

15.9-1. Vorbereiding door de sollicitant op psychologisch onderzoek¹

Effecten van Impression Management, oefening en coaching

Dr. W. Bloemers

¹ Dit artikel is gebaseerd op hoofdstuk 12 van de cursus Selectiepsychologie: Selectie en Assessment van de Open Universiteit Nederland, 2006. De auteur dankt Wim Hofstee voor diens commentaar op eerdere versies van dit manuscript.

Inhoud

1	Inleiding	15.9-1.03
2	Impression management, oefening en coaching	15.9-1.03
2.1	Oefening en coaching: argumenten pro	15.9-1.04
2.2	Oefening en coaching: argumenten contra	15.9-1.05
3	De procedure bij het psychologisch onderzoek als buffer tegen onwenselijke gevolgen van oefening en coaching	15.9-1.06
4	Drie vragen: invloeden op predictorscores, fairness en predictieve validiteit	15.9-1.07
5	Personeelsselectie: predictoren en hun voorspellende waarde	15.9-1.08
6	De intelligentietest: voorbereidingseffecten	15.9-1.10
7	De Assessment Center Methode (ACM)	15.9-1.12
7.1	Effecten van coaching en oefening op ACM-simulaties	15.9-1.12
7.2	Impression Management Tactieken bij de ACM	15.9-1.13
7.3	Coachingseffecten op prestaties tijdens de leiderloze groepsdiscussie (LGD)	15.9-1.15
7.4	Coachingseffecten op de In Basket Opdracht (IBO)	15.9-1.15
8	Het selectie-interview	15.9-1.16
8.1	Wat wordt er beoordeeld in selectie-interviews?	15.9-1.17
8.2	Selectie-interviews en IM	15.9-1.18
8.3	Effecten van coaching op selectie-interviewuitkomsten	15.9-1.19
9	Persoonlijkheidsvragenlijsten	15.9-1.20
9.1	Faking op persoonlijkheidsvragenlijsten	15.9-1.21
9.2	Faking op integriteitsvragenlijsten	15.9-1.23
10	Trainingseffecten op overige selectie-instrumenten	15.9-1.25
11	Conclusie	15.9-1.26
11.1	Effecten: algemeen overzicht	15.9-1.26
11.2	Fairness	15.9-1.27
11.3	Predictieve validiteit	15.9-1.28
11.4	Coaching en faking	15.9-1.28
12	Tot besluit: een nieuw selectiescenario?	15.9-1.29
	Literatuur	15.9-1.30

Auteur:

Dr. W. (Wim) Bloemers is universitair docent Arbeids- en Organisationspsychologie aan de Open Universiteit Nederland. Ook adviseert hij bedrijven en instanties op het gebied van selectie en assessment.

E-mail: wimbloemers@wxs.nl

1 Inleiding

In de jaren tachtig ontstond er een kleine rel binnen het Nederlands Instituut van Psychologen (NIP). Er was (nota bene) een NIP-psycholoog die tegen betaling sollicitanten trainde door hen (authentieke) psychologische tests te laten oefenen. Dit kon niet door de beugel omdat openbaar maken van testopgaven aan een beperkt aantal individuen de *fairness* en de kwaliteit van selectieprocedures aantast. Verder waren het blijkbaar vooral de laatlunkende uitlatingen van deze psycholoog over zijn collega's en hun zogenaamd wetenschappelijk verantwoorde testmethoden die er toe leidden dat hij als NIP-lid geroyeerd werd (voor een uitgebreide beschrijving van deze kwestie, zie Bloemers, 1988, 1989).

Tegenwoordig wordt de soep niet meer zo heet gegeten. Oefening op psychologische tests in het kader van sollicitaties is algemeen goed, zij het niet met de authentieke tests maar met afgeleide vormen. Het *world wide web* staat bol van de oefensites en er is een scala aan oefenboeken voor psychologische tests, assessmentoefeningen en interviews op de markt (Kok & De Jongh, 2004; Bloemers, 2001, 2005; Van Minden, 2006). Ook zijn er verschillende bureaus die coaching (oefening) op een psychologisch onderzoek of assessment bieden waaronder ook het bureau van de toenmalige psycholoog (zij het niet langer als NIP-lid).

Het is zelfs zo dat veel selectiebureaus tegenwoordig voorbereiding op een psychologisch onderzoek expliciet aanraden. Op de sites van selectiebureaus staat meestal een lijstje met geadviseerde voorbereidingsliteratuur. Blijkbaar heeft de testpsycholoog van de 21e eeuw liever een protoprofessionele, goed voorbereide sollicitant, dan een naïeve leek tegenover zich. Oriëntatie en oefening zijn echter nog geen coaching. Hieronder wordt eerst ingegaan op de verschillen tussen Impression Management, oefening en coaching.

Vervolgens worden effecten van coaching en oefening besproken voor de meestgebruikte selectiemiddelen in instrumenten. Afgesloten wordt met een idee voor een nieuw scenario voor de selectiepraktijk, waarin de lijn van de autonome, goed geïnformeerde cliënte wordt doorgetrokken naar een horizontaal selectieparadigma.

2 Impression management, oefening en coaching

Verschillende auteurs hebben gepoogd om de verschillende voorbereidingsvormen op een psychologische instrument te rubriceren. Als er in een selectieonderzoek iets op het spel staat, ook wel *high stakes testing* genoemd, (zie Sackett, Schmitt, Ellingson, & Kabin, 2001) mag worden aangenomen dat kandidaten zullen proberen een zo goed mogelijk beeld van zichzelf te scheppen, zodat er sprake is van Impression Management (IM). Onder IM kan worden verstaan: '...the process whereby people seek to control the image others have of them' (Rosenfeld, Giacalone & Riordan, 1995a). Meer algemeen gaat het bij IM om het beïnvloeden van de perceptie en het gedrag van anderen door het beheersen en sturen van de informatie die we afgeven omtrent onszelf en ons gedrag (zie ook Rosenfeld, Giacalone & Riordan, 1995b). IM kan dus gezien worden als een normaal onderdeel van een sociaal interactieproces waarin mensen informatie kunnen sturen en beheersen om een zo goed mogelijk beeld van zichzelf te

scheppen. In de selectiesituatie speelt dit vooral als het gaat om *typical* behavior (persoonlijkheidsvragenlijsten en interviews). Bij *maximum* behavior (prestatietests) is informatie door de kandidaat per definitie veel minder of zelfs niet stuurbaar en beheersbaar (Sackett, Zedeck & Fogli, 1988). IM bij prestatietests is dan een afgeleide of een gevolg van oriëntatie en oefening om tot een maximale score op de desbetreffende predictor te komen. Meer extreme vormen van IM bij prestatietests leiden tot verstoring van de fairness: er is dan sprake van voorkennis of fraude.

Er zijn talloze mogelijkheden voor kandidaten om zich op een of andere wijze voor te bereiden op het instrumentarium van het gemiddeld personeelselectieonderzoek. In het overzichtsartikel van Sackett, Burris en Ryan (1989) wordt een onderscheid gemaakt naar *inhoud* (oefenen met vergelijkbare items); vergroten testwijsheid; verbeteren probleemoplossend vermogen (Anastasi, 1981), naar *acceptabiliteit* (onacceptabel versus acceptabel) (Mehrens, 1989), naar *functie* (verhogen predictorscores versus het verbeteren van onderliggende vaardigheden) (Messick, 1981) en naar al of niet aanwezig zijn van een *externe interventie* (Kulik et al. 1984).

Is dit laatste het geval dan is er sprake van coaching; in alle andere gevallen is er sprake van oefening, waarbij het leren van de eigen ervaring bij het maken van oefenopgaven centraal staat. Deze laatste vorm van voorbereiding is in Nederland algemeen geaccepteerd en wordt zelfs zoals gezegd gestimuleerd. Sackett et al. (2001) sluiten zich wat betreft een hanteerbare definitie van voorbereiding aan bij Kulik et al. (1984). Zo ontstaat wat betreft voorbereiding op een selectieprocedure een onderscheid tussen *algemene oefenprogramma's* (een kortdurende algemene oriëntatie op test en selectieonderdelen, waarbij de kandidaat kennismaakt met voorbeelden van testopdrachten) en een meer intensieve vorm van voorbereiding genaamd *coaching*. Onder coaching verstaan Sackett et al. een vorm van voorbereiding op (elementen van) een psychologisch onderzoek waarbij er wordt geoefend met relevant materiaal en waarbij ook feedback wordt gegeven op de resultaten.

2.1 OEFENING EN COACHING: ARGUMENTEN PRO

Wat betreft de Nederlandse situatie kan gesteld worden dat voorbereiding door middel van algemene oefenvoorzieningen inmiddels gemeengoed is. De vraag is dan ook niet of coaching/oefening mag, het gebeurt immers gewoon. De vraag is welke vormen van voorbereiding acceptabel zijn of zelfs positieve effecten hebben en hoe deze positieve varianten van oefening/coaching gepropageerd kunnen worden. Naast het 'fait accompli' van het bestaan en gebruik van voorbereidingsmogelijkheden zijn er argumenten te geven waarom een gereguleerde vorm van oefening positieve effecten kan hebben.

