

De relatie tussen persoonlijkheid en effectiviteit van assessoren

Assessor personality and effectiveness

Marije J.E. Tummers-Poels

Eerste begeleider: dr. T. Houtmans

Tweede begeleider: dr. W. Bloemers

Begeleider Saxion: dr. W. Schoonman

Oktober 2007

Faculteit Psychologie

Afstudeerrichting Arbeids- en organisatiepsychologie

Open Universiteit Nederland

VOORWOORD

Deze scriptie dient als afstudeeronderzoek voor mijn studie Arbeids – en Organisationspsychologie aan de Open Universiteit Nederland. Voor dat ik met deze scriptie begon heb ik stage gelopen bij H&G Personeel en Organisatie. Hier heb ik ervaren dat assessment een veelomvattend en boeiend onderwerp is. Om een assessment te organiseren is veel deskundigheid nodig. Mijn stagebegeleiders drs. P. Gruijters en drs. H. van de Westelaken wil ik bedanken voor hun adviezen en aanmoedigingen om meer te weten te komen over assessment.

Het onderwerp van deze scriptie, een onderzoek naar de relatie tussen de persoonlijkheid en de effectiviteit van assessoren, is uitgevoerd op verzoek van de Kenniskring Assessment van Saxion Hogescholen. Graag wil ik drs. H. van der Stam bedanken voor zijn volledige medewerking gedurende het onderzoek en dr. W. Schoonman, lector Assessment, voor zijn hulp en begeleiding bij deze scriptie.

Vanuit de Open Universiteit ben ik bijgestaan door dr. T. Houtmans. Haar wil ik bedanken voor haar motiverende en verhelderende steun en hulp. Evenzeer wil ik dr. W. Bloemers bedanken voor zijn inbreng in deze scriptie.

Uiteraard ook hartelijk dank aan alle deelnemers aan dit onderzoek die hun resultaten beschikbaar hebben gesteld. Zonder hun inbreng zou dit onderzoek niet mogelijk zijn geweest.

INHOUDSOPGAVE

1. INLEIDING	6
1.1. Assessment in Het Hoger Beroepsonderwijs.....	6
1.2. Assessor bij gedragsproeven	8
1.3. De persoonlijkheid van de assessor	10
1.4. Persoonlijkheid en de wijze van beoordelen	12
1.5. Probleemstelling en hypothesen	15
2. METHODE	18
2.1. Onderzoeksgroep/respondenten	18
2.2. Meetinstrumenten.....	20
2.3. Procedure.....	24
3. RESULTATEN	24
3.1 Persoonlijkheidsvragenlijst	25
3.2 Gedragsoefeningen.....	26
3.3. Analyse hypothesen.....	27
3.3. Analyse onderzoeksvragen.....	28
4. DISCUSSIE	34
4.1. Conclusies	34
4.2. Beperkingen van dit onderzoek.....	36
4.3 Aanbevelingen.....	37
LITERATUUR	38
BIJLAGE 1: observatieformulier ‘Apies’	43
BIJLAGE 2: observatieformulier ‘Verlof 1’ en ‘Verlof 2’	44
BIJLAGE 3: Normscore experts	46
BIJLAGE 4: Coëfficiënten regressieanalyse	48

SAMENVATTING

In het competentiegericht onderwijs betekent assessment het multidimensionaal beoordelen van de geschiktheid van een persoon voor een maatschappelijke functie. Een instrument dat hiervoor gebruikt kan worden is de gedragsproef. Deze wordt beoordeeld door assessoren. De Kenniskring Assessment van Saxion Hogescholen (KAS) heeft een persoonlijkheidsprofiel voor assessoren opgesteld. In verband hiermee is de volgende onderzoeksvraag opgeworpen: heeft dit assessorprofiel voorspellende waarde voor de wijze van beoordelen door de assessor? Onderzocht is in hoeverre een bepaalde score op de dimensies van de NEO-FFI tot een bepaalde beoordeling cq. beoordelingstendentie leidt (Bartels en Doverspike, 1997; Bernardin, Cooke en Villanova, 2000).

Om de onderzoeksvraag te kunnen beantwoorden zijn van maart tot en met december 2006 gegevens verzameld van 225 aspirant assessoren die een persoonlijkheidsvragenlijst hebben ingevuld en vijftal assessmenttoefeningen hebben gemaakt. Uit het onderzoek naar de samenhang hiertussen blijkt dat voor twee van de vijf schalen van de NEO-FFI significante resultaten zijn gevonden die voldoen aan de verwachtingen van het assessorprofiel. Deze resultaten gelden voor de factor Openheid en voor de factor Consciëntieusheid, elk voor één assessmenttoefening. Op grond van één assessmenttoefening is het echter niet mogelijk stellige uitspraken te doen over de samenhang tussen de persoonlijkheid van een assessor en zijn manier van beoordelen. Bij het toetsen van drie exploratieve onderzoeksvragen blijkt dat persoonlijkheidsfactoren en demografische factoren geen significante bijdragen leveren aan de predictie van de assessmentscore. Bij het vergelijken van de assessmentscores bij verschillende groepen zijn ‘geslacht’ en ‘leeftijd’ positieve predictoren.. Het lijkt erop dat oudere assessoren beter beoordelen. De meest voor de hand liggende verklaring is dat ervaring een rol speelt bij de wijze van beoordelen. Aanbevolen wordt om verder onderzoek te doen met assessmenttoefeningen die overeenkomen met een werkelijke assessortaak, met meer assessmentindicatoren (bijvoorbeeld cijfers van tentamens, stagebeoordelingen) en met ervaren assessoren.

SUMMARY

In competence-based education, assessment refers to the multidimensional evaluation of the suitability of a person for a social task or function. One instrument which may be used is the behaviour test, which is evaluated by assessors. The Assessment Centre of Expertise (*Kenniskring Assessment*) of Saxion Universities of Professional Education has constructed a personality profile for assessors. This assessor profile gives indications about the personality characteristics of a good assessor, and has prompted the following research question: does this assessor profile predict the quality of the evaluation of behaviour tests by the assessor? This study focuses on the degree to which a particular score on the NEO-FFI dimensions leads to a particular evaluation or evaluation tendency (Bartels en Doverspike, 1997; Bernardin, Cooke en Villanova, 2000.) In order to answer the research questions, data were collected from 225 participants during twelve workshops/assessment training courses held between March and December 2006. The assessor trainees filled in a personality questionnaire and completed five assessment exercises. Correlation analysis showed significant results for two of the five NEO-FFI scales and assessment exercise performance. These results relate to the factors Openness and Conscientiousness, both for one assessment exercise. However, it is not possible to make definitive pronouncements about the correlation between the personality of an assessor and his manner of evaluation on the basis of a single assessment exercise. Furthermore, three research questions were explored. Personality factors and demographic factors did not significantly contribute to the prediction of the assessment score. With regard to the differences between different groups, sex and age were positive predictors for assessment evaluation outcome. It seems that the older the assessor is, the better his or her evaluation. An explanation could be that experience plays a part in assessment. It is advisable to carry out further research into assessment exercises that conform to a real assessor task and that have different assessment indicators (like examination marks) and experienced assessors.

De relatie tussen persoonlijkheid en effectiviteit van assessoren

1. INLEIDING

Het toepassen van assessment in het onderwijs neemt een hoge vlucht, waardoor de behoefte aan goed opgeleide assessoren (beoordelaars) steeds groter wordt. Beoordelen van studenten in het Hoger Beroepsonderwijs dient op een betrouwbare wijze te gebeuren, omdat dit consequenties heeft voor de student. Bij een professionele assessor wordt verondersteld dat hij in staat is valide oordelen uit te spreken over prestaties van anderen. Het blijkt dat mensen dit in het algemeen niet zo goed kunnen, hoewel zij denken dat dit wel het geval is (Hofstee, 1999; Schoonman, 2004). In dit onderzoek wordt de effectiviteit van de assessor op het beoordelen van gedragsproeven onderzocht. Dit onderzoek vindt plaats bij Saxion Hogescholen, een van de grootste hogescholen in Oost-Nederland met vestigingen in Deventer, Enschede en Apeldoorn en ongeveer 20.000 studenten.

1.1. Assessment in Het Hoger Beroepsonderwijs

In het Hoger Beroepsonderwijs hebben de laatste jaren grote veranderingen plaats gevonden. De onderwijsbehoefte van de individuele student staat steeds meer centraal (Wijnhoven, 2001). Binnen de docentfunctie ontstaan nieuwe taken zoals coach en assessor. Er wordt vanuit een veranderende visie op het onderwijs minder de nadruk gelegd op het overbrengen van kennis en vaardigheden en meer op het ontwikkelen van aantoonbare beroepscompetenties (Hendriks & Schoonman, 2006).

In het bedrijfsleven maakt men bij het beoordelen van kandidaten voor een nieuwe functie gebruik van de assessment center methode (Roe, 1983; Zonderland, 1995; Guidelines, 2000). Het belangrijkste kenmerk van een assessment center (AC) is de meervoudigheid van assessments: er dienen verschillende beoordelingstechnieken gehanteerd te worden zoals: tests, interviews, vragenlijsten, sociometrische middelen en simulaties. Welke tests en oefeningen er gebruikt

worden, hangt af van de doelstellingen van een AC (selectie, ontwikkeling) en vooral van de competenties waarover men informatie wenst te verzamelen (Guidelines, 2000).

Hogescholen ontwikkelen competentiegericht onderwijs waarbij competenties van studenten beoordeeld worden (Cuppen, 2000; Stichting HON, 2000). De Onderwijsraad (2002) concludeert dat er geen generieke definitie voor een competentie te geven is, maar dat competentie altijd te maken heeft met:

- het toepassen van wat geleerd is;
- de integratie van twee of meer van de componenten kennis, vaardigheid en attitude;
- persoonlijke eigenschappen en stabiele bekwaamheden.

In het onderwijs gebruikt men de term assessment voor het multidimensionaal beoordelen van de geschiktheid van een persoon voor een maatschappelijke taak of functie. Multidimensionaal wil hier zeggen dat er meerdere instrumenten worden gebruikt en dat er vanuit meerdere perspectieven beoordeeld wordt (Elshout-Mohr, Oostdam, Dietze, & Snoek, 2001). Straetmans (1998) onderscheidt drie klassen van assessmentmethodes:

1. Hands-on: assessmentmethodes waarbij kandidaten in een realistische of zelfs reële werksituatie en met gebruik van echte apparatuur, gereedschappen of instrumenten kenmerkende taken uitvoeren.

2. Hands-off: methodes waarbij taken worden voorgelegd waaruit moet blijken of kandidaten de cognitieve component van een competentie beheersen.

3. Simulaties: onder nagebootste werkzaamheden demonstreert de kandidaat zijn bekwaamheid.

Hierbij maakt men gebruik van al dan niet realistische apparatuur.

Een simulatie (gedragsproef) dient aan de functie gerelateerd te zijn. Bij de ontwikkeling van een simulatie dient er zorg voor te worden gedragen dat er voldoende gelegenheid geboden wordt om verscheidene competentie-relevante gedragingen uit te lokken (Guidelines, 2000).

