worden opgevraagd bij de auteurs.