1 Doordat sollicitanten door oefening beter begrijpen wat de bedoeling van een selectie-instrument is, zullen de constructvaliditeit en de predictieve validiteit toenemen. Dit argument geldt echter niet voor algemene intelligentietests, die per definitie het vermogen (*g*) meten om met nieuwe onbekende informatie om te gaan. Daar zou coaching juist het 'zuiver' meten van het *g* gehalte verminderen, daar coaching zich richt op aan te leren regels, procedures en trucjes (Zie bijv. Te Nijenhuis, Voskuil & Schijve, 2001; Hough & Ones, 2001).

2 Door gereguleerde oefenmogelijkheden worden verschillen in testervaring geminimaliseerd. Elke sollicitant is in dezelfde mate voorbereid op de test, zodat er sprake is van een faire selectiesituatie. Dit argument gaat echter voorbij aan differentiële effecten van voorbereiding (o.a. Hausknecht, Trevor & Farr, 2002): minder intelligente mensen boeken door oefening minder winst dan intelligentere mensen.

3 Door oefening wordt testangst gereduceerd. Sollicitanten zullen met meer zelfvertrouwen de test ingaan en een optimale prestatie kunnen en willen leveren (Zeidner, 1998; McCarthy & Goffin, 2005). Bovendien verdwijnt door het afnemen van testangst een errorvariantiebron: de constructvaliditeit en de predictieve validiteit nemen toe. Het gebeurt nu nogal eens dat kandidaten door testangst een te lage score krijgen en dus ten onrechte niet worden aangenomen.

4 Oefening is een teken van verhoogde motivatie, het gaat hierbij om kandidaten die de baan echt willen. Deze kandidaten zullen ook in de functie zelf een verhoogde motivatie tonen. In die zin werken oefenmogelijkheden als een zelfselectiemechanisme en als middel om optimaal te presteren voor sollicitanten (Anderson, Born & Cunningham-Snell, 2001).

5 Oefening, als element van transparantie, verhoogt de acceptabiliteit van de selectieprocedure. Er ontstaat een meer harmonieuze relatie waarbij de sollicitant het idee heeft dat hij zelf invloed kan uitoefenen op de test.

6 Oefening bevordert zelfselectie: door te oefenen met relevant selectiemateriaal krijgen sollicitanten een reëel beeld van de functievereisten en hun geschiktheid. Via zelfselectie blijven alleen de meer geschikte en gemotiveerde sollicitanten over. Dit bevordert de efficiency en de uitkomsten van selectieprocedures (zie bijv. Hoff Macan, Avedon, Paese & Smith, 1994; Chan, Schmitt, Sacco & DeShon, 1998; Lievens, De Corte & Brysse, 2003; Anderson, 2004; Chan & Schmitt, 2004).

Bovengenoemde zes voordelen gelden uitsluitend voor oefenmogelijkheden waarbij sollicitanten kunnen *leren* van hun ervaringen met relevante afgeleide voorbeelden van selectiemiddelen. Het geldt niet voor die vormen van coaching waarbij wordt geoefend met echte of parallelvormen van bestaand selectiemateriaal (testopgaven) en waarbij een externe partij betrokken is. Deze laatste vormen van *coaching* zijn alleen acceptabel als er sprake is van een aantoonbare lagere score op selectie-instrumenten terwijl het vermoeden bestaat dat de score op de onderliggende latente trek hoger is, zoals bij minderheidsgroepen soms het geval is. Coaching kan in zo'n geval discriminatie of bias reduceren of ongedaan maken (Sackett et al., 2001).

2.2 OEFENING/COACHING: ARGUMENTEN CONTRA

De gedachte dat oefening/coaching leidt tot een verhoogde score die op haar beurt leidt tot een te positief advies ligt voor de hand. In dat geval zou oefening/coaching het aantal *false positives* en *false negatives* in een selectieprocedure kunnen verhogen. Meer kandidaten worden onterecht aangenomen en dus andere onterecht afgewezen, bij vergelijkende selectie althans. De gevolgen hiervan zijn drieërlei:

- Economische gevolgen voor de *geadviseerde organisatie*. Deze krijgt een kandidaat die in de praktijk minder presteert dan het advies (de voorspelling) aan geeft. Dit kan tot aanmerkelijke economische schade voor betrokken bedrijven leiden (zie bv. Schmidt & Hunter, 1983; Van der Maesen de Sombreff, 1997).
- Menselijke en materiële schade voor *sollicitanten*. In die zin dat te hoog geadviseerde sollicitanten te maken krijgen met loopbaanafbreuk met alle mogelijke psychische en materiële gevolgen van dien. Onterecht afgewezen kandidaten krijgen eveneens te maken met loopbaanafbreuk en raken geblokkeerd in hun professionele, persoonlijke en materiële ontwikkeling.
- Economische en professionele schade voor het *betrokken adviesbureau*. Het bureau levert kandidaten die voor een gedeelte false positives (en false negatives) blijken te zijn. Dit kan tot een verstoring van de relatie met de opdrachtgever leiden, met economische gevolgen. De professionele status van het bureau kan in het geding raken door een groot aantal 'foute' adviezen. Ook kunnen er klachtenprocedures opgestart worden door onterecht afgewezen kandidaten die de validiteit van de selectieprocedure gaan aanvechten.

De negatieve gevolgen van trainingseffecten kunnen dus zowel in economisch als in menselijk als in professioneel opzicht aanzienlijk zijn. Maar zijn ze dat ook? Kunnen sollicitanten door oefening hun prestaties op de verschillende selectie-instrumenten verhogen? En geldt dat voor alle selectie-instrumenten in gelijke mate? Deze vraag betreft de vraag naar de effecten van oefening en coaching op verschillende selectie-instrumenten en -middelen. En wat zijn de gevolgen van oefening en coaching met betrekking tot de fairness. Daarnaast kunnen ook de fairness van de procedure, de predictieve validiteit en zelfs de gemeten constructen beïnvloed worden door oefening, coaching en faking (Hough & Ones, 2001).

3 De procedure bij het psychologisch onderzoek als buffer tegen onwenselijke gevolgen van oefening en coaching

Op voorhand kan wat betreft mogelijke effecten van training/coaching bij personeelsselectie een nuancerende opmerking gemaakt worden. Deze betreft de bufferwerking van de algemene procedure bij psychologisch onderzoek van sollicitanten. Een sollicitant wordt zelden geadviseerd op basis van bijvoorbeeld een geïsoleerde intelligentietestscore. Dit gebeurt nog wel eens bij selectie voor een bepaalde opleiding, maar zelden of nooit bij psychologisch onderzoek in het kader van loopbaanselectie. Daar bestaat de procedure gemiddeld gesproken uit een of meer interviews, vragenlijstonderzoek, een intelligentietest en soms rollenspellen en casussen. Een door oefening verhoogde intelligentiescore zal dus bij een selectiepsycholoog al snel tot argwaan leiden omdat deze 'onnatuurlijk' hoge score niet strookt met andere gedragsindicatoren van de kandidaat. In die zin kan training op losse predictoren een averechts effect hebben. Hetzelfde geldt voor training gericht op persoonlijkheidsvragenlijsten beïnvloeding, daar deze uitkomst wordt gerelateerd aan de bevindingen uit het interview. Kortom: de professioneel werkende selectiepsycholoog heeft voldoende controlemiddelen in huis om geïsoleerde effecten als gevolg van training/coaching op te sporen en te corrigeren. Iets anders is het als de kandidaat in staat is om alle predictorscores door oefening op te hogen, maar het voor de

hand liggende antwoord is dan dat deze kandidaat waarschijnlijk ook wel in staat is om de criteriumscore overeenkomstig op te hogen (de werkprestaties dus), en in het verleden dus onderschat is wat betreft arbeidsprestaties.

Een eerste bespiegeling over de actuele situatie van psychologisch onderzoek voor loopbaanselectie doet vermoeden dat het in de praktijk met de onwenselijke effecten van voorbereiding/training wel mee zal vallen. Daar komt nog bij dat veel kandidaten pas tot coaching zullen besluiten als het moment suprême van de eindbeslissing nabij is. Dat wil zeggen, training/coaching vindt pas plaats in de laatste fase van het selectieproces, als de kandidaat is doorgedrongen tot de laatste selectieronde waarbij zeg nog maximaal vijf kandidaten in het spel zijn uit een oorspronkelijke poule van 50-100. Op basis van voorselectie zijn alle resterende vijf kandidaten acceptabel bevonden, en de rest is lood om oud ijzer. Coaching/training in deze fase is hoogstens te betitelen als 'vuil spel' ten opzichte van de medekandidaten. Daar zou bijvoorbeeld sprake van zijn als de uiteindelijke beslissing genomen wordt op basis van (weer) een intelligentietest, welke dan door een van de kandidaten van te voren geoefend wordt. Maar welk bureau zou in deze fase (opnieuw een intelligentietest als doorslaggevend predictor inzetten? Aangenomen mag worden dat er al eerder op intelligentie geselecteerd is (voor een rationele inkleding van selectieprocedures, zie Bloemers, 1997; en Schmidt en Hunter, 1998). Algemeen bekend is dat de sterkste predictoren (die met de hoogste validiteit) het eerst moeten worden ingezet. Zo blijven de meest geschikte kandidaten zo lang mogelijk in de procedure.