Wanneer iemand in een gesimuleerde beroepssituatie het bedoelde gedrag vertoont, is het

waarschijnlijk dat hij dit gedrag ook in de functie zelf zal vertonen. Bij de beoordeling wordt als het ware een steekproef (sample) getrokken uit het gedragsrepertoire dat de persoon in de functie moet vertonen. Van de Sande (1999) geeft aan dat gedrag geobserveerd dient te worden aan de hand van vooraf bepaalde gedragscategorieën en achteraf geïnterpreteerd moet worden op basis van voor iedereen gelijke criteria. Bij Saxion Hogescholen is per 2004 een Lectoraat en een bijbehorende Kenniskring Assessment opgericht. Een van de taken van de kenniskring is het ontwikkelen en beschikbaar maken van assessment instrumenten, hier opgevat als ‘gedragsproeven’ (Schoonman, 2004).

1.2. Assessor bij gedragsproeven

Assessoren beoordelen gedrag. De kwaliteit van de assessor bepaalt voor een groot deel de uiteindelijke kwaliteit van een assessment center (Lievens, 2000). Een assessor is een (vak)specialist die zelf jaren ervaring heeft in het te beoordelen vakgebied, kandidaten kan beoordelen en een assessment-procedure goed kan uitvoeren (Schuler & ter Wee, 2001). Aan de hand van de literatuur (Competentieprofiel, 2002; Van Dam, Klaijnsen & Frietman, 2003) zijn vijf competenties voor assessoren benoemd:

- a) Omgang met kandidaten en medeassessoren ;
- b) Professionele beroepshouding;
- c) Hanteren van de assessmentprocedure;
- d) Hanteren van assessmenttechnieken;
- e) Persoonlijke kenmerken.

De laatste competentie kan problemen opleveren. De menselijke beoordelaar is kwetsbaar en feilbaar. Of het nu om producten of mensen gaat, steeds weer blijkt dat beoordelen erg moeilijk is. De beoordelaar is feilbaar, maar onmisbaar (Hofstee, 1999). Beoordelaars gebruiken een eigen taal om kandidaten te beschrijven (Lievens & De Fruyt, 2001). Bij het beoordelen van een

gedragsproef vormt de zwakste schakel de individuele assessor (Schoonman, 2005). Menselijke beoordelaars kunnen te maken hebben met een aantal storende variabelen:

- Halo- versus horn-effecten; dit houdt in dat een kandidaat die een aangename en plezierige indruk maakt (versus onaangenaam en onplezierig) ook op andere punten b.v. zijn initiatief in zijn werk, te hoog respectievelijk te laag wordt beoordeeld (Heider, 1958; De Groot, 1994). Beoordelaars letten op in het oog springende kenmerken in plaats van zorgvuldig verschillende werkprestaties te beoordelen (Borman, 1975).
- Het sequentie-effect. De objectiviteit van de beoordeling van een bepaalde kandidaat kan verminderd worden door de nawerking van beoordelingen van voorafgaande kandidaten (De Groot, 1994). Roossink (2006) noemt dit het volgorde-effect.
- Persoonlijke vergelijking: de objectiviteit kan verminderd worden doordat beoordelaars verschillende (persoonsgebonden) strategieën hanteren (De Groot, 1994). Heider (1958) merkt op dat mensen er impliciete persoonlijkheidstheorieën op na houden – dat wil zeggen dat hun oordeel is gebaseerd op een onderliggend concept of ‘theory’. De ‘theorie’ bestaat in wezen uit een heel netwerk van vooronderstellingen over persoonlijke eigenschappen en de daarbij behorende gedragingen en de stellige verwachting dat mensen zich volgens die veronderstellingen gedragen. Zo verwacht men niet dat de ‘aardige’ buurman een geweldsmisdrijf pleegt.
- Het contaminatie-effect. De objectiviteit kan verminderd worden doordat beoordelaars er bewust of onbewust verschillende agenda's tegelijkertijd op na houden (De Groot, 1994; Roossink, 2006). De beoordeling dient dan een meerledig doel, namelijk enerzijds het eigenlijke doel van de beoordeling (b.v. beoordeling van arbeidsprestaties van de kandidaat) en anderzijds een doel dat door de betreffende beoordelaar wordt gesteld (b.v. zijn kans op carrièremogelijkheden).

- Centrale tendentie. De neiging om voor het gemak in het midden te gaan zitten in plaats van extreme waarderingen te geven die om meer onderbouwing vragen (Van der Maesen de Sombreff, 1999).
- Mildheidstendens (leniency). Het blijkt moeilijk om anderen negatief te beoordelen. (Van der Maesen de Sombreff, 1999). Sommige beoordelaars zijn geneigd om consequent hoge scores te geven ('lenient raters') terwijl anderen juist strenger zijn in hun beoordelingen ('severe raters') (Lunz, Wright & Linacre, 1990).
- Signifisch effect (Sadler, 1983; De Groot, 1994; Roossink, 2006). Dit betekent dat er beoordelingsfouten ontstaan als gevolg van verschillen van opvatting over de taak. De criteria laten ruimte voor de beoordelaar waardoor subjectief wordt beoordeeld.

Het gebruik van meerdere assessoren en het 'combineren' van hun informatie en oordelen kan het voordeel hebben dat individuele vertekeningen opgeheven worden (Herriot, 2003).

Assessortrainingen hebben een positief effect op het beoordelen van gedrag (Lievens, 2000). Om assessoren te helpen hebben Cito, Saxion Hogescholen en Helicon in samenwerking met Celbe Consultancy en de certificatie-instelling Certiforce een internationaal erkend en gewaarborgd persoonscertificaat voor de functie van assessor ontwikkeld.

1.3. De persoonlijkheid van de assessor

De onafhankelijke variabele in dit onderzoek is de persoonlijkheid van de assessor. In de klassieke oudheid probeerden de Grieken en Romeinen greep te krijgen op de vele verschillen tussen mensen door de mensheid onder te brengen in typen of categorieën. De arts Hippocrates (vierde eeuw voor Christus) onderscheidde al vier typen personen naar overheersend lichaamssap (Kouwer, 1963).

Tegenwoordig bestaat er een rijke verscheidenheid aan persoonlijkheidstheorieën waaronder de trektheorieën (Costa & McCrae, 1985). De trektheorieën onderscheiden een wisselend aantal trekken. Een trek is een duurzaam kenmerk van een persoon dat als verklaring

kan dienen voor geobserveerde regelmatigheden en consistenties in gedrag (Reber, 1997). Trekken zijn consistent over tijd en ze zijn betrekkelijk onafhankelijk van situaties. De meest invloedrijke trektheorie is waarschijnlijk die van Cattell (1965), die zich baseert op een reeks “brontrekken” die in elk individu in bepaalde mate zouden voorkomen en die de “echte structurele invloeden achter de persoonlijkheid” zouden vormen. Cattell wilde een omvattende gedragstheorie maken. Gedrag in een bepaalde situatie is volgens hem een functie van drie dimensies: geschiktheden/vaardigheden, temperament en motieven. Het theoretische werk van Cattell resulteerde onder meer in de 16 persoonlijkheidsvragenlijst waarin 16 universele gedragsdeterminanten onderscheiden worden. De 16 PF van Cattell staat aan de basis van de “Big Five”, een trektheorie die op dit moment in de belangstelling staat. Bij de “Big Five” (Costa & McCrae, 1985) wordt uitgegaan van vijf elementaire trekdimensies ofwel vijf persoonlijkheidsfactoren. De oorsprong van de vijf persoonlijkheidsfactoren ligt in de “lexicale hypothese” (Goldberg, 1990; McCrae, 1990). Deze hypothese is gebaseerd op het idee dat individuele verschillen die van belang worden geacht in de onderlinge communicatie, hun weerslag zullen vinden in de dagelijkse taal. Zoekend naar een zo volledig mogelijke bron van persoonlijkheidsbeschrijvende termen (adjectieven) kiezen de meest aanhangers van de lexicale hypothese voor het woordenboek als bron voor het selecteren van deze termen. Vervolgens wordt een selectie van termen uit het woordenboek voorgelegd aan een groot aantal personen en wordt aan deze personen gevraagd zichzelf en/of anderen in alledaagse taal te beoordelen op deze persoonlijkheidsbeschrijvende termen (Slotboom & Elphick, 1999).

Tegenwoordig heerst in toenemende mate overeenstemming dat vier of vijf replicerbare factoren de verschillen tussen mensen met betrekking tot de meeste van die eigenschappen op adequate wijze samenvatten. Sterker nog, terwijl onderzoekers in het verleden deels verschillende labels hanteerden, begint het nu algemeen gebruik te worden naar de eerste vier factoren te verwijzen als Extraversion, Agreeableness, Conscientiousness, en Emotional Stability

(Nederlandse labels: Extraversie, Mildheid, Ordelijkheid, en Emotionele Stabiliteit). Ten aanzien van de vijfde factor zijn de discussies nog gaande, niet alleen met betrekking tot het beste label daarvoor, maar ook over de vraag of het hier wel een repliceerbare factor betreft. Costa en McCrae (1985) gebruiken voor de vijfde factor het label Openness to Experience. Het Nederlandse label voor de vijfde factor is volgens Hoekstra, Ormel en De Fruyt (1996) Openheid.

Het onderzoek naar de link tussen persoonlijkheid en functievervulling is sinds de introductie van het Five Factor Model sterk geïntensiveerd. Barrick en Mount (1991) hebben in een meta-analyse het verband onderzocht tussen de Big Five persoonlijkheidsdimensies en drie werkcriteria: werkbekwaamheid, trainingsbekwaamheid en personeelsgegevens (zoals de hoogte van het salaris). Voorspellingen van werkprestaties zijn gedaan voor professionals, politie-agenten, managers, verkopers en geschoolde en semi-geschoolde arbeiders. Het belangrijkste resultaat van het onderzoek van Barrick en Mount (1991) is dat Consciëntieusheid consistent de meest valide predictor is voor prestaties in alle beroepsgroepen en criteriumtypen. Tett, Jackson en Rothstein (1991) hebben in een meta-analyse de relatie onderzocht tussen persoonlijkheid en functievervulling. Het belangrijkste resultaat van deze meta-analyse is dat Altruïsme, Openheid en Neuroticisme meer valide predictoren zijn dan Consciëntieusheid. Deze resultaten verschillen met de resultaten van het onderzoek van Barrick en Mount (1991). Over de oorzaak van deze verschillen zijn verschillende artikelen geschreven (Ones & Mount, 1994; Tett & Jackson, 1994). Voor dit onderzoek is van belang dat beide onderzoeken aangeven dat er samenhang is tussen de Big Five persoonlijkheidsdimensies en functievervulling. De persoonlijkheid kan een gunstige of een storende variabele zijn bij de uitvoering van een functie.