4 Drie vragen: invloeden op predictorcores, fairness en predictieve validiteit

In dit artikel staan drie vragen centraal die nauw met elkaar verweven zijn.

In de eerste plaats: Wat zijn de invloeden van voorbereiding op de verschillende *predictoren* die deel uitmaken van een psychologisch onderzoek (tegenwoordig ook wel 'assessment' genoemd)? Met andere woorden: kan men door oefening/coaching de score op bijvoorbeeld een intelligentietest verhogen? Dit is in feite de primaire vraag want als oefening/coaching nauwelijks of geen effecten heeft is er ook geen invloed op respectievelijk de fairness van de selectieprocedure en de predictieve validiteit.

In de tweede plaats: wat is de invloed van coaching/oefening op de *fairness* van selectieprocedures? Als coaching/oefening effecten heeft en met name differentiële effecten (de ene kandidaat profiteert meer dan de andere) dan doorkruist coaching/oefening de fairness of eerlijkheid/objectiviteit van selectieprocedures. Kan en of moet coaching/oefening bij bewezen differentiële effecten een rol spelen en zo ja hoe en in welke vorm? Deze vraag is actueel bij predictorbias, als kandidaten worden afgewezen vanwege hun te lage score op een predictor terwijl blijkt dat ze in de praktijk evengoed als of beter dan niet afgewezen kandidaten presteren. Vooral bij selectie van minderheden is deze vraag actueel (Sackett et al., 2001).

In de derde plaats gaat het om de vraag naar de invloed van coaching op de *predictieve validiteit* van de betrokken instrumenten. Als coaching effecten heeft op predictorcores, wat zijn dan de gevolgen voor de predictieve validiteit? Heeft coaching/oefening schadelijke gevolgen voor het doen van voorspelling?

gen over het functioneren van de kandidaat? Zoals hiervoor betoogd is dit het geval bij differentiële effecten die door coaching kunnen ontstaan. Bij een algemeen of lineair effect is er geen invloed op de predictieve validiteit en wordt ook de fairness niet aangetast: iedereen profiteert in gelijke mate.

Zoals te zien is, hangen bovengenoemde drie vragen nauw met elkaar samen waarbij de vraag naar het mogelijk differentieel effect van coaching primair is. Selectie moet fair zijn, dat wil zeggen de kandidaten moeten een gelijkwaardige uitgangspositie hebben bij een psychologisch onderzoek wat betreft hun te leveren prestaties op de selectiemiddelen en instrumenten. Dat veronderstelt gelijke toegang tot beschikbare voorbereidingsmiddelen. Gegeven het feit dat iedereen zich kan voorbereiden wordt, bij bewezen effecten van coaching, voorbereiding haast een sine qua non. Maar als er sprake is van differentiële effecten (sommige kandidaten profiteren meer van voorbereiding dan andere) is coaching problematisch. Niet alleen in de zin dat deze dan de eis van fairness doorkruist, maar ook vanwege het feit dat dan de predictieve validiteit van de procedure kan worden aangetast.

Op basis van een overzicht van de absolute en relatieve sterkte van mogelijke predictoren bij personeelsselectie (Tabel 1) zullen we de mogelijkheden en invloeden van training/coaching op deze predictoren bespreken (vraag 1). Besloten wordt met een conclusie over de huidige situatie wat betreft oefenen/coaching ten behoeve van personeelsselectie. Ook zal een perspectief geschetst worden voor een alternatief scenario wat betreft coaching/training in het kader van personeelsselectie. Dit alternatieve scenario dat uitgaat van de sollicitant als opdrachtgever maakt in principe de problemen wat betreft fairness en mogelijke verstoring van de predictieve validiteit overbodig (zie de vragen 2 en 3 hiervoor).

5 Personeelsselectie: predictoren en hun voorspellende waarde

Welke predictoren en/of selectiemiddelen worden er gebruikt bij personeelsselectie en wat is er bekend van de effecten van oefening/coaching ten aanzien van de prestaties op deze instrumenten/middelen?

Tabel 1: De operationele en relatieve validiteit van verschillende selectiemiddelen (predictoren) voor het voorspellen van werkprestaties (Gebaseerd op Schmidt en Hunter, 1998, p.265 en Ryan et al., 1999).

Voorspeller voor prestaties in werktrainingen	Operationele validiteit	Toegevoegde validiteit naast gebruik intelligentietest	Gebruik in Nederland* 1=nooit 5=altijd
Intelligentietest	.51	-	3.8
Arbeidsproeven	.54	24%	?
Gestructureerd Interview	.51	24%	4.3**
Tests voor functie kennis	.48	14%	?
Integriteitsvragenlijsten	.41	27%	3.3***
Ongestructureerd Interview	.38	8%	4.3**

Voorspeller voor prestaties in werktrainingen	Operationele validiteit	Toegevoegde validiteit naast gebruik intelligentietest	Gebruik in Nederland* 1=nooit 5= altijd
Assessment Center Methode	.37	4%	2.8
Collega-beoordelingen	.36	14%	55%
Biografische gegevens	.35	2%	1.5
Consciëntieusheidsvragenlijsten	.31	18%	3.3**
Referenties	.26	12%	2.7
Genoten jaren opleiding	.10	2%	4.7****
Interesse	.10	2%	?
Jaren werkervaring	.18	6%	?
Leeftijd	-.01	0%	?

* Gebaseerd op Ryan et al. (1999). In dit onderzoek werden bedrijven onderzocht met 1000 of meer medewerkers. Cijfers voor midden- en kleinbedrijf kunnen dus aanzienlijk afwijken (zie bijvoorbeeld Dany & Torchy, 1994).

** Ryan et al. maken onderscheid tussen one-to-one interviews (3.8) en group/panel interviews (4.3). In Tabel 1 is uitgegaan van de waarde van group/panel interviews.

*** Ryan et al. maken geen onderscheid tussen personality questionnaires en meer specifieke persoonlijkheidsvragenlijsten, zoals integriteitsvragenlijsten. In de tabel is de algemene waarde voor persoonlijkheidsvragenlijsten opgenomen.

**** Ryan et al. hanteren de term educational qualifications, deze term slaat op de hoogte van de genoten opleiding.

Uit tabel 1 valt af te lezen dat de sterkste predictor voor werkprestaties¹ de algemene intelligentietest is. De arbeidsproef scoort weliswaar in absolute zin iets hoger, maar levert een lager rendement vanwege de hoge ontwikkelings- en uitvoeringskosten. Andere voorspellers die nog substantieel iets toevoegen naast gebruik van een intelligentietest zijn het gestructureerde interview en consciëntieusheids- en integriteitsvragenlijsten.

We concentreren ons op die predictoren waarop sollicitanten door oefening/coaching of door herhaalde afname in principe in staat zijn een hogere (intelligentietest) of meer wenselijke score (vragenlijsten) tot stand te brengen. Bovendien moeten die predictoren zijn die relevant zijn voor de Nederlandse situatie. We zullen van bovenstaande tabel achtereenvolgens de coachingseffecten bespreken voor intelligentietests, interviews, integriteits- en persoonlijkheidsvragenlijsten en de assessmentcenterprocedure (ACM).

Ten aanzien van de andere voorspellers zijn nauwelijks of geen coachingsresultaten bekend, of coaching is niet relevant, zoals bij referenties. Ten aanzien van biografische gegevens, die kunnen worden ontleend aan een CV, geldt dat de absolute voorspellende waarde vrij hoog is en dat coaching, of liever gezegd faking, voor de sollicitant dus rendabel lijkt. Er zijn tegenwoordig echter gespecialiseerde bureaus die voor een paar honderd euro CV's op waarheidsgehalte toetsen. Door een CV te faken neemt de sollicitant dus een aanzienlijk risico. Een bespreking van het faken van een CV valt buiten het bestek van dit hoofdstuk.

¹ Officieel gaat het om prestaties in werktrainingen, maar het verschil met werkprestaties is minimaal, al kunnen werktrainingsprestaties door een wat hogere betrouwbaarheid, ook tot hogere predictieve validiteitscoëfficiënten leiden.

6 De intelligentietest: voorbereidingseffecten

De intelligentietest is verreweg de belangrijkste predictor voor arbeidsprestaties. Trouwens, voor alle cognitieve prestaties, ook die in het dagelijks leven, zoals het kunnen lezen van een medicijnenetiket (zie bijvoorbeeld Gottfredson 1997; Gottfredson & Deary, 2003; Gordon, 1997).

Onderzoek naar de effecten van coaching/training op intelligentietestscores is vooral gedaan in de Verenigde Staten, met name ten aanzien van de Scholastic Aptitude test (SAT) en de Graduate Record Examination (GRE), twee tests die als entreeselectiemiddel bij colleges en universiteiten standaard worden gebruikt in de VS. Er hangt veel van af om goed te scoren op deze tests omdat zij toegang geven tot het hoger onderwijs. De SAT en de GRE zijn geen intelligentietests pur sang, in de zin dat het alleen om het meten van de 'g'-factor (algemene intelligentie) gaat. Beide tests bevatten ook 'achievement'-elementen, gericht op het in kaart brengen van verworven kennis (vergelijk de CITO-toets). Dit aspect heeft betrekking op wat wel 'crystallized intelligence' genoemd wordt (Carroll, 1993).

Er zijn inmiddels verschillende overzichtsstudies gepubliceerd wat betreft de effecten van coaching op de SAT en de GRE en *aptitude* en *achievement* tests in het algemeen (Bangert-Drowns, Kulik & Kulik, 1983; Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik & Bangert, 1984; Salomon, 1985; Powers, 1986; Sackett, Burris & Ryan, 1989; Ryan, Ployhart, Greguras & Schmitt, 1998; Allalouf & Ben-Shakar, 1998; Powers & Rock, 1999; Hausknecht, Trevor & Farr, 2002; Lievens, Buyse & Sackett, 2005).

De resultaten van bovengenoemde studies wijzen in het algemeen op een klein tot middelgroot effect van oefening/coaching bij intelligentietests (effectscores tot .5 SD). Meestal gaat het daarbij niet om onderzoek bij personeelsselectie, maar om selectie voor onderwijsdoeleinden. Hausknecht, Trevor en Farr onderzochten vooral het effect van het herhaald afnemen van een intelligentietest in een selectiecontext. Zij vonden een effectscore tussen de eerste en de derde afname van de intelligentietest van .76 SD. Dit is een groot effect. Het betekent dat een kandidaat die bij test 1 in het 50e percentiel scoort, bij test 3 in het 75e percentiel komt (bij gelijke gemiddelden van de scoredistributie). Deze verschuiving kan grote gevolgen hebben voor de selectiebeslissing. Deze resultaten stemmen overeen met die uit de studie van Kulik, Kulik en Bangert uit 1984. Zij vonden een effect van .96 SD van de eerste naar de vierde testafname. Uit de studie van Kulik e.a kwam nog een ander interessant gegeven naar voren: mensen die hoog scoren op een intelligentietest profiteren het meeste van herhaalde afname. Hoog intelligente mensen gingen .80 SD vooruit bij de tweede afname van dezelfde test, gemiddeld intelligente mensen .40 SD en laag intelligente mensen .17 SD. In IQ-scores uitgedrukt ($M=100$, $SD=15$) zou dit betekenen dat hoog intelligente mensen hun IQ-score bij een tweede testafname met 12 IQ-punten kunnen verhogen, gemiddeld intelligente mensen met 6 IQ-punten en laag intelligente mensen met 3 IQ-punten.

In Nederland is een studie uitgevoerd naar het effect van testcoaching, door Van der Molen, Keen en te Nijenhuis (1995). Zij onderzochten het effect van het lezen van een voorbereidingsboek en coaching op de NAT (Drenth en Hoolwerf, 1970) en de VAT (Drenth en Van Wieringen, 1969), twee subtests uit de bij personeelsselectie veelgebruikte Drenth serie.

Er werden in deze studie geen significante test-hertesteffecten gevonden, maar wel een duidelijk oefeningseffect (niets doen, boek lezen of training). Het lezen van het boek leidde tot een significante scoretoename en het volgen van de training leidde tot een nog grotere scorewinst (tot 1 SD op de NAT). Dit experiment maakt vooral duidelijk dat een specifieke en gedetailleerde training die gebruikmaakt van een heuristiek, effecten kan hebben. Dat werkt echter alleen bij tests met een duidelijke cognitieve structuur, zoals de NAT (zie ook: Drasgow, 2003).

Concluderend kunnen we stellen dat dat het effect van coaching en herhaalde testafname bij intelligentietests gemiddeld tot groot is, met een toename in score tot maximaal 1 SD. Van een effect van 1 SD mag worden aangenomen dat daardoor de predictieve validiteit van de procedure wordt beïnvloed. Conclusies als dat de validiteit ongevoelig zou zijn voor coachingseffecten, terwijl er wel meer false positives zijn (en bij 1SD kan dat haast niet anders), impliceren dat r geen goede indicator is voor de predictieve validiteit. Daarbij moet wel de kanttekening gemaakt worden dat de coachingsmethoden over de verschillende studies heen nogal varieerden, van kort tot lang en van minder tot meer intensief. Ook het gehanteerde criterium: vooruitgang op een achievementtest versus een 'echte' intelligentietest (g) kan verschil maken. Daarbij kunnen de resultaten van studies waarbij studenten als proefpersoon worden ingezet afwijken van resultaten gevonden in de A&O-sfeer. De motivatie van 'echte' sollicitanten kan hoger zijn en dus samengaan met een verhoogd score-effect (Clause, Delbridge, Schmitt, Chan & Jennings, 2001).

Hoe een maximale scoretoename het beste en op de meest efficiënte wijze bereikt kan worden is nog niet duidelijk. Volgens Te Nijenhuis, Voskuil & Schijve (2001) wordt het maximumeffect van coaching na een paar uur bereikt. Maar uit de studie van Bangert-Drowns, Kulik & Kulik (1983) naar de effecten van coaching op een 'achievement test' bleek dat herhaalde oefening een verhogend effect had, vooral intensieve oefening gericht op het verbeteren van algemene cognitieve vaardigheden. De effecten hingen direct samen met het aantal uren genoten coaching. Een extreem 'drill'-experiment wordt beschreven door Stankov (1986). In dit experiment ondergingen Joegoslavische studenten gedurende drie jaar een intensieve training gericht op het verbeteren van cognitieve vaardigheden. Na zo ongeveer 500 uur getraind te zijn was de scorewinst 8 IQ-punten (.55 SD, gemeten met de Raven).

Het is dus onduidelijk wat beter werkt: een 'drill'-effect van herhaald, langdurig oefenen of een korte intensieve coachingscursus gericht op inzicht, of een combinatie van beide. Gezien het feit dat het *world wide web* een overvloed aan 'drill'-mogelijkheden geeft, kan de moderne sollicitant op efficiënte wijze een maximale scorewinst boeken op intelligentietests. Dit gegeven zou het zogenaamde Flynn-effect, het verschijnsel dat de gemiddelde scores op intelligentietests per generatie zo'n 3-5 punten vooruitgaan, nog kunnen versterken. Door efficiëntere en effectievere communicatie stijgt het rendement van de 'testwiseness' van de moderne sollicitant.

7 De Assessment Center Methode (ACM)

Van de lijst met selectie-instrumenten is het assessment center, officieel de Assessment Center Methode (ACM) de meest recente ontwikkeling. Ontwikkeld door verschillende militaire inlichtingendiensten tijdens de tweede wereldoorlog is de ACM tegenwoordig ook in Nederland gemeengoed geworden, zij het vooral als selectie- en ontwikkelingsmiddel voor de hogere managementfuncties (Tabel 1). In het geval van ontwikkeling spreekt men wel van een Development Center Methode (DCM). De term assessment betekent zoveel als 'het in kaart brengen van' en nodigt dus uit tot verwarring. Zowel tests, als interviews, vragenlijsten, opdrachten en casussen kunnen deel uitmaken van een ACM. Wat eigen is aan een ACM is eigenlijk het observeren van het spontane gedrag van kandidaten, bijvoorbeeld door middel van een rollenspel, presentatie of opdracht, zoals de In Basket. Wat de ACM dus feitelijk toevoegt aan een conventionele selectieprocedure is het observeren van het gedrag van de kandidaat in functierelevante oefeningen. Dit kunnen individuele oefeningen zijn (tweegesprek) of groepsoefeningen, zoals de Leiderloze groepsdiscussie (LGD). Gezien de directe koppeling met functierelevant gedrag heeft de ACM een hoge face validity, wat haar populair maakt bij de kandidaten. De predictieve validiteit van de Overall Assessment Rating (OAR) is heel behoorlijk, maar het rendement is relatief laag vanwege de hoge kosten die met de ACM verbonden zijn. Uit tabel 1 blijkt bovendien dat de incrementele validiteit boven een intelligentietest minimaal is. Bekendste probleem wat betreft validiteit is het verschijnsel dat dezelfde dimensies binnen verschillende simulaties lager met elkaar samenhangen dan verschillende dimensies binnen één simulatie, reden dat ook wel gesproken wordt van een overall situationeel effect (exercise effect) in plaats van dimensionele effecten (Zie bijv. Thornton & Byham, 1982; Gaugler et al., 1987; Lievens & Klimoski, 2001; Lievens, 2002).