1.4. Persoonlijkheid en de wijze van beoordelen

In de literatuur is onderzoek gedaan naar de link tussen persoonlijkheid en de wijze van beoordelen. Bartels en Doverspike (1997) onderzochten de relatie tussen de persoonlijkheid van

de assessor (aan de hand van de 16 PF vragenlijst van Cattell) en beoordelingen in assessment centers. Het resultaat van dit onderzoek is dat milde beoordelingen positief correleren met 'Gevoeligheid' ($r = .61, p < .05$) en 'Warmte' ($r = .50, p < .05$) en negatief correleren met 'Ongevoelige Houding' ($r = -.46, p < .05$). Deze drie factoren komen het meest overeen met 'Agreeableness' bij de Big Five vragenlijst. Bernardin, Cooke en Villanova (2000) bestudeerden de relatie tussen twee factoren van de Big Five en het geven van hoge beoordelingen door medestudenten. Ze waren hierin geïnteresseerd omdat ze bij eerder onderzoek hadden ontdekt dat een meerderheid van supervisors en administrateurs onnauwkeurig beoordeelden om confrontaties met ondergeschikten te vermijden (het leniency effect: te vriendelijk willen zijn). Uit hun onderzoek blijkt dat de relatie tussen 'Agreeableness' en het beoordelingsniveau positief en statistisch significant is ($r = .33, p < .01$). De correlatie tussen 'Conscientiousness' en de beoordelingen is negatief en statistisch significant ($r = -.37, p < .01$).

Dit onderzoek bouwt voort op de suggestie van Hofstee (1999) die suggereert dat de beste beoordelingen verkregen worden door het gemiddelde van meerdere beoordelaars te nemen. Om te bepalen of iemand een goede assessor is wordt in navolging van Hofstee in dit onderzoek bij één oefening gebruik gemaakt van de gemiddelde beoordeling. Hofstee beargumenteert verder dat door collectieve dwaling de gemiddelde beoordelaar het bij het verkeerde eind kan hebben. Ter voorkoming van deze 'collectieve dwaling' wordt in dit onderzoek gebruik gemaakt van 'expert-rating' (Hofstee, 1999; Van de Sande, 1999). In een meta-analyse (bij 50 studies) concluderen Gaugler, Rosenthal, Thornton en Bentson (1987) dat in assessment centers een hogere validiteit wordt gevonden wanneer psychologen assessor zijn dan wanneer managers assessor zijn ($r = -.21, p < .05$). Ook al wordt voor dit resultaat door Gaugler et al. (1987) geen verklaring gegeven, in dit onderzoek zullen een ervaren psycholoog, een psycholoog in opleiding en een ervaren cursusleider van assessortrainingen voor de oefeningen een expert score bepalen.

Voor dit onderzoek is van belang dat tussen enkele factoren van de Big Five (Altruïsme en Consciëntieusheid) en de wijze van beoordelingen samenhang is gevonden.

De leden van KAS in Enschede, Deventer en Apeldoorn hebben op basis van literatuur (Hoekstra et al., 1996; Bernardin et al., 2000) en eigen ervaring onderstaand assessorprofiel opgesteld.

	1	2	3	4	5	6	7	8	9
N									
E									
O									
A									
C									

	OK
	Aandacht
	Pas op!

Afbeelding © 2006 Kenniskring en Lectoraat Assessment Saxion Hogescholen

Neuroticisme hoog: mogelijk instabiel, emotioneel of stemmingsgevoelig

Extraversie hoog: mogelijk dominant, weinig zelfinzicht of agressief

Openheid laag: mogelijk te rigide, dogmatisch of nauw denkend

Openheid hoog: mogelijk te creatief, theoretisch of onrealistisch.

Altruïsme laag: mogelijk te egoïstisch, niet menslievend of self-centered.

Altruïsme hoog: mogelijk te aardig, opofferend of niet willen kwetsen.

Consciëntieusheid laag: mogelijk te slordig, niet exact of onmethodisch.

Dit assessorprofiel geeft een indicatie over de persoonlijkheid van een goede assessor en geeft zijn/haar kwetsbaarheid /gevoeligheid aan, vandaar dat wordt gesproken over gevoeligheidsanalyse. Aan de hand van deze gevoeligheidsanalyse kunnen (kandidaat)assessoren zich bewust worden van mogelijke risico's in hun functioneren. Dit persoonlijkheidsprofiel voor assessoren geeft indicaties over de kenmerken van een goede assessor, maar is nog niet empirisch

onderzoekt. In dit onderzoek wordt geprobeerd vast te stellen of het persoonlijkheidsprofiel van een goede assessor, opgesteld door KAS, een goede indicator is.

1.5. Probleemstelling en hypothesen

Het doel van dit onderzoek is inzicht te krijgen in de relatie tussen persoonlijkheid en de effectiviteit van assessoren. Dit is maatschappelijk relevant omdat het beoordelen van studenten in het onderwijs op een betrouwbare manier dient te gebeuren. Voor het onderwijs is de praktische relevantie dat men de juiste mensen als assessor inzet. De assessoren krijgen inzicht in hun zwakheden en valkuilen. Dit onderzoek is theoretisch relevant omdat het een aanvulling kan zijn in het onderzoek naar de samenhang tussen de vijf factoren van de NEO-FFI en het beoordelingsgedrag van mensen. In hoeverre leidt een bepaalde score op een dimensie van de NEO-FFI tot een bepaalde beoordeling cq. beoordelingstendentie.

De hoofdvraag van dit onderzoek luidt als volgt: zijn assessoren met een persoonlijkheidsprofiel dat overeenkomt met het assessorprofiel opgesteld door de Kenniskring Assessment betere assessoren dan assessoren met een afwijkend profiel? De onafhankelijke variabele in dit onderzoek is de persoonlijkheid van de assessor (beoordelaar), de afhankelijke variabele is de wijze van beoordelen; in dit geval observeren en registreren van waargenomen gedrag bij de oefeningen `Puzzel`, `Glimlachen`, `Apies kijken`, `Verlof 1` en `Verlof 2`. Vanuit de Big Five persoonlijkheidsdimensies en de gevoeligheidsanalyse, opgesteld door KAS, zijn de volgende hypothesen opgesteld:

1. Een goede assessor van gedragsproeven scoort laag op Neuroticisme.
2. Een goede assessor van gedragsproeven scoort laag op Extraversie.
3. Een goede assessor van gedragsproeven scoort gemiddeld op Openheid.
4. Een goede assessor van gedragsproeven scoort gemiddeld op Altruïsme.
5. Een goede assessor van gedragsproeven scoort hoog op Consciëntieusheid.

Ook zijn er drie exploratieve onderzoeksvragen geformuleerd:

6. In welke mate beïnvloeden de verschillende persoonlijkheidsfactoren de wijze van beoordelen.

7. In welke mate dragen deze persoonlijkheidsfactoren bij aan de wijze van beoordelen nadat is gecontroleerd is voor geslacht, leeftijd, opleiding, ervaring, training, bevolkingsgroep en intelligentie (gemeten aan de hand van de oefening puzzel).
8. In welke mate zijn er verschillen in assessment voor verschillende groepen, te weten mannen en vrouwen, mensen van verschillende leeftijden, mensen met een verschillend aantal jaren ervaring in assessment taken, training in assessment taken, hoogst genoten opleiding, bevolkingsgroep en intelligentie (gemeten aan de hand van oefening puzzel).

Onderbouwing hypothese factor Neuroticisme. Verondersteld wordt dat naarmate iemand lager scoort op Neuroticisme hij/zij geneigd zal zijn om kalm, zeker, rustig en rationeel te reageren (Costa & McCrae, 1985; John & Srivastava, 1999). Een lage score op Neuroticisme betekent voor een assessor dat hij/zij gewoonlijk vrij is van stress- en schuldgevoelens, minder onderhevig is aan stemmingswisselingen en verlangens (aandrang) tot het geven van te hoge of te lage waarderingen weerstaat.

Onderbouwing hypothese factor Extraversie. Volgens de definitie betekent een lage score op Extraversie dat iemand een serieuze rustige persoon is die schrijven of e-mail prefereert boven praten (Costa & McCrae, 1985; John & Srivastava, 1999). Een lage score op Extraversie betekent voor een assessor dat hij/zij meer ruimte biedt aan anderen en zich daardoor meer zal focussen op de kandidaat.

Onderbouwing hypothese factor Openheid. Volgens de definitie betekent een gemiddelde score op Openheid dat iemand gematigd en pragmatisch is, maar open staat voor verandering, voor nieuwe ervaringen en om nieuwe oplossingen voor een probleem te onderzoeken als daartoe aanleiding is (Costa & McCrae, 1985; John & Srivastava, 1999). Een gemiddelde score op Openheid betekent voor een assessor dat hij/zij zowel innovatie als efficiëntie waardeert en daardoor realistisch zal beoordelen.

Onderbouwing hypothese factor Altruïsme. Volgens de definitie betekent een gemiddelde score op Altruïsme dat iemand in staat is om te wisselen/schakelen tussen competitieve en samenwerkingsgerichte situaties (Costa & McCrae, 1985; John & Srivastava, 1999). Een

gemiddelde score op Altruïsme betekent voor een assessor dat hij/zij niet overdreven afhankelijk, noch onafhankelijk is en daardoor objectief beoordeelt.

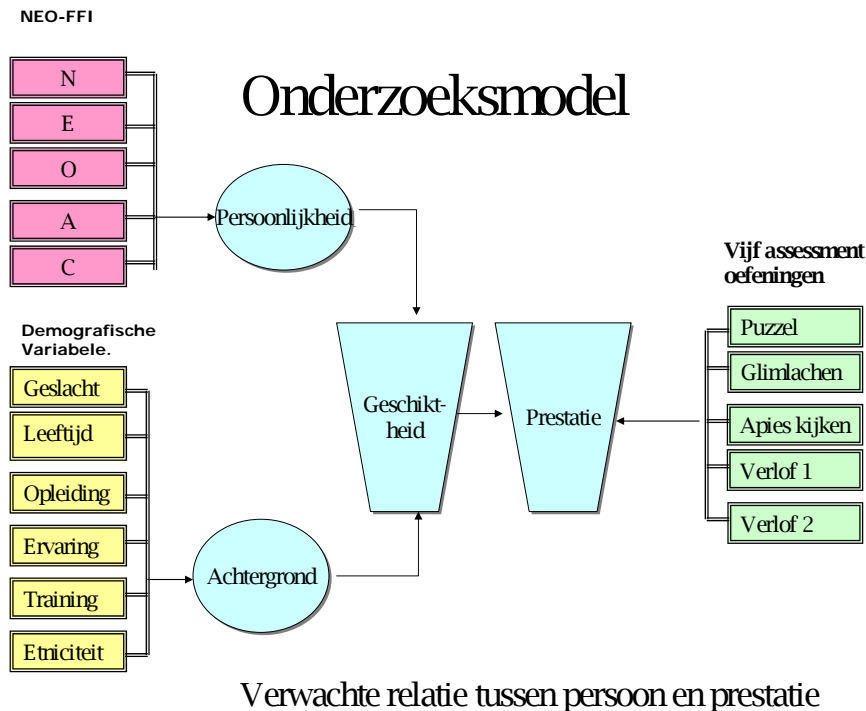
Onderbouwing hypothese factor Consciëntieusheid. Volgens de definitie betekent een hoge score op Consciëntieusheid dat iemand geneigd is zich op een vlijtige, gedisciplineerde en betrouwbare wijze op doelen te richten (Costa & McCrae, 1985; John & Srivastava, 1999). Een hoge score op Consciëntieusheid betekent voor een assessor dat hij/zij nauwgezet zal observeren en beoordelen.

Onderbouwing onderzoeksvraag persoonlijkheidskenmerken. Verondersteld wordt dat persoonlijkheidskenmerken bijdragen aan de wijze van beoordelen. Welke persoonlijkheidskenmerk het meeste bijdraagt is onduidelijk.

Onderbouwing onderzoeksvraag demografische variabelen. Verondersteld wordt dat de variabelen geslacht, leeftijd, opleiding, ervaring, training en bevolkingsgroep (demografische variabelen) van invloed kunnen zijn op de manier van beoordelen van de assessor.

Onderbouwing onderzoeksvraag verschillen in assessment voor verschillende groepen.

Verondersteld wordt dat verschillende groepen, te weten mannen en vrouwen, mensen van verschillende leeftijden, mensen met een verschillend aantal jaren ervaring in assessment taken, training in assessment taken, hoogst genoten opleiding en intelligentie (gemeten aan de hand van de oefening puzzel) verschillen in het beoordelen van de oefeningen.



2. METHODE

2.1. Onderzoeksgroep/respondenten

De dataverzameling heeft plaats gevonden van maart tot en met december 2006 tijdens twaalf assessment trainingen die verzorgd werden door leden van KAS. Elke workshop/training bestond uit 10-20 cursisten. Deelnemers waren medewerkers van onderwijsinstellingen zoals ROC Zwolle, Haagse Hogeschool, Universiteit Twente, Saxion Hogescholen, Hogeschool van Hall Larenstein, samenwerkingsnetwerk NL-opleidingen en medewerkers van diverse bedrijven zoals Defensie, Brandweer, Ingenieursvereniging en Celbe Consultancy. De duur van de trainingen liep sterk uiteen, van een halve dag tot enkele dagen (M 10; SD 4.67). In totaal hebben 225 respondenten meegewerkt aan het onderzoek. De respondenten hebben schriftelijk een aantal vragen beantwoord over de volgende demografische variabelen: geslacht, leeftijd, hoogste voltooide opleiding, jaren ervaring als assessor, afkomst (allochtoon dan wel autochtoon). Bovendien werd gevraagd of men een assessortraining gevolgd had.

De aspirant assessoren maken deel uit van een theoretische populatie, omdat de functie van assessor een nieuwe functie is. De veronderstelling is dat de samenstelling van de

toekomstige populatie met betrekking tot geslacht en leeftijd ongeveer gelijk is aan deze steekproef. De deelnemers van de trainingen zijn 141 (63%) mannen en 84 (37%) vrouwen. De gemiddelde leeftijd is 43.6 jaar (SD 9.6; range 20-66). De variabele 'leeftijd' is gehercodeerd in vier leeftijdsgroepen (zie tabel).

Tabel 1

Leeftijdsofbouw vrouwen en mannen (N=225)

Leeftijd	Vrouwen	Mannen	Totaal
20 t/m 30 jaar	6.0% (14)	7.0% (16)	13% (30)
31 t/m 40 jaar	8.5% (19)	9.5% (21)	18% (40)
41 t/m 50 jaar	16.5% (37)	28.5% (64)	45% (101)
51 jaar en ouder	6.0% (14)	28.0% (40)	24% (54)
Totaal	37.0% (84)	63.0% (141)	100% (225)

De opleiding van de respondenten is verdeeld in drie categorieën: WO 83 (37%) HBO 96 (43%) en anders 46 (20%). De respondenten die als opleiding post-HBO of HBO+ hebben opgegeven zijn gerangschikt onder de rubriek HBO.

Tabel 2

Opleiding vrouwen en mannen (N=225)

Opleiding	Vrouwen	Mannen	Totaal
WO	13% (30)	24% (53)	37% (83)
HBO	20% (44)	23% (52)	43% (96)
Anders	4% (10)	16% (36)	20% (46)
Totaal	37% (84)	63% (141)	100% (225)

Geen ervaring als assessor hebben 174 respondenten tegenover 51 respondenten die wel ervaring als assessor hebben. De 51 respondenten met ervaring hebben gemiddeld 9.7 jaar ervaring als assessor (SD 8.3; range 0.25 – 33). Het beoordelen van gedragsproeven door assessoren in het HBO is een recente ontwikkeling. Een respondent die aangeeft 33 jaar ervaring als assessor te hebben bedoelt hier waarschijnlijk mee dat hij 33 jaar ervaring heeft met het beoordelen van studenten. Deze vraag had duidelijker gesteld moeten worden namelijk: de vraag naar ervaring als assessor van gedragsproeven. Ervan uitgaande dat in het HBO sinds een jaar of vijf gewerkt wordt met assessoren is een opgave van vijf jaar ervaring reëel. Een ervaringstijd boven vijf jaar is in dit verband niet reëel en wordt niet in de analyse meegenomen. Het aantal respondenten zonder ervaring is 89% (174) en met 5 jaar of minder ervaring is 11 % (22) (SD 1.09; range 0-5). Van de 225 respondenten hebben 206 respondenten (92%) geen training gevolgd en 19 respondenten (8%) hebben wel een training gevolgd. Bij de vraag naar etniciteit is door 6 respondenten (3%) geantwoord met allochtoon, door 173 respondenten met autochtoon (77%) en 46 respondenten (20%) hebben deze vraag niet beantwoord. De variabele ‘bevolkingsgroep’ wordt niet in de verdere analyse betrokken omdat de verdeling tussen allochtoon en autochtoon te ongelijk is. Bovendien hebben te veel respondenten de vraag niet beantwoord. Waarschijnlijk is de vraag naar etniciteit een gevoelig onderwerp.

2.2. Meetinstrumenten

Persoonlijkheidsvragenlijst

De kandidaten beantwoordden de vragen van de NEO Five Factor Inventory (NEO FFI) (1996). Dit is de geautoriseerde Nederlandstalige versie van de NEO-FFI (Five Factor Inventory) van Costa en McCrae (1985), door Hoekstra et al. (1996). De NEO FFI Big Five meet de vijf belangrijkste dimensies of domeinen in de persoonlijkheid van volwassenen en bestaat uit 60 vragen. De testitems bestaan uit stellingen in de ik-vorm (voorbeeld: ‘Ik ben geen tobber’),

waarop gereageerd dient te worden aan de hand van een vijfpuntschaal, die van ‘helemaal oneens’ via ‘neutraal’ naar ‘helemaal eens’ loopt. De NEO-FFI wordt individueel met pen-en-paperversie afgenomen. In dit onderzoek wordt gebruik gemaakt van de ruwe scores. Door gebruik te maken van de ruwe scores is de variantie groter omdat er een groter bereik is (range 12 - 60). Als een factor ontbrekende waarden heeft (een item is niet ingevuld) wordt de ontbrekende waarde vervangen door het gemiddelde van alle waarnemingen voor die factor. In onderstaande tabel staan de alfa coëfficiënten van de NEO-FFI van Hoekstra et al. (1996) en van Costa en McCrae (1985) en van dit onderzoek .

Tabel 3

Cronbach's Alfa van de subschalen van de NEO-FFI.

	Hoekstra et al. (<i>N</i> = 2736)	Costa & McCrae (<i>N</i> = 1539)	Tummers (<i>N</i> = 190)
Neuroticisme	.84	.68	.82
Extraversie	.77	.77	.75
Openheid	.69	.73	.73
Altruïsme	.68	.81	.66
Consciëntieusheid	.77	.86	.75

Gedragsoefeningen

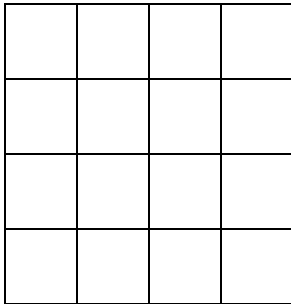
De kandidaten maken vijf oefeningen (Puzzel, Glimlachen, Apies kijken, Verlof 1 en 2).

De in dit onderzoek uitgevoerde oefeningen zijn ontwikkeld door KAS en worden gebruikt bij het trainen van assessoren. De validiteit en betrouwbaarheid van de oefeningen is niet onderzocht. Wel is van de gedragsproef “Apies kijken” bij een klein experiment (Hendriks & Schoonman, 2006) de beoordelaarovereenstemming berekend (de gemiddelde correlatie van niet-

getrainde beoordelaars en getrainde beoordelaars met de ware score). Voor de ongetrainde groep bedraagt deze .37 en voor de getrainde groep .47. De oefeningen zijn opgenomen op video zodat steeds hetzelfde gedrag wordt geobserveerd.

Puzzel: hoeveel vierkanten ziet men. Doel: observeren (kijken naar wat men ziet).

De eerste oefening ‘Puzzel’ is een observatietaak waarbij maar één oplossing de juiste is.



Antwoordmogelijkheden: 16, 17, 21, 26, 30, 36, 64. Het juiste antwoord is 30 (objectieve norm).

De score van elke respondent wordt vergeleken met de normscore. Voor elke respondent is de absolute score berekend $|30 - \text{score respondent}|$. Hoe kleiner de score hoe minder de afwijking is van het juiste antwoord (normscore), hoe beter de respondent scoort. Deze oefening observeert geen gedrag van mensen, maar geeft een indicatie van het ruimtelijk inzicht (een vorm van intelligentie) van de assessoren. Intelligentie speelt een rol bij de wijze van beoordelen. In alle functies vanaf een bepaald niveau is intelligentie een sterke predictor voor succes (Schmidt, Hunter & Outerbridge, 1986; Ree & Carretta, 1998). Deze oefening wordt gebruikt om te onderzoeken of een hogere score op intelligentie (in dit geval iemand die goed scoort op deze oefening) een betere score heeft bij het beoordelen van gedragsproeven.

Glimlachen: observeren of iemand glimlacht.

Er zijn tien foto's met gezichten van verschillende personen. De personen vertonen expressies die op glimlachen lijken (A tot en met J). Er wordt gewerkt met een observatieformulier.

Kandidaten moeten aangeven of een persoon glimlacht (ja = 1) of niet (nee = 0). Als normscore bij deze oefening is de modus bepaald van de scores van alle respondenten op elke foto. De score van elke respondent wordt vergeleken met de normscore. Voor elke respondent wordt per item de

absolute score berekend, dat wil zeggen dat de score van de respondent afgetrokken wordt van de normscore waarna de scores bij elkaar worden opgeteld. Hierdoor ontstaat een totale absolute score per respondent; hoe hoger de score, hoe minder de respondent overeenkomt met de normscore. Bij deze oefening is van te voren geen definiëring van het begrip ‘glimlachen’ gegeven. Hierdoor is een expertscore niet mogelijk.

Apies kijken: observeren welke gedragingen hoe vaak plaatsvinden.