Studies van Lance et al. (2000, 2004) suggereren echter een radicale *gestaltshift* wat betreft de onderbouwing van de werking van de ACM. Zij betogen dat de exercise effects wel degelijk criteriumgerelateerd zijn en dat de ACM dus niet zozeer dimensies meet maar de prestatie op functiegerelateerde taken. In hun ogen kan het ACM het beste gezien worden als een oefening waarbij de kandidaten in korte tijd functierelevant gedrag moet laten zien (*maximum* in plaats van *typical* behavior). Voor trainingseffecten zou dit vanuit het fairness principe betekenen dat (entree)kandidaten van te voren uitgebreid gebriefd moeten worden over relevante functieaspecten, daar anders kandidaten met meer functie-kennis in het voordeel zouden zijn. Voorbereiding voor een ACM wordt dan een sine qua non en bij studies naar effecten zal moeten worden gecontroleerd voor relevante werkervaring.

7.1 EFFECTEN VAN COACHING EN OEFENING OP ACM-SIMULATIES

Gezien het feit dat de ACM de jongste loot vormt binnen het palet van selectiemiddelen en instrumenten, is er nog weinig onderzoek gedaan naar effecten van daarop gerichte oefening en coaching.

We zullen eerst een aantal algemene oefeneffecten bespreken en dan een aantal effecten voor specifieke onderdelen: de In Basket en leiderloze groepsdiscussie.

Sackett et al. (1989) noemen bij gebrek aan empirisch onderzoek naar coachings- en oefeningseffecten het zogenaamde 'grapevine coaching effect'. Daar- onder verstaan zij het verschijnsel dat binnen organisaties kandidaten die een assessment hebben ondergaan informatie hierover doorspelen aan nieuwe kandidaten. Dit valt onder informele coaching die ook kan optreden bij andere selectiemiddelen (zie ook Conlon, 2004).

Tegenwoordig kan via het web, met name in discussiefora, dit verschijnsel zich ook buiten en tussen organisaties voordoen. Ook resultaten ten aanzien van het voorkomen en de effecten van grapevine coaching zijn echter onduide- lijk (Dodd, 1977; Cohen, 1978; Dulewicz, Fletcher & Wood, 1983).

In een recente studie (2003) onderzochten Kolk, Born en Van der Flier het ef- fect van het openbaar maken van de onderzochte dimensies op de prestaties van kandidaten (studenten en sollicitanten) in een ACM. Het openbaar maken van dimensies heeft een aantal voordelen, vooral vanuit het perspectief van *fairness*: alle kandidaten beschikken over dezelfde informatie betreffende de op- drachten en kunnen dus in dezelfde mate dimensierelevant gedrag vertonen. Bovendien wordt de kans vergroot dat ook daadwerkelijk dimensierelevant ge- drag vertoond wordt: de kandidaten weten nu immers welk gedrag van hen verwacht wordt.

De conclusie was dat transparantie bij de ACM tot meer consistentie in ge- drag op de gebruikte dimensies lijkt te leiden, maar dat alleen transparantie de gemiddelde prestaties van kandidaten niet beïnvloedt. Feedback op vertoond gedrag zou eventueel wel tot een gedragsverandering kunnen leiden.

7.2 IMPRESSION MANAGEMENT TACTIEKEN BIJ DE ACM

In alle situaties van een selectieprocedure waarbij het gaat om 'typical behavior' zullen gemotiveerde kandidaten proberen om een zo gunstig mogelijk beeld van zichzelf neer te zetten. Dat geldt zelfs voor het curriculum vitae, dat tegen- woordig steeds vaker op waarheid wordt onderzocht en getoetst. IM kan een probleem worden als het beeld dat de kandidaat van zichzelf schept te veel af- wijkt van zijn of haar daadwerkelijke capaciteiten, vaardigheden en of kennis (Knowledge, Skills, Abilities & Other characteristics). Beoordelingen op deze door IM opgeblazen aspecten hebben een positieve bias en de predictieve validiteit van selectieprocedures kan hierdoor negatief worden beïnvloed.

McFarland et al. (2003, 2005) onderzochten het voorkomen en het effect van Impression Managementtechnieken (IM) bij verschillende assessmentonderde- len. De verwachting was dat de aard van de assessmentsimulatie (meer tech- nisch of meer sociaal gericht) van invloed zou zijn op het optreden en het effect van IM.

McFarland et al. onderscheiden verschillende vormen van IM: verbale tactie- ken, waaronder assertieve en defensieve tactieken; non-verbale tactieken, zoals gebruik van oogcontact, glimlachen, hand- en hoofdgebaren (voor een uitge- breid overzicht, zie McFarland et al., 2005, p. 953-954).

In eerste instantie keken McFarland et al. (2003) naar het gebruik van verba- le IM-tactieken in twee verschillende ACM-simulaties: een rollenspel en een in- terview. Het bleek dat, in lijn met de hypothese, kandidaten in het situationele interview meer IM-tactieken gebruikten dan in het rollenspel. In beide simula- ties gebruikten de kandidaten meer tactieken gericht op directe beïnvloeding

van de ander, dan defensieve tactieken. Wat betreft de invloed van het gebruik van IM-tactieken op de uitkomsten van de simulaties bleek dat gebruik van IM in het interview wel samenging met betere beoordelingen en betere promotie-scores maar in het rollenspel niet.

In hun studie van 2005 gingen de onderzoekers meer verfijnd te werk: het gebruik en het effect van IM werden onderzocht in samenhang met de inhoud van de simulatie (meer technisch, zoals een presentatie, of meer sociaal, zoals een rollenspel). In een eerste experiment vonden McFarland et al. minder gebruik en minder spreiding van verbale IM-tactieken in een tactische assessmentopdracht dan tijdens een rollenspel. Gebruik van verbale IM tactieken had echter in geen van beide simulaties invloed op de ACM-beoordelingen van kandidaten.

Daarnaast onderzochten McFarland et al. in een tweede experiment ook het effect van non-verbale IM. De kandidaten werden in dit tweede experiment van tevoren geïnformeerd over de dimensies waarop zij beoordeeld zouden worden. In dit tweede experiment werd in de meer sociale ACM-opdrachten (rollenspel) een gemiddeld effect voor non-verbale tactieken gevonden van .71 ten opzichte van het gebruik van non-verbale IM-tactieken in een tactische simulatie.

Tabel 2: *Verskil in effect van IM-tactieken in sociale versus tactische ACM-simulaties*

IM-tactiek	Gemiddelde effectscore (d waarde)
direct gericht op beïnvloeden ander	1.29
zelfpromotie	.62
defensief	.41
non-verbaal	.71

Verder bleek dat in de presentatie minder gebruik werd gemaakt van zelfpromotie en defensieve tactieken dan in de rollenspellen, maar juist meer van directe beïnvloeding van de ander en non-verbale tactieken.

In lijn met de uitkomsten van het eerste experiment bleek dat gebruik van verbale en non-verbale IM geen invloed had op het assessoroordeel van de rollenspellen en de tactische opdracht. Echter, gebruik van verbale en non-verbale IM had wel invloed op het assessoroordeel van de presentatie. Gebruik van defensieve IM-tactieken had hier een negatieve invloed op het assessoroordeel. Het effect van het gebruik van IM-tactieken op de assessorbeoordelingen bij de presentatie was voor de onderzoekers onverwacht: men had juist een effect van het gebruik van IM verwacht op het assessoroordeel van de rollenspellen.

De auteurs concluderen dat de aard van de competenties die opgeroepen worden tijdens een simulatie (technisch versus sociaal) van invloed is op het voorkomen van IM. Verder doen zij de suggestie dat ACM-opdrachten moeten worden herzien vanuit het oogpunt welke competenties door een simulatie bij de kandidaat worden opgeroepen. De presentatie, ontworpen vanuit het oogpunt om technische kennis en vaardigheden bij kandidaten te beoordelen, gaf een onverwacht hoog gebruik van IM te zien en bovendien was dit de simulatie waarin gebruik van IM het sterkste samenhing met het oordeel van de assessor.

7.3 COACHINGSEFFECTEN OP PRESTATIES TIJDENS DE LEIDERLOZE GROEPSDISCUSSIE (LGD)

De LGD maakt regelmatig onderdeel uit van een ACM. Een klein groepje kandidaten wordt bij elkaar gezet met een gegeven (controversieel) discussieonderwerp. Opdracht kan zijn om bijvoorbeeld in maximaal twintig minuten een consensus over het onderwerp te verkrijgen. Het gedrag van de deelnemers wordt geobserveerd en beoordeeld op bijvoorbeeld leiderschapskwaliteiten en sociale sensitiviteit. Recente studies over coachingseffecten op prestaties tijdens de LGD zijn er niet. Sackett et al. (1989) bespreken een aantal (oudere) studie-uitkomsten. Over het algemeen zijn de gevonden trainingseffecten ten aanzien van prestaties tijdens de LGD niet overtuigend. Inzet als assessor blijkt geen invloed te hebben op LGD-prestaties. In een onderzoek van Petty (1974), bleek dat het gevolgd hebben van de LGD-presentatie, waarin deelnemers werd verteld wat er van ze verwacht werd en hoe ze hun gedrag tijdens de LGD konden maximaliseren (dus een simulatiespecifieke training), een significant effect had op alle criteriummaten. Dit in tegenstelling tot resultaten van eerdere studies waarin overigens niet zozeer getraind was op simulatiespecifieke kenmerken, maar op algemene dimensies van leiderschapsgedrag. De resultaten van een studie van Denning en Grant (1979) sluiten aan bij de onderzoeksresultaten van Petty. Denning en Grant vonden geen effect op beoordelingen van gedrag van kandidaten die van tevoren waren ingelicht over de gebruikte beoordelingsdimensies, vergeleken met het gedrag van kandidaten die werden voorgelicht over algemene kenmerken van de ACM.