Voor de oefening ‘Apies kijken’ wordt gebruik gemaakt van een categorieënsysteem met verschillende gedragsindicatoren en er wordt non-verbaal gedrag geobserveerd. Bij deze oefening wordt gebruik gemaakt van een fragment van drie minuten uit de film van Bert Haanstra ‘*Chimps onder elkaar*’ (1984). Hoofdrolspeler is de aap Wouter. Het gaat om zijn gedrag. Het geobserveerde gedrag is non-verbaal en heeft vooral betrekking op grote lichaamsbewegingen. Dit geobserveerde gedrag wordt opgedeeld in gedragsindicatoren. Een gedragsindicator heeft een rating-karakter; bijvoorbeeld de gedragsindicator “verplaatst boomstronk horizontaal gedurende tien seconden” moet worden geteld. Er wordt gewerkt met een observatieformulier (zie bijlage 1). Kandidaten turven hoe vaak ze het gedrag hebben gezien. Niets ingevuld wil zeggen dat het gedrag niet gezien is en levert score 0. De totaalscore per item wordt bepaald door het aantal keren dat gedrag gezien is. Als normscore voor deze oefening wordt uitgegaan van het gemiddelde van de scores van de experts (N=3) op elk item.

Verlof 1 en Verlof 2: observeren hoe vaak gedrag plaatsvindt en de kwaliteit van het vertoonde gedrag evalueren. Voor de oefeningen ‘Verlof 1’ en ‘Verlof 2’ wordt gebruik gemaakt van een categorieënsysteem met verschillende gedragsindicatoren. Er wordt non-verbaal en verbaal gedrag geobserveerd. Het betreft twee verschillende video-opnamen. Bij beide films wordt gebruik gemaakt van dezelfde vier categorieën en gedragsindicatoren. Er wordt gewerkt met een observatieformulier (zie bijlage 2). Behalve het observeren wordt in deze oefeningen ook geëvalueerd (waarde verbinden aan het geobserveerde gedrag). In ‘Verlof 1’ is het gesprek

zo gemaakt dat het bedoelde gedrag niet of nauwelijks voorkomt. In ‘Verlof 2’ is het bedoelde gedrag wel zichtbaar. Het zichtbare geobserveerde gedrag wordt van een waardeoordeel voorzien. Eerst wordt het gedrag geturfd met gezien (1) of niet gezien (0). Als een gedragsindicator niet is ingevuld wordt er vanuit gegaan dat dit gedrag niet is gezien (0). Als normscore bij deze oefeningen wordt de modus bepaald van de scores van de experts (N=3) op elk item. De kwaliteit van het gedrag wordt aangegeven met + (positief); 0 (neutraal = niet positief of negatief); - (negatief). De totaalscore van “kwaliteit van gedrag” wordt bepaald door de + score minus de – score. Als normscore bij deze oefening wordt de modus bepaald van de scores van de experts op elk item.

2.3. Procedure

Aan het begin van de training vullen de deelnemers anoniem de persoonlijkheidsvragenlijst in en maken de oefeningen. Aan de deelnemers wordt gevraagd of ze er bezwaar tegen hebben om deel te nemen aan dit onderzoek. Als ze geen bezwaar hebben tegen deelname aan dit onderzoek kunnen ze de formulieren (resultaat persoonlijkheidsvragenlijst + observatieformulieren) afgeven aan de cursusleider. De resultaten zijn handmatig ingevoerd in het computerprogramma SPSS.

3. RESULTATEN

Voor twee van de vijf schalen van de NEO/FFI (Openheid en Consciëntieusheid) zijn significante resultaten gevonden die overeenstemmen met de verwachtingen van het door KAS opgestelde assessorprofiel, elk voor één assessmentoefening. Persoonlijkheidsfactoren en demografische factoren leveren geen significante bijdragen aan de predictie van de assessmentscore. Bij het vergelijken van de assessmentscores bij verschillende leeftijdsgroepen blijkt dat de factor leeftijd (ervaring) een rol speelt bij de wijze van beoordelen.

3.1 Persoonlijkheidsvragenlijst

Tabel 4

Statistische kengetallen (N = 190) en Cronbach's alfa van de subschalen van de NEO-FFI.

	Items	Min.score	Max.score	<i>M</i>	<i>SD</i>	α
Neuroticisme	12	14	49	26.90	6.36	.82
Extraversie	12	25	56	43.68	5.60	.75
Openheid	12	26	57	41.35	6.17	.73
Altruïsme	12	27	57	44.73	4.76	.66
Consciëntieusheid	12	24	58	46.11	5.20	.75

Om vast te stellen of er samenhang is tussen de factoren, werd een correlatieanalyse uitgevoerd.

Tabel 5

Correlaties tussen de vijf verschillende factoren van de NEO-FFI (N=223)

Subschaal	2	3	4	5
1. Neuroticisme	-.38**	.09	.02	-.17**
2. Extraversie	-	.10	.25**	.14*
3. Openheid		-	.20**	-.21**
4. Altruïsme			-	.05
5. Consciëntieusheid				-

* $p < .05$. ** $p < .01$.

De vijf schalen van de Big Five worden geacht onafhankelijk van elkaar te zijn (Costa en McCrae, 1985). In tegenstelling met deze assumptie correleren in dit onderzoek verschillende factoren met elkaar. De oorzaak is waarschijnlijk de beperkte, niet representatieve steekproef.

3.2 Gedragsoefeningen

Tabel 6

Statistische kengetallen van de scores van de respondenten op 7 oefeningen en Cronbach's alfa (expertscore N=3)*

	Items	Min. score	Max. score	<i>M</i>	<i>SD</i>	<i>N</i>	<i>a</i>
1. Puzzel	1	16	64	8.59	8.20	221	-
2. Glimlachen	10	1	10	2.78	1.55	224	.61
3. Apies kijken	4	1	33	6.75	3.80	209	.70
4. Verlof 1	11	0	11	2.75	1.36	107	.52
4a. Kwaliteit Verlof 1	11	-4	5	2.14	1.76	107	-
5. Verlof 2	11	5	11	2.37	1.25	118	.16
5a. Kwaliteit Verlof 2	11	-4	9	4.69	1.54	118	-

* Zie voor de expertscore op de oefeningen Apies, Verlof 1 en Verlof 2 bijlage 3

De gegevens zijn verzameld tijdens verschillende workshops waarbij gebruik is gemaakt van verschillende oefeningen. Hierdoor is het aantal deelnemers per oefening verschillend.

Tabel 7

Correlatie tussen scores respondenten op verschillende oefeningen

	2	3	4	4a	5	5a
1. Puzzel	.00	.07	-.02	-.06	.00	.17
2. Glimlachen	-	.17*	-.20*	-.16	-.03	.06
3. Apies kijken		-	.03	.02	.16	-.08
4. Verlof 1			-	.52**	.11	-.08
4a. Kwaliteit verlof 1				-	-.10	.00
5. Verlof 2					-	-.07
5a. Kwaliteit verlof 2						-

* $p < .05$. ** $p < .01$.

Om te kijken of er een samenhang is tussen de scores van de respondenten op de verschillende oefeningen is een correlatieanalyse uitgevoerd, waarbij als normscore de expertscore geldt. Er zijn drie correlaties significant. De oefening 'Glimlachen' hangt positief samen met de oefening 'Apies kijken'. 'Glimlachen' en 'Verlof 1' hangen negatief samen en de samenhang tussen 'Verlof 1' en 'Kwaliteit verlof 1' is positief.

3.3. Analyse hypothesen

Bij het analyseren van de gegevens zijn allereerst de persoonlijkheidsfactoren gehercodeerd in drie rubrieken (laag – gemiddeld - hoog) en is bij de persoonlijkheidsfactoren Openheid en Altruïsme de gemiddelde score gehercodeerd naar de hoogste score. Voor de verdeling van de scores op de persoonlijkheidsfactoren wordt uitgegaan van een normaalverdeling. Een gemiddelde score is gedefinieerd als de scores tussen één standaarddeviatie onder het gemiddelde en één standaarddeviatie boven het gemiddelde. De hoge scores zijn de scores vanaf $M+1Sd$; de lage scores zijn de scores vanaf $M-1Sd$. Om de scores op de oefeningen met elkaar te kunnen vergelijken en een totaalscore te berekenen voor alle oefeningen samen zijn de absolute scores op de oefeningen getransformeerd naar standaardscores: de standaardscore is gelijk aan de totaalscore van een respondent op een oefening min de gemiddelde score van alle respondenten op die oefening gedeeld door de standaardafwijking van de score op die oefening. Om de samenhang tussen enerzijds de scores op de vijf factoren van de NEO-FFI en anderzijds de scores op elke oefening apart en de totaalscore van de respondenten op alle oefeningen samen te berekenen is correlatieanalyse uitgevoerd.

Tabel 8

Correlatieanalyse tussen scores op de persoonlijkheidsfactoren en de getransformeerde scores op de oefeningen

NEO-FFI	Oefeningen							
	Puzzel	Glimlachen	Apies	Verlof 1	Kwal 1	Verlof 2	Kwal 2	Totaal
1. Neuroticisme	.03	-.04	-.13	-.05	.09	-.25**	.16	-.11
2. Extraversie	-.11	-.09	-.01	.03	-.02	.03	-.04	-.13
3. Openheid	-.01	-.01	.04	-.25*	-.07	.06	.07	.00
4. Altruïsme	.09	.02	-.01	-.17	-.11	.02	.13	-.04
5. Consciëntieusheid	-.19**	-.02	.12	.25*	.07	.05	-.13	.16

* $p < .05$. ** $p < .01$.

Hypothese 1 wordt niet bevestigd. Er is wel een negatieve correlatie tussen Neuroticisme en de oefening ‘Verlof 2’.

Hypothese 2 wordt niet bevestigd. Er blijkt geen significante correlatie te zijn tussen Extraversie en de oefeningen.

Hypothese 3 wordt bevestigd. Er is een negatieve samenhang tussen Openheid en de oefening ‘Verlof 1’.

Hypothese 4 wordt niet bevestigd. Er is geen significante correlatie tussen Altruïsme en de oefeningen.

Hypothese 5 wordt bevestigd, maar de uitkomst is ambivalent. Er is een negatieve samenhang tussen Consciëntieusheid en de oefening ‘Puzzel’, echter tussen Consciëntieusheid en de oefening ‘Verlof 1’ is een positieve samenhang.

3.3. Analyse onderzoeksvragen

Onderzoeksvraag 6 veronderstelt dat persoonlijkheidskenmerken bijdragen aan de wijze van beoordelen. Om dit verband te toetsen werd meervoudige regressieanalyse uitgevoerd.

De totaalscore (standaardscore) op de oefeningen is de afhankelijke variabele. Er wordt onderzocht welke persoonlijkheidsfactoren bijdragen aan de verklaring van de afhankelijke variabele. De variantie R^2 is de variantie die alle vijf factoren samen verklaren.

Tabel 9

Lineaire regressieanalyse met als predictoren persoonlijkheidsfactoren (n = 105)

Model	R	R ²	Adjusted R ²	Std. Error of the Estimate
1	.25	.06	.02	2.99894

De persoonlijkheidsfactoren verklaren 6 % van de variantie van de scores op de oefeningen.

Het percentage verklaarde variantie is niet significant: $F(5,99) = 1.34$; $p = .25$.

Uit de regressieanalyse met als predictoren de persoonlijkheidsfactoren voor de afhankelijke variabele blijkt dat de persoonlijkheidsfactoren geen significante invloed hebben op de totale assessmentscore (zie bijlage 4, tabel 1).