Een studie van McIntyre (1980) bevestigt nog eens het belang van feedback voor gedragsverandering. In zijn studie ontvingen kandidaten zowel schriftelijke als mondelinge feedback over hun gedrag tijdens een eerste LGD. Tijdens een tweede LGD scoorden de kandidaten op zeven van de twaalf beoordelingscriteria significant hoger. Opvallend was verder dat de stabiele gedragsaspecten geen verandering vertoonden, terwijl situationele gedragingen zoals probleemanalyse en management control de meeste verbetering vertoonden. In totaal ging 67% van alle beoordelingen na de tweede LGD omhoog.

Kureca et al. (1982) voerden een experiment uit met drie groepen. Een groep die een verslag te horen kreeg van een persoonlijke en gekleurde LGD-ervaring (surrogaat voor *grapevine coaching*); een groep die voorbeelden kreeg van effectief gedrag en feedback op een daadwerkelijke LGD-ervaring en een controlegroep die alleen een verslag kreeg van de geschiedenis van de LGD. De groep met de meest volledige coaching deed het significant beter dan de twee andere groepen, die qua resultaten gelijk scoorden.

Sackett et al. concluderen dat wat betreft coaching voor de LGD het geven van feedback essentieel lijkt om een experimenteel effect te bereiken. Dit sluit aan bij de eerdere uitkomsten van bijvoorbeeld sociale-vaardigheidstraining. Algemene informatie over de werking van een LGD heeft nauwelijks of geen effect. Slechts simulatiespecifieke informatie, gecombineerd met feedback over het vertoonde gedrag, kan de prestaties van deelnemers in een LGD verbeteren.

7.4 COACHINGSEFFECTEN OP DE IN BASKET OPDRACHT (IBO)

Tijdens de In Basket Opdracht (IBO) moeten kandidaten de rol van manager vervullen. Zij krijgen het postbakje van een (afwezige) manager met de opdracht

om dit zo effectief mogelijk af te wikkelen. Gedrag wordt onder andere beoordeeld op analyse, planning en het kunnen stellen van prioriteiten. Uit het weinige onderzoek dat naar coachingseffecten met betrekking tot de IBO gedaan is blijkt dat zolang er geen adequate gedragsfeedback gegeven wordt, coaching weinig of geen effect heeft. In die zin lijkt zich wat betreft de mogelijkheden om gedrag tijdens een ACM te verbeteren de tendens af te tekenen dat het geven van persoonlijke feedback aan de deelnemers een voorwaarde sine qua non is om gedrag tijdens ACM simulaties te verbeteren. Specifieke ervaring kan soms van invloed zijn op ACM beoordelingen, zoals in het geval van een planningssimulatie, maar ervaring speelde weer geen rol voor gedrag tijdens het houden van een presentatie of het schrijven van een brief (Dulewicz & Fletcher, 1982).

Concluderend kunnen we stellen dat de effecten van oriëntatie, training en coaching voor het verbeteren van gedrag in ACM-simulaties nogal variëren. Wel moet de kanttekening gemaakt worden dat er weinig onderzoek gedaan is en dat het onderzoek dat gedaan is, nogal wat methodologische manco's vertoont zoals kleine onderzoeksgroepen, vaak bestaande uit studenten, wat van invloed kan zijn op de motivatie om goed te presteren. Een absolute voorwaarde om gedrag in een ACM te verbeteren lijkt het verkrijgen van persoonlijke en simulatiespecifieke feedback zodat een kandidaat weet welk specifiek gedrag van hem of haar verwacht wordt in een simulatie en hoe dit gewenste gedrag zich verhoudt tot zijn of haar 'natuurlijke' (base rate) gedrag. Gebruik van IM-tactieken kan daarbij van nut zijn, vooral als dat gedrag gericht is op directe beïnvloeding van de assessor of gericht op het scheppen van een positief beeld van zichzelf. Defensief gedrag dient zoveel mogelijk vermeden te worden. Is aan deze voorwaarden voldaan, dan kan gesteld worden dat, analoog aan de effecten gevonden in sociale vaardigheidstrainingen (Van de Molen et al., 1995; Schönrock, 2002), maximale gedragsveranderingseffecten van ongeveer 1 SD op assessment oefeningen gericht op het in kaart brengen van sociale competenties haalbaar zijn.

8 Het selectie-interview

In bijna alle sollicitatieprocedures krijgt de sollicitant te maken met een interview (Tabel 1). Hoewel het interview meestal ook deel uitmaakt van een ACM, rechtvaardigt de standaardrol van het interview in selectieprocedures een aparte bespreking. De voorspellende waarde van het interview is heel redelijk (zie Tabel 1). Vooral als het interview in gedragsgerichte en gestructureerde vorm wordt afgenomen is de voorspellende waarde goed te noemen. Kan door training of coaching de prestatie tijdens een interview verhoogd worden? En zo ja, leidt dit tot aantasting van de predictieve validiteit?

Er worden verschillende soorten interviews onderscheiden. Het meest bekende onderscheid is dat tussen het gestructureerde interview en het ongestructureerde interview. Naast dit onderscheid worden ook wel situationele interviews en gedragsgerichte interviews onderscheiden (voor een overzicht van interviews, wat ze meten en hun voorspellende waarde zie bijvoorbeeld McDaniel, Whetzel, Schmidt & Maurer, 1994; Huffcut, Conway, Roth & Stone, 2001; Eder & Harris, 1999; Salgado & Moscoso, 2002).

8.1 WAT WORDT ER BEOORDEELD IN SELECTIE-INTERVIEWS?

Huffcutt, Conway, Roth, Stone (2001) concluderen dat de meeste interviews gericht zijn op het in kaart brengen van sociale vaardigheden en persoonlijkheidseigenschappen (deze twee omvatten meer dan 60% van alle beoordeelde attributen). De interviewers proberen voornamelijk af te leiden hoe de geïnterviewden zich zullen gedragen tijdens het werk en hoe zij zullen of kunnen samenwerken met toekomstige collega's. Wat betreft persoonlijkheidseigenschappen ligt de nadruk op vragen naar consciëntieusheid of varianten daarvan (verantwoordelijkheid, betrouwbaarheid, initiatief en volharding).

Vragen naar cognitieve capaciteiten, kennis en vaardigheden zijn verantwoordelijk voor 25% van de beoordeelde attributen. De resterende 15% gaat over interesses, person-organization fit en fysieke kenmerken.

Een belangrijk onderscheid dat Huffcutt et al. aantreffen was dat de mate van gestructureerdheid van het interview samenhang vertoonde met de beoordeelde interviewonderwerpen. Interviews met een lage structuur richtten zich meestal op constructen als cognitieve capaciteiten, opleiding en training, ervaring en interesse. Interviews met een hoge structuur richtten zich meer op constructen als relevante werkkennis en vaardigheden, person-organization fit, sociale vaardigheden en probleemoplossende en beslissingsvaardigheid. Volgens Huffcutt et al. is dit onderscheid gebaseerd op het feit dat gestructureerde interviews meer gebaseerd worden op een formele taakanalyse. Dit onderscheid is in overeenstemming met de gegevens uit tabel 1, waaruit blijkt dat gestructureerde interviews een grotere (incrementele) predictieve validiteit hebben ten aanzien van werkprestaties: er wordt gepoogd constructen te inventariseren die (onafhankelijk van intelligentie) een grotere samenhang hebben met arbeidsprestaties.

Salgado en Moscoso (2002) maken wat betreft interviews onderscheid tussen conventionele interviews en gedragsgerichte (situationele) interviews. Deze indeling komt in redelijke mate overeen met die van ongestructureerd versus gestructureerd.

Uit hun bevindingen blijkt dat conventionele interviews vooral gericht zijn op het in kaart brengen van intelligentie, emotionele stabiliteit, extraversie, openheid voor ervaring en sociale vaardigheden. Gedragsgerichte interviews focussen op werkervaring, functiekennis, beoordelingsvermogen en sociale vaardigheden. In grote lijnen komen deze resultaten overeen met die van Huffcutt et al. (2001). Ongestructureerde, conventionele interviews lijken qua aanpak en inhoud inderdaad andere interviews dan gestructureerde gedragsgerichte interviews (zie ook Roth, van Iddekinge, Huffcutt & Eidson jr., 2005). In de eerste ligt de nadruk meer op intelligentie en persoonlijkheidsaspecten, in de tweede gaat het vooral om relevante functiekennis. De enige overeenkomst is dat in beide interviews substantieel gevraagd wordt naar en beoordeeld wordt op sociale vaardigheden (zie ook: Moscoso, 2000). Dit heeft waarschijnlijk te maken met de aard van de sociale interactie: de Big Five dimensies vriendelijkheid en extraversie spelen daarbij een prominente rol (Van Dam, 2003).

8.2 SELECTIE-INTERVIEWS EN IM

In het interview staat beïnvloeding centraal. Of, zoals Eder en Harris (1999) stellen: 'When the interviewer is trying to evaluate objectively, the applicant is trying to sell; and when the applicant is trying to evaluate objectively the interviewer is trying to sell' (Eder en Harris, 1999, p. 346). Het gaat dus om een wederzijds beïnvloedings- en beoordelingsproces en de vraag is of de sollicitant in staat is door training en oefening dit beïnvloedings- en beoordelingsproces in zijn voordeel te benutten.

De al eerder genoemde IM-tactieken zullen vooral in het interview een belangrijke rol spelen. Dat zal met name het geval zijn als het interview zich richt op 'soft constructs', zoals sociale vaardigheden en persoonlijkheidseigenschappen. McFarland, Ryan en Kriska (2002) onderzochten het algemene gebruik van beïnvloedingstechnieken door brandweerlieden tijdens sollicitatiegesprekken. Het bleek dat in de interviews meer softe dan harde beïnvloedingstechnieken werden gebruikt (innemendheid en overtuigingskracht tegenover druk uitoefenen of een verbond aangaan). Ook bleek het gebruik van softe technieken samen te gaan met een betere beoordeling van de interviewprestaties. Ellis, West, Ryan en DeShon (2002) verfijnden dit onderzoek nog door bij gestructureerde interviews te onderzoeken of het stellen van specifieke vragen invloed had op het gebruik van specifieke IM-tactieken. Het kan immers zijn dat specifieke vragen een beroep doen op een specifiek soort IM bij de kandidaat. Bij situationele vragen werd meer gebruikgemaakt van innemendheid, maar bij vragen naar werkervaring werd vooral zelfpromotie gebruikt. Zowel het gebruik van innemendheid als dat van zelfpromotie beïnvloedde de evaluatie van de interviewuitkomsten in positieve zin. Het bleek dat bijna alle onderzochte kandidaten (n = 119) wel op een of andere manier gebruikmaakten van IM en dan vooral van assertieve zelfpromotie en IM-tactieken.

Defensieve tactieken werden veel minder gebruikt.

Stevens en Kristof (1995) komen na een overzicht van IM-tactieken en een vrij uitgebreide studie, waarin zij observaties ten aanzien van IM van sollicitanten, interviewers en observatoren met elkaar vergeleken, tot ongeveer dezelfde conclusie. Ook zij vonden dat het voorkomen van IM samenging met betere interviewbeoordelingen. Zij vonden ook dat kandidaten in het interview veel meer gebruikmaakten van assertieve tactieken dan van defensieve tactieken. Binnen de assertieve tactieken werd zelf-promotie weer meer gebruikt dan innemendheid. Wel werpt hun studie vraagstekens op naar de constructvaliditeit van de gebruikte IM-tactieken: de drie groepen waren het niet altijd eens over de betekenis van een gebruikte tactiek. Dat kan komen door een verschil in perspectief op de gebruikte tactiek: sollicitanten hebben een perspectief van binnenuit, de intentie staat voorop. Interviewers hebben een perspectief van buiten en zijn het doel van de gebruikte IM-tactiek; observatoren hebben weer een ander perspectief en missen de visuele cues die interviewers wel zien bij de sollicitant.

Het klinkt misschien contra-intuïtief, maar ook wat betreft meer 'harde maten', zoals de vraag naar cognitieve capaciteiten, heeft de geïnterviewde in het interview de mogelijkheid om via IM-tactieken zijn 'waargenomen' intelligentie op te poetsen. Uit een studie van Caldwell & Burger (1998) bleek bijvoor-

beeld dat (verhoogde) extraversie samenhang met een meer positieve indruk wat betreft intelligentie van de geïnterviewde.

8.3 EFFECTEN VAN COACHING OP SELECTIE-INTERVIEWUITKOMSTEN

In een overzicht naar de effecten van coaching op interviewprestaties komen Sackett et al. (1989) tot de conclusie dat er sprake is van behoorlijke effecten, ook al gaan veel studies mank door te kleine steekproeven en een onvolledige experimentele opzet. Er kan dus geen uitsluitel gegeven worden over een generaliseerd of 'waar' effect van interviewcoaching. Ook lijkt feedback een belangrijke voorwaarde om een substantieel trainingseffect tot stand te brengen. Sackett et al. (ibid) vonden geen verschil in coachingseffecten wat betreft de structuur, aard en doel van het interview.

Maurer, Solamon en Troxtel (1998) onderzochten het effect van coaching op prestaties in een realistisch gestructureerd situationeel interview waarbij zij controleerden voor motivatie en functie-kennis. De coaching bestond uit een uitgebreide voorlichting over de werking en inhoud van het interview. Relevante onderwerpen werden besproken en er werden tips gegeven hoe zich te gedragen in het interview. Ook werd er van tevoren geoefend middels rollenspelen waarbij uitgebreide feedback gegeven werd. De proefpersonen waren brandweerlieden en politiemensen die konden solliciteren op vier functies (n varieerde per groep van 49 tot 217). Het bleek dat coaching een significant effect had op de interviewresultaten bij drie van de vier onderzochte functies (gemiddelde $R^2 = .09$), onafhankelijk van motivatie en functie-kennis. Er was echter geen controlegroep dus er konden geen effectmaten (d) berekend worden. In een tweede onderzoek (Maurer, Solamon, Andrews en Troxtel, 2001) werd onderzocht hoe de kandidaten de coaching gebruikten om hun prestaties in het interview te verbeteren. Het bleek dat na het volgen van het coachingsprogramma de kandidaten in het interview een betere strategie gebruikten om hun antwoorden te organiseren en dat gebruik van zo'n strategie leidde tot betere interviewprestaties.

Van Iddekinge et al. (2005) vergeleken 'respons inflation' of 'faking' in een interview met response inflation op een persoonlijkheidsvragenlijst (NEO-PIR). Volgens Van Iddekinge et al. zal het voor sollicitanten lastiger zijn om een interview te faken dan een persoonlijkheidsvragenlijst. In een interview is de cognitieve belasting voor sollicitanten groter. Gefakete antwoorden moeten in overeenstemming zijn met eerder gegeven antwoorden en ook mag de interviewer niet merken dat de sollicitant de zaak bedriegt. Er is immers sprake van continue sociale interactie in een interview. Bovendien kan een gefakete antwoord weer aanleiding zijn tot nieuwe vragen, zodat de frauderende sollicitant verstrikt raakt in zijn eigen leugens. Faken veronderstelt naast 'automatic procession' een proces van 'controlled information processing'. Dit stelt extreme eisen aan de cognitieve belastbaarheid van een kandidaat (Bargh, 1994). Bovendien is er in een interview sprake van tijdsdruk, iets wat niet of veel minder speelt bij een persoonlijkheidsvragenlijst.

Uit de resultaten van de studie van Van Iddekinge et al. blijkt dat de instructie om te faken een zeer grote invloed kan hebben en dan met name op de uitkomsten van persoonlijkheidsvragenlijsten. Bij de NEO zelfbeoordeling versus de NEO ingevuld door een ander was het effect in de 'fake good' ten opzichte

van de 'eerlijke' groep negen keer zo groot. Ook bleek de instructie om te faken de factorstructuur van de vragenlijst te beïnvloeden: faken kan dus ook de constructvaliditeit van een vragenlijst aantasten. Wat overduidelijk bleek was dat persoonlijkheidsvragenlijsten veel gemakkelijker te manipuleren zijn dan interviewbeoordelingen.

Concluderend kunnen we stellen dat interviewresultaten verbeterd kunnen worden door gebruik te maken van IM, of door coaching. De effecten zijn wisselend, mede doordat de settings verschillen (veel experimenten met studenten) en kleine steekproeven. Wat betreft het onderscheid tussen conventionele interviews en situationeel (gedragsgerichte) gestructureerde interviews zijn conventionele interviews beter te manipuleren. Conventionele interviews focussen meer op persoonlijkheidsvariabelen, situationele interviews meer op werkrelevante constructen. Relevante kennis van de specifieke functie lijkt in het laatste geval een voorwaarde om een beter interviewresultaat te bereiken.

9 Persoonlijkheidsvragenlijsten

Persoonlijkheidsvragenlijsten hebben een matig voorspellende predictieve validiteit (Tabel 1). Voor de voorspelling van arbeidsprestaties lijken consciëntieusheid (en specifieke integriteitsschalen) samen met emotionele stabiliteit de grootste voorspellende waarde te hebben. Meestal worden persoonlijkheidsvragenlijsten afgenomen als een soort veiligheidsclausule (opsporen voorkeur voor extreme gedragingen) en als leidraad voor het interview. De moderne persoonlijkheidsvragenlijst is meestal een Big Five variant, zoals de NEO of de FFPI.