Onderzoeksvraag 7 veronderstelt dat behalve de persoonlijkheidsfactoren ook de demografische variabelen van invloed kunnen zijn op de wijze van beoordelen. De totaalscore (standaardscore) op de oefeningen is de afhankelijke variabele.

Tabel 10

Lineaire regressieanalyse met als predictoren demografische factoren en persoonlijkheidsfactoren (n = 82)

Model	R	R ²	Adjusted R ²	Std. Error of the Estimate
1	.20(a)	.04	-.02	3.02840
2	.29(b)	.08	-.05	3.06507

a = training als assessor, geslacht, ervaring (max. 5 jaar), opleiding en leeftijd

b = training als assessor, geslacht, ervaring (max. 5 jaar), opleiding, leeftijd, Neuroticisme, Extraversie, Openheid, Altruïsme en Consciëntieusheid

Om dit verband te toetsen wordt meervoudige regressieanalyse uitgevoerd in twee stappen. De demografische factoren verklaren 4 % van de variantie van de scores op de oefeningen (model 1). Toevoeging van de persoonlijkheidsfactoren leidt tot een grotere verklaarde variantie. Nu wordt 8 % van de variantie van de scores op de oefeningen verklaard (model 2). De percentages verklaarde variantie zijn niet significant. Model 1: $F(5,76) = .65$; $p = .66$; model 2: $F(10,71) = .64$; $p = .78$.

Uit de regressieanalyse met als predictoren de demografische factoren en de persoonlijkheidsfactoren voor de assessmentscore blijkt dat de demografische factoren niet significant bijdragen aan de wijze van beoordelen (zie bijlage 4, tabel 2).

De oefening puzzel kan gezien worden als een maat voor intelligentie. Intelligentie kan van invloed zijn op de wijze van beoordelen. De regressieanalyse wordt opnieuw uitgevoerd, maar nu met de oefening puzzel toegevoegd aan de demografische variabele. De totaalscore (standaardscore) op de oefeningen (Apies kijken, Glimlachen, Verlof 1 + Kwaliteit verlof 1, Verlof 2 + Kwaliteit Verlof 2) is de afhankelijke variabele

Tabel 11

Lineaire regressieanalyse met als predictoren demografische factoren en persoonlijkheidsfactoren (n=82)

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. Error of the Estimate
1	.24(a)	.06	-.02	2.42882
2	.31(b)	.10	-.05	2.45831

a = puzzel, training als assessor, ervaring (max. 5jaar), geslacht, leeftijd en opleiding

b = puzzel, training als assessor, ervaring (max. 5jaar), geslacht, leeftijd, opleiding, Neuroticisme, Extraversie, Openheid, Altruïsme en Consciëntieusheid

De demografische factoren verklaren 6 % van de variantie van de scores op de oefeningen (model 1). Toevoeging van de persoonlijkheidsfactoren leidt tot een grotere verklaarde variantie.

Nu wordt 10 % van de variantie van de scores op de oefeningen verklaard (model 2). De percentages verklaarde variantie zijn ook hier niet significant.

Model 1: $F(6,74) = .74$; $p = .62$; model 2: $F(11,69) = .69$; $p = .75$.

Toevoeging van de oefening ‘puzzel’ bij de demografische factoren heeft geen significante invloed op de wijze van beoordelen (zie bijlage 4, tabel 3).

Onderzoeksvraag 8 veronderstelt dat er verschillen zijn in assessment voor verschillende groepen, te weten mannen en vrouwen, mensen van verschillende leeftijden, mensen met een verschillend aantal jaren ervaring in assessment taken, training in assessment taken, hoogst genoten opleiding en intelligentie (gemeten aan de hand van de oefening puzzel).

Geslacht. Door middel van de t-toets wordt gekeken of de gemiddelde scores van mannen en vrouwen gelijk zijn. De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standaard score op alle oefeningen gebruikt. Van de respondenten hebben 60 (57%) mannen en 46 (43%) vrouwen alle oefeningen gemaakt. De gemiddelde score van de mannen is -1.13 , van de vrouwen is dat -1.20 . Het resultaat van de t-toets: $t(104) = 1.83$; $p = .07$. Bij het hanteren van een significantieniveau van 5 % blijkt dat mannen niet significant beter scoren dan vrouwen, er is wel een trend in die richting.

Leeftijd. Door middel van variantie-analyse wordt gekeken of er verschillen zijn in scores voor de verschillende leeftijdsgroepen. De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standardscore op alle oefeningen gebruikt.

De gemiddelde score voor de leeftijdscategorie 20 tot en met 30 jaar ($n = 14$) is -2.15 , voor de leeftijdscategorie 31 tot en met 40 jaar ($n = 19$) is 1.38 , voor de leeftijdscategorie 41 tot en met 50 jaar ($n = 41$) is -1.26 en voor de leeftijdscategorie 51 jaar en ouder ($n = 32$) is $-.23$. In de laagste leeftijdscategorie is het verschil met de absolute score het grootst; in de hoogste

leeftijdscategorie het laagst. Er blijkt een significant verschil tussen de groepen te zijn $F(3,102) = 5.32$; $p = .002$. Er zijn minimaal twee groepsgemiddelden ongelijk aan elkaar. Door middel van een LSD toets wordt gekeken welke groepen van elkaar verschillen

Tabel 12

LSD-toets: meervoudige vergelijking variabele leeftijd (n=106)

	2.	3.	4.
1. 20 t/m 30 jaar	-3.52*	-.89	-1.92*
2. 31 t/m 40 jaar	-	2.63*	1.60
3. 41 t/m 50 jaar		-	-1.03
4. 51 jaar en ouder			-

* $p < .05$

Deze analyse toont aan dat er een significant verschil in scores is tussen leeftijdscategorie 20 tot en met 30 jaar en leeftijdscategorie 31 en 40 jaar en tussen leeftijdscategorie 20 tot en met 30 jaar met leeftijdscategorie 51 jaar en ouder. Ook is er een significant verschil in scores tussen leeftijdscategorie 31 en 40 jaar en tussen 41 en 50 jaar.

Ervaring met assessment taken. Door middel van de t-toets wordt gekeken of de gemiddelde score van respondenten met ervaring (max. 5 jaar) gelijk is aan de gemiddelde score van respondenten zonder ervaring. De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standaardscore op alle oefeningen gebruikt. Van de respondenten die alle oefeningen hebben gemaakt hebben 72 personen (87%) geen ervaring en 11 personen (13%) wel ervaring. De gemiddelde score van de personen zonder ervaring is $-.59$ en van de personen met ervaring is dat -1.55 . Het resultaat van de t-toets: $t(81) = .99$; $p = .33$. Bij het hanteren van een significantieniveau van 5% blijken de groepsgemiddelden niet van elkaar te verschillen.

Training. Door middel van de t-toets wordt gekeken of de gemiddelde score van respondenten die een assessor training hebben gevolgd en respondenten die geen assessor training hebben gevolgd gelijk zijn. De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standaardscore op alle oefeningen gebruikt.

Van de respondenten die alle oefeningen hebben gemaakt hebben 101 personen (95%) geen training gevolgd en 5 personen (5%) wel een training. De gemiddelde score van de personen zonder ervaring is $-.53$ en van de personen met ervaring is dat -1.89 .

Het resultaat van de t-toets: $t(104) = .99; p = .33$. Bij het hanteren van een significantieniveau van 5% blijken de groepsgemiddelden niet van elkaar te verschillen.

Hoogst genoten opleiding. Door middel van variantie-analyse wordt gekeken of er verschillen zijn in scores voor de verschillende opleidingsgroepen (WO – HBO - anders). De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standaardscore op alle oefeningen gebruikt. De gemiddelde score voor de categorie WO ($n = 52$) is $-.72$, voor de categorie HBO ($n = 46$) is dat $-.43$ en voor de categorie anders ($n = 8$) is dat $-.70$.

Het resultaat van de variantie-analyse: $F(2,103) = .12; p = .89$. Bij het hanteren van een significantieniveau van 5% blijken de groepsgemiddelden niet van elkaar te verschillen.

Puzzel (intelligentie). Door middel van de t-toets wordt gekeken of er verschillen zijn in de scores voor respondenten die de oefening puzzel goed hebben beantwoord en respondenten die de oefening puzzel fout hebben beantwoord. De afhankelijke variabele is de wijze van beoordelen. Hiervoor wordt de totale standaardscore op alle oefeningen (zonder de oefening puzzel) gebruikt. Van de respondenten die deze oefeningen hebben gemaakt hebben 16 personen (15%) deze oefening goed beantwoord en 89 (85%) personen fout beantwoord. De gemiddelde score van de personen met een goede score is -1.03 van de personen met een foute score is dat $-.29$. Het resultaat van de t-toets: $t(103) = .09; p = .28$. Bij het hanteren van een significantieniveau van 5% blijken de groepsgemiddelden niet van elkaar te verschillen.

4. DISCUSSIE

Het doel van dit onderzoek was inzicht te krijgen in de relatie tussen persoonlijkheid en effectiviteit van assessoren. Personen met een persoonlijkheidsprofiel dat overeenkomt met het assessorprofiel zouden betere assessoren zijn dan assessoren met een afwijkend profiel.

Om vast te stellen wat een goede score op een oefening was, werd gebruik gemaakt van een expert-rating. Als controle voor de expertscore werd ook een normscore berekend aan de hand van de gemiddelde score van de respondenten. Er waren slechts enkele verschillen tussen de analyse waarbij de expertscore de normscore was en de analyse waarbij de gemiddelde score van de respondenten de normscore was. Daarom werd er in dit onderzoek alleen gebruik gemaakt van de expertscore.

4.1. Conclusies

Het onderzoek naar de relatie tussen persoonlijkheid en de effectiviteit van assessoren heeft geen overtuigend resultaat opgeleverd voor de voorspellende waarde van het door de KAS opgestelde assessorprofiel. Voor twee van de vijf schalen van de NEO-FFI zijn wel significante resultaten gevonden die overeenstemmen met de verwachtingen van het assessorprofiel.

- een goede assessor scoort voor de assessmentoefening 'Verlof 1' gemiddeld op Openheid;
- een goede assessor scoort voor de assessmentoefening 'Puzzel' hoog op Consciëntieusheid.

De kans dat bij twee van de vijf schalen significante resultaten worden gevonden kan ook op toeval berusten. (toevalskans 5% van 7×5 correlaties = 1.75). Daarbij is het op grond van één assessmentoefening niet mogelijk stellige uitspraken te doen over de samenhang tussen de persoonlijkheid van een assessor en zijn manier van beoordelen. Een mogelijke verklaring voor het grotendeels ontbreken van relaties tussen persoonlijkheidsfactoren en assessmentscores kan zijn dat de in dit onderzoek gebruikte oefeningen geen sterke voorspellers zijn. Een simulatie (gedragsproef) dient gerelateerd te zijn aan de functie (Guidelines, 2000). Dit is niet het geval met de oefeningen die bij dit onderzoek gebruikt zijn. De oefeningen komen te weinig overeen

met een echte assessortaak. De oefening die het meeste overeenkomt met een assessortaak (Verlof 2) heeft een lage interne consistentie (tabel 6). In een assessmentcenter staat een gedragsproef niet op zich zelf, maar maakt deel uit van een totale assessmentprocedure, die onder andere bestaat uit tests en interview(s). Schmidt en Hunter (1998) geven aan dat een assessmentuitkomst maar 4% verklaarde variantie toevoegt boven de score op een (goede) intelligentietest. Voor het beoordelen van competenties zijn verschillende assessmentmethodes nodig (Straetmans, 2004). Gedragsproeven dienen gecombineerd te worden met andere indicatoren zoals stagebeoordelingen om assessments te kunnen valideren.

Behalve vijf hypothesen zijn er ook drie exploratieve onderzoeksvragen onderzocht. Hieruit blijkt dat persoonlijkheidsfactoren en demografische factoren geen significante bijdrage leveren aan de wijze van beoordelen. Ook toevoeging van de oefening 'puzzel' als maat van intelligentie bij de demografische factoren heeft geen significante invloed op de wijze van beoordelen. Voor de factoren 'leeftijd' en 'geslacht' blijken er significante verschillen tussen de groepen te zijn. Bij de factor 'leeftijd' is in de laagste leeftijdscategorie het verschil met de absolute score het grootst en in de hoogste leeftijdscategorie het laagst. Dit betekent dat bij de laagste leeftijdscategorie er een groot verschil is tussen de score op de assessmentoefeningen en de normscore en bij de hoogste leeftijdscategorie er een klein verschil is tussen de score op de assessmentoefeningen en de normscore. Het lijkt erop dat oudere assessoren beter beoordelen. Bij de factor 'geslacht' is er de tendens dat mannen beter beoordelen dan vrouwen. In tabel 1 is te zien dat in de hoogste leeftijdscategorie mannen meer vertegenwoordigd zijn dan vrouwen. Aangezien persoonlijkheid er niet toe lijkt te doen is de meest voor de hand liggende verklaring dat het gaat om ervaring. Dit sluit aan bij onderzoek van Spence en El-Ansari (2004) en onderzoek van Adrales, Park, Chu, Witzke, Donnelly, Hoskins, Mastrangelo en Gandsas (2003) die ook constateerden dat ervaring een belangrijke rol speelt bij het goed uitvoeren van werkzaamheden. Straetmans (2006) beschrijft dat een competentie niet het automatische resultaat is van de

verworvenheid van de bekende trits kennis, vaardigheid en attitude, maar dat voor het verwerven van een competentie veel ervaring nodig is.

4.2. Beperkingen van dit onderzoek

1. Training assessoren. Dit onderzoek werd uitgevoerd tijdens workshops en trainingen aan aspirant assessoren. De duur van een workshop en/of training was heel verschillend, waardoor niet alle assessoren dezelfde oefeningen hebben gemaakt. Dit kan van invloed zijn geweest op de resultaten van dit onderzoek.
2. Predictoren (NEO-FFI). Bij deze workshops en trainingen werd gebruik gemaakt van de NEO-FFI (60 items; duur afname 10-15 min.). De aspirant assessoren verwerkten de gegevens van de vragenlijst zelf. De NEO-FFI is afgeleid van de NEO-PI-R die 240 items bevat, waardoor per domein nog met subschalen gewerkt kan worden. Zo is het domein Neuroticisme onderverdeeld in de subschalen N1 Angst; N2 Ergernis; N3 Depressie; N4 Schaamte; N5 Impulsiviteit en N6 Kwetsbaarheid. Door het gebruik van de NEO-PI-R zou het assessorprofiel gedetailleerder beschreven kunnen worden zodat te zien is welk facet van de persoonlijkheid samenhangt met de wijze van beoordelen. In het kader van de workshops/trainingen was het niet mogelijk om de NEO-PI-R te gebruiken. Het zou teveel tijd kosten (240 items, duur afname 40-50 min.) en de interpretatie zou door een psycholoog uitgevoerd moeten worden.
3. Criteriumscores. De assessmentoefeningen in dit onderzoek werden gebruikt bij assessortrainingen met het doel assessoren bewust te maken van de moeilijkheden bij het observeren. Deze oefeningen stonden te ver af van de onderwijspraktijk, waardoor de inhoudsvaliditeit te gering was. Daarbij was de interne consistentie voor de oefening ‘Verlof 2’ te laag (tabel 6) en de samenhang tussen de scores op de verschillende oefeningen te gering (tabel 7).

4. Methode. Voor de expertscore is gebruik gemaakt van drie experts. De expertscore had krachtiger kunnen zijn als er gebruik was gemaakt van meer experts.

Voor de scores op de persoonlijkheidsvragenlijst waren de ruwe scores ingedeeld in klassen. Door deze indeling vond variantieverlies plaats en werden de correlaties gedrukt. Ook de factoren 'leeftijd' en 'opleiding' zijn ingedeeld in klassen, waardoor variantieverlies optrad.

5. Respondenten. De respondenten waren aspirant assessoren, die weinig ervaring hadden met assessortaken. Bij ervaren assessoren zouden de resultaten meer overeen kunnen komen met het KAS-assessorprofiel.

4.3 Aanbevelingen

In dit onderzoek zijn gedragsoefeningen gebruikt die weinig overeen komen met de werkelijke assessortaak. Verder onderzoek naar de voorspellende waarde van het assessorprofiel zou zich moeten richten op gedragsoefeningen die overeenkomen met een werkelijke assessortaak.

Assessmentbeoordelingen moeten gecorreleerd worden met andere indicatoren zoals cijfers van tentamens en stagebeoordelingen. Verder onderzoek naar de relatie tussen persoonlijkheid en de effectiviteit van assessoren zou zich moeten richten op het gebruik van meerdere onderwijsindicatoren. Uit dit onderzoek blijkt dat het waarschijnlijk is dat meer levenservaring of meer praktijkervaring effect heeft op de wijze van beoordelen. Verder onderzoek naar de voorspellende waarde van het assessorprofiel zou daarom moeten gebeuren met ervaren assessoren.

LITERATUUR

- Adrales, G.L., Park, A.E., Chu, U.B., Witzke, D.B., Donnelly, M.B., Hoskins, J.D., Mastrangelo, M.J., & Gandsas, A. (2003). A valid method of laparoscopic simulation training and competence assessment. *Journal of Surgical Research*, 114 (2), 156-162.
- Barrick, M.R., & Mount, M.K. (1991). The Big Five Personality Dimensions and Job Performance: a meta-analysis. *Personnel Psychology: a journal of applied research*, 44(1), 1-26.
- Bartels, L.K., & Doverspike, D. (1997). Assessing the Assessor: The Relationship of Assessor Personality to Leniency in Assessment Center Ratings. *Journal of Social Behaviour and Personality*, 12(5), 179-190.
- Bernardin, H.J., Cooke, D.K., & Villanova, P. (2000). Conscientiousness and Agreeableness as Predictors of Rating Leniency. *Journal of Applied Psychology*, 85(2), 232-234.
- Borman, W. (1975). Effects of Instruction to Avoid Halo Error on Reliability and Validity of Performance Evaluation Ratings. *Journal of Applied Psychology*, 60(5), 556-560.
- Cattell, R.B. (1965). *The scientific analysis of personality*. Hammondsworth: Penguin.
- Competentieprofiel interne assessoren (2002) [Elektronische versie]. VAPRO-OVP, 1-13.
- Costa, P.T., Jr., & McCrae, R.R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.
- Cuppen, H. (2000). Toetsen van feitjes maakt plaats voor beoordelen van gedrag. *HBO-journaal*, (juni/juli), 32-35.
- Dam, E. van, Klaijisen, A., & Frietman, J. (2003). *De rol van assessor in EVC-procedures* [Elektronische versie]. Nijmegen: Kenniscentrum Beroepsonderwijs Arbeidsmarkt.
- Elshout-Mohr, M., Oostdam, R., Dietze, A., & Snoek, M. (2001). Assessment van competenties. *Velon: tijdschrift voor lerarenopleiders*, 22(2), 48-54.

- Gaugler, B.B., Rosenthal, B., Thornton III, G.C., & Bentson, C. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology Monograph*, 72(3), 493-511.
- Goldberg, L.R. (1990). An Alternative “Description of Personality”: The Big-Fiver factor Structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Groot, A.D. de (1994). *Methodologie: grondslagen van onderzoek en denken in de Gedragwetenschappen* (12^e dr). Assen: Van Gorcum.
- Guidelines and Ethical Considerations for Assessment Center Operations: International Task Force on Assessment Center Guidelines (2000). *Public Personnel Management*, 29(3), 315-332.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley & Sons.
- Hendriks, P., & Schoonman, W. (Eds.) (2006). *Handboek assessment deel 1, gedragsproeven, ontwikkeling, implementatie en evaluatie* [met bijbehorende website www.assessment-pro.nl.] Assen: Van Gorcum.
- Herriot, P. (2003). Assessments by groups: Can value be added? *European Journal of Work and Organizational Psychology*, 12, 131-145.
- Hoekstra, H.A., Ormel, J., & De Fruyt, F. (1996). *Handleiding bij de NEO Persoonlijheids Vragenlijsten NEO-PI-R en NEO-FFI* [Neo Personality Inventories NEO-PI-R and NEO-FFI: Manual]. Lisse: Swets & Zeitlinger.
- Hofstee, W.K.B. (1999). *Principes van beoordeling. Methodiek en ethiek van selectie, examinering en evaluatie*. Lisse: Swets & Zeitlinger.
- John, O.P., & Srivastava, S. (1999). The Big Five Trait Taxonomy: History, Measurement and Theoretical Perspectives. In L.A. Pervin & O.P. John (Eds.), *Handbook of Personality: Theory and Research*, (2nd ed.) (pp. 102-138). London: The Guilford Press.
- Kouwer, B. (1963). *Het spel van de persoonlijkheid. Theorieën en systemen in de psychologie van de menselijke persoon*. Utrecht: Erven Bijleveld.

- Lievens, F. (2000). De invloed van verschillende types assessorentraining op observaties en beoordelingen in assessment center oefeningen. *Gedrag en Organisatie*, 13(4), 185-196.
- Lievens, F., & Fruyt, F. De (2001). Assessor's use of personality traits in descriptions of assessment centre candidates: a five factor model perspective. *Journal of Occupational and Organizational Psychology*, 74, 623-636.
- Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the Impact of Judge Severity on Examination Scores. *Applied Measurement in Education*, 3, 331-394.
- Maesen de Sombreff, P.E.A.M. van der (1999). Personeelsbeoordeling. In G.H.M. Evers & F. Kluytmans (Eds.), *Personeelsmanagement*. Heerlen: Open Universiteit.
- McCrae, R.R. (1990). Traits and Trait Names: How well is Openness Represented in Natural Languages? *European Journal of Personality*, 4, 119-129.
- Onderwijsraad (2002). *Competenties: van complicaties tot compromis*. Over schuifjes en begrenzers. Den Haag: Onderwijsraad.
- Ones, D.S., & Mount, M.K. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, 47 (1), 147-157
- Reber, A.S. (1997). *Woordenboek van de psychologie*. Amsterdam: Bert Bakker.
- Ree, M.J., & Carretta, T.R. (1998). General Cognitive Ability and Occupational Performance. *International review of industrial and organizational psychology*, 13, 159-184.
- Roe, R.A. (1983). *Grondslagen der personeelsselectie*. Assen: Van Gorcum.
- Roossink, H.J. (2006). *Valkuilen voor beoordelaars*. Enschede: Universiteit Twente.
Retrieved June 20, 2006, from <http://www.utwente.nl>
- Sadler, D.R. (1983). Evaluation and the Improvement of Academic Learning. *Journal of Higher Education*, 54, 60-79.
- Sande, J.P. van de (1999). *Gedragsobservatie. Een inleiding tot systematisch observeren*. Groningen: Martinus Nijhoff.

- 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 (2), 262-274.
- Schmidt, F.L., Hunter, J.E., & Outerbridge, A.N. (1986). Impact of Job Experience and Ability on Job Knowledge, Work Sample Performance, and Supervisory Ratings of Job Performance. *Journal of Applied Psychology*, 71(3), 432-439.
- Schoonman, W. (2004). *Assessment voor en door iedereen*. Enschede: Saxion Hogescholen (lectorale rede).
- Schoonman, W. (2005). De zwakste schakels: over assessment door assessoren. *Onderzoek van Onderwijs*, 34(maart), 10-13.
- Schuler, Y., & Wee, E. ter (2001). Assessoren aan het werk. *Velon: tijdschrift voor lerarenopleiders*, 22(2), 43-47.
- Slotboom, A., & Elphick, E. (1999). Constructie van een Big Five persoonlijkheidsvragenlijst. *De psycholoog*, 33(jan), 21-22.
- Spence, W., & El-Ansari, W. (2004). Portfolio assessment: practice teacher's early experience. *Nurse Education Today*, 24 (5), 388-401.
- Stichting HON (2000). *Onderwijs op maat in een activerend klimaat: onderwijsvisie van de Saxion Hogescholen* [Elektronische versie]. Enschede, Deventer: Stichting HON.
- Straetmans, G.J.J.M. (1998). Toetsing van competenties. In P.W.J. Schramade (Red.). *Handboek Effectief Opleiden*, (pp. 67-88) 's-Gravenhage: Elsevier Bedrijfsinformatie
- Straetmans, G.J.J.M.(2004). Protocol Portfolio Scoring: praktisch artikel. *Onderwijsinnovatie*, 4, 2, 17-27.
- Straetmans, G.J.J.M. (2006). *Bekwaam beoordelen en beslissen: beoordelen in competentiegerichte beroepsopleidingen*. Enschede: Saxion Hogescholen (lectorale rede).

Tett, R.P., & Jackson, D.N. (1994). Meta-analysis of personality-job performance relations: a reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology*, 47 (1), 157-171.

Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44 (4), 703-742.

Wijnhoven, L. (2001). *Maatwerk 3: voortgangsrapportage* [Elektronische versie]. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.

Zonderland, B. (1995). *Kwaliteit van assessment-oordelen "de beoordelaar beoordeeld"*. Utrecht: Universiteit Faculteit Sociale wetenschappen.

BIJLAGE 1: observatieformulier 'Apies'

Apies kijken

Gedragsobservatie

Hoofdrolspeler: Wouter

*(het gaat om **zijn** gedrag!)*



Indicatoren	Aantal
	(turven)
Gedrag	
Rolt boomstronk om	
Zet boomstronk rechtop	
Beklimt boomstronk	
In rust	
Totaal	

Omschrijvingen	
Rollen	Verplaatst boomstronk horizontaal gedurende tien seconden
Rechtop zetten	Beweegt en/of zet boomstronk in een verticale positie
Beklimmen	Verplaatst zichzelf via boomstronk omhoog
In rust zijn	Vertoont geen zichtbare fysieke activiteit

© 2005, Saxion



Kenniskring en
Lectoraat Assessment

SAXION
Hogeschole

BIJLAGE 2: observatieformulier 'Verlof 1' en 'Verlof 2'

Het verlof



Gedragsobservatie – Adviesgesprek

Indicatoren	Gezien	Kwaliteit
	1 of 0	+ 0 -
Relatie opbouwen		
Neemt tijd het ijs te breken		
Stelt cliënt op z'n gemak		
Vertoont uitnodigende lichaamshouding		
Cliëntvraag formuleren		
Vraagt door over de precieze hulpvraag		
Herformuleert de hulpvraag		
Checkt de geherformuleerde hulpvraag		
Oplossingen aanbieden		
Biedt cliënt meerdere oplossingen		
Onderzoekt met cliënt meerdere oplossingen		
Laat cliënt de ruimte te kiezen		
Voortgang borgen		
Herhaalt gemaakte afspraken		
Spreekt vervolgacties af		
Totaal		

Omschrijvingen	
Relatie opbouwen	
Neemt tijd het ijs te breken	Welkom, small talk, humor, algemeen gesprek- onderwerp, biedt iets te drinken aan
Stelt cliënt op z'n gemak	Ik heb ... tijd gereserveerd. Het besprokene blijft tussen ons.
Vertoont uitnodigende lichaamshouding	Voorovergebogen, kijkt cliënt aan,
Cliëntvraag formuleren	
Vraagt door over de precieze hulpvraag	Waarom? Hoe zit dat precies? Waarvoor?
Herformuleert de hulpvraag	Dus je zoekt een oplossing voor ... Betekent dit dat... Waar je op uit wil komen is ...
Checkt de geherformuleerde hulpvraag	Klopt dat? Is het dus zo dat? Bedoel je dus dat
Oplossingen aanbieden	
Biedt cliënt meerdere oplossingen	Er zijn meerdere mogelijkheden. We kunnen denken aan ... en ... Waar ik aan zit te denken is ... of ...
Onderzoekt met cliënt meerdere oplossingen	Als je dit kiest, betekent dat. Welke spreekt je het meest aan? Wat past het best?
Laat cliënt de ruimte te kiezen	Welke richting spreekt jou het meest aan? Wat heeft jouw voorkeur? Misschien wil je er over nadenken?
Voortgang borgen	
Herhaalt gemaakte afspraken	Dus dan doen we dit en dit. Jij doet dit, ik doe dat. We hebben dus afgesproken dat ...
Spreekt vervolgacties af	Dan maken we een nieuwe afspraak over ... Welk nieuwe afspraak ... Wanneer spreken we elkaar weer?

BIJLAGE 3: Normscore experts**Apies kijken***Normscore experts (N=3) voor de oefening Apies kijken*

Item	Normscore
Rolt boomstronk om	5
Zet boomstronk rechtop	5
Beklimt boomstam	6
In rust	4

Kwaliteit verlof 1*Normscore experts (N=3) oefening Verlof 1 en Kwaliteit Verlof 1*

Item	Verlof 1	Kwaliteit verlof 1
1. Neemt tijd het ijs te breken	1	0
2. Stelt cliënt op zijn gemak	0	
3. Uitnodigende houding	1	0
4. Vraagt door	0	
5. Herformuleert	0	
6. Checkt	0	
7. Biedt meerdere oplossingen	1	-1
8. Onderzoekt oplossingen	0	
9. Laat ruimte om te kiezen	0	
10. Herhaalt afspraken	0	
11. Spreekt vervolgacties af	0	

Van het zichtbare geobserveerde gedrag is de kwaliteit vastgesteld. Score -1 wil zeggen dat de kwaliteit van het geobserveerde gedrag negatief is; score 0 wil zeggen dat de kwaliteit van het geobserveerde gedrag neutraal is en score +1 wil zeggen dat de kwaliteit van het geobserveerde gedrag positief is. Als er niets is ingevuld is het gedrag niet gezien.

Kwaliteit verlof 2

Van het zichtbare geobserveerde gedrag is de kwaliteit vastgesteld. Score -1 wil zeggen dat de kwaliteit van het geobserveerde gedrag negatief is; score 0 wil zeggen dat de kwaliteit van het geobserveerde gedrag neutraal is en score +1 wil zeggen dat de kwaliteit van het geobserveerde gedrag positief is. Als er niets is ingevuld is het gedrag niet gezien.

Normscore experts(N=3) oefening Verlof 2 en Kwaliteit Verlof 2

	Verlof 2	Kwaliteit verlof 2
1. Neemt tijd het ijs te breken	1	+1
2. Stelt cliënt op zijn gemak	0	
3. Uitnodigende houding	1	0
4. Vraagt door	1	+1
5. Herformuleert	1	0
6. Checkt	1	0
7. Biedt meerdere oplossingen	1	+1
8. Onderzoekt oplossingen	1	0
9. Laat ruimte om te kiezen	1	+1
10. Herhaalt afspraken	0	
11. Spreekt vervolgacties af	1	0

BIJLAGE 4: Coëfficiënten regressieanalyse*Tabel 1*

Lineaire regressieanalyse met als predictoren persoonlijkheidsfactoren voor de afhankelijke variabele totaalscore op alle oefeningen (n = 105)

Variabele	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Neuroticisme	-.76	.55	-.14	.17
Extraversie	-.90	.53	-.17	.09
Openheid	.19	.42	.05	.50
Altruïsme	.10	.42	.03	.81
Consciëntieusheid	.83	.56	.16	.14

Tabel 2

Lineaire regressieanalyse met als predictoren demografische factoren en persoonlijkheidsfactoren voor de afhankelijke variabele totaalscore op alle oefeningen

(n = 82)

Variabele	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Stap 1				
Geslacht	-.95	.69	-.16	.17
Leeftijd	.23	.33	-.08	.50
Opleiding	-.16	.54	-.03	.77
Ervaring (max. 5 jaar)	-.10	.28	-.04	.73
Training	.20	2.24	.01	.93
Stap 2				
Geslacht	-.84	.76	-.14	.27
Leeftijd	.40	.37	.14	.28
Opleiding	-.18	.58	-.04	.76
Ervaring (max. 5 jaar)	-.06	.32	-.03	.84
Training	-.23	2.34	-.01	.92
Neuroticisme	-.25	.74	-.05	.74
Extraversie	-.51	.69	-.10	.46
Openheid	.09	.51	.02	.86
Altruïsme	-.02	.46	-.00	.97
Consciëntieusheid	.96	.66	-.18	.15

Tabel 3

Lineaire regressieanalyse met als predictoren demografische factoren (+ oefening puzzel) en persoonlijkheidsfactoren voor de afhankelijke variabele totaalscore op de oefeningen

(n = 82)

Variabele	<i>B</i>	<i>SE B</i>	β	<i>p</i>
Stap 1				
Geslacht	-.75	.56	-.16	.19
Leeftijd	-.04	.27	-.02	.87
Opleiding	.03	.44	.01	.95
Ervaring (max. 5 jaar)	.27	.23	.13	.25
Training	1.68	1.8	.11	.35
Puzzel	.01	.04	.04	.71
Stap 2				
Geslacht	-.93	.62	-.19	.39
Leeftijd	.08	.30	.03	.80
Opleiding	.11	.47	.03	.81
Ervaring (max. 5 jaar)	.26	.25	.13	.31
Training	1.80	1.88	.12	.34
Puzzel	.02	.04	.07	.58
Neuroticisme	.55	.60	.13	.36
Extraversie	.17	.56	.03	.84
Openheid	.09	.41	.03	.82
Altruïsme	-.14	.40	-.05	.72
Consciëntieusheid	.77	.53	.19	.15