In hun overzichtsartikel uit 1989 concluderen Sackett et al. nog dat zij geen coachingsstudies met betrekking tot het gebruik van persoonlijkheidsvragenlijsten in A&O-settings aantreffen. De enige effecten die zij in de testboekjes aantreffen waren test-hertesteeffecten van maximaal .10 SD. Het betrof in alle gevallen laboratoriumstudies met studenten.

In het kader van 'coaching' is het wat vreemd om van coaching op een persoonlijkheidsvragenlijst te spreken. Coaching zou hier letterlijk inhouden dat men uitleg verschaft bij de items zodat de sollicitant de precieze betekenis van een item duidelijk wordt, waardoor een meer constructvalide scoring kan ontstaan.

In de praktijk worden persoonlijkheidsvragenlijsten afgenomen om het persoonlijkheidsprofiel van de sollicitant te vergelijken met de eisen die de functie stelt aan persoonlijkheid (gedrag). Het gaat dus om 'person-function fit' en in bredere zin om 'person-organisation fit'. Kennis van het vereiste profiel kan een voordeel zijn. Het is meestal niet erg ingewikkeld om een persoonlijkheidsvragenlijst zodanig in te vullen dat het gewenste profiel naar voren komt. Ook zijn er boeken die de sollicitant uitleg geven over sociale wenselijkheidsschalen en hoe persoonlijkheidsvragenlijsten werken qua structuur (Bloemers, 2005).

In zekere zin valt de mogelijkheid om een gewenst profiel op een persoonlijkheidsvragenlijst tot uitdrukking te laten komen onder IM. In een meer extreme situatie is er sprake van 'faking' of bedrog. De sollicitant veinst posities op onderliggende latente trekken die niet corresponderen met zijn of haar ware scores. In de inleiding van dit hoofdstuk is al gewezen op de bufferende werking van de procedure bij personeelsselectie: naast de vragenlijstuitkomst

sten heeft de selecteur ook informatie uit het interview, zodat de uitkomsten van beide met elkaar vergeleken kunnen worden.

Een meer realistische procedure zou zijn om de output van de vragenlijst te gebruiken als input voor het interview: zo weet de sollicitant meteen dat bedrog al snel kan terugslaan (backfiring).

9.1 FAKING OP PERSOONLIJKHEIDSVRAGENLIJSTEN

We zullen ons concentreren op enkele studie-uitkomsten naar de effecten van bedrog op persoonlijkheidsvragenlijsten. In de paragraaf hiervoor is al gewezen op de mogelijkheid dat bedrog bij persoonlijkheidsvragenlijsten, vooral als de expliciete instructie gegeven wordt om een bepaald profiel tot uitdrukking te brengen, grote tot zeer grote effecten kan opleveren (Van Iddekinge et al., 2005).

In het midden van de jaren negentig van de vorige eeuw ontstaan de eerste studies naar het 'faken', waarschijnlijk onder invloed van een toenemend gebruik van Big Five vragenlijsten. Een meta-analyse van Viswesvaran en Ones (1999) over 51 deelstudies wijst uit dat alle dimensies van de Big Five in dezelfde mate aan bedrog onderhevig kunnen zijn. De auteurs keken zowel naar effecten van een 'fake good', als van effecten van een 'fake bad'-instructie. De effectgroottes van zowel de within als de between subject designs waren heel behoorlijk, met een gemiddelde rond de .5 SD. In de fake good conditie gaven Emotionele stabiliteit, Consciëntieusheid en Openheid het grootste effect te zien (.76-.93SD). Extraversie en vriendelijkheid bleken het minst gevoelig voor faking instructions (.54 en .47 SD resp.). De grootste effecten werden gevonden voor de dimensie sociale wenselijkheid (2.26 SD). Ook blijkt het gemakkelijker om 'verkeerd' te bedriegen dan om 'goed' te bedriegen, in de 'fake bad'-conditie waren de effecten wel twee tot drie keer zo groot. Een kanttekening wat betreft de gevonden effecten in deze studie is het feit dat de proefpersonen expliciet de instructie kregen om te bedriegen. In een realistische A&O-situatie is dat niet het geval, daar is bedrog een individuele keuze. Het is dus nog de vraag in welke mate 'bedrog' bij persoonlijkheidsvragenlijsten zich voordoet in 'real life' selectiesituaties.

Griffin, Hesketh en Grayson (2004) concluderen dat transparante items makkelijker te faken zijn. Wat betreft de Big Five is de dimensie Openheid het minst gemakkelijk te faken en de dimensie consciëntieusheid het gemakkelijkst. Ook zijn er aanwijzingen dat mensen die hoger scoren op neuroticisme en lager op integriteit en consciëntieusheid in meerdere mate faken op persoonlijkheidsvragenlijsten. Verder moet ook het plafond effect niet vergeten worden: mensen die een 'ware score' hebben die vrij hoog ligt op de onderliggende latente trek, hebben minder ruimte om (goed) te faken.

Het is nog onduidelijk in welke mate sollicitanten nu echt faken bij vragenlijstonderzoek: de verschillende onderzoeksuitkomsten spreken elkaar tegen. Afhankelijk van het geloof in de eerlijkheid van de selectieprocedure kan het fakinggedrag van sollicitanten verschillen. Ook kan de neiging tot faking een persoonsgebonden eigenschap zijn waarbij intelligentie, sociale vaardigheden, de behoefte aan erkenning, religie, waarden en morele codes een rol spelen. Schmitt en Ryan (1993) suggereren dat faking de factorstructuur van een test kan aantasten. Normaal gesproken zou een Big Five vragenlijst bij factoranaly-

se vijf onafhankelijke factoren te zien moeten geven, maar in selectieonderzoek komt vaak een zesde algemene factor naar voren. Schmitt en Ryan noemen dit de 'ideal employee'-factor. Items van alle dimensies laden op deze factor, behalve items van de dimensie 'openheid'. Dit is dus een soort sociale wenselijkheidsfactor die vaak (maar niet altijd) bij persoonlijkheidsvragenlijsten, gebruikt in de selectiecontext, naar voren komt.

Het effect van faking op de predictieve en constructvaliditeit blijkt vooralsnog onduidelijk: ook hier spreken de verschillende onderzoeksresultaten elkaar tegen (ibid. p. 1547).

In hun eigen studie gaan Schmitt en Ryan in op de rol van twee Big Five facetten: Consciëntieusheid (gemakkelijkst te faken) en Openheid (moeilijkst te faken). Als instrument werd de NEO PI-R gebruikt (Costa & McCrae, 1992). Twee steekproeven werden gebruikt: studenten en laatste jaars studenten/sollicitanten. Het was de onderzoekers vooral te doen om differentiële item functioning (DIF) op te sporen. DIF is het verschijnsel dat verschillende mensen, die een gelijke score hebben op een onderliggende latente eigenschap, op eenzelfde vragenlijstitem verschillend antwoorden vanwege de invloed van een of meer andere factoren (Zickar en Robie, 1999). DIF kijkt dus op itemniveau en niet op schaalniveau. Als DIF op testniveau leidt tot differentiële predictie is er sprake van bias in de testresultaten. Dat kan bijvoorbeeld bij minderheden het geval zijn. Zij kunnen anders reageren op testitems waardoor hun score vertekend wordt (zie bijv. Te Nijenhuis en Van de Flier, 2000). Volgens de onderzoekers is er wat betreft de neiging tot faking op persoonlijkheidsvragenlijsten sprake van een continuüm: studenten scoren het laagst en sollicitanten het hoogst. Werknemers zullen ergens in het midden scoren. DIF bleek inderdaad op te treden, sollicitanten bleken zich positiever te omschrijven dan studenten wat betreft consciëntieusheid. De scores op Openheid kwamen overeen. Na controle voor geslacht en leeftijd werd een gemiddeld DIF-effect aangetroffen in vier van de zes facetschalen van Consciëntieusheid, en tevens werd een licht tot gemiddeld DIF-effect aangetroffen in alle zes facetten van Openheid.

De auteurs concluderen dat 'faking' de meeteigenschappen van een test beïnvloedt en daarom niet per dimensie moet worden onderzocht maar per item en per facet. Voorgaande studies die geen verschillen aantreffen op dimensieniveau hebben misschien verschillen op facetniveau genegeerd. Dit kan met name het geval zijn voor de dimensie Openheid, waarop sollicitanten sommige facetschalen interpreteren als negatieve indicatoren van werkprestaties. Dit kan ook verklaren waarom openheid in een aantal onderzoeken niet of negatief blijkt samen te hangen met arbeidsprestaties (Barrick, Mount & Judge, 2001; Griffin & Hesketh, 2004).

De auteurs adviseren om naast gebruik van persoonlijkheidsvragenlijsten ook biodata en assessmentcenter oefeningen te gebruiken om de persoonlijkheid van sollicitanten in kaart te brengen. Ze noemen het interview niet expliciet, maar aan te nemen valt dat dit onder de assessmentonderdelen valt. Zoals eerder betoogd lijkt het interview een zeer geschikte methode om de gegevens uit de persoonlijkheidsvragenlijst te valideren.

9.2 FAKING OP INTEGRITEITSVRAGENLIJSTEN

Alliger en Dwight (2000) onderzochten of integriteitstests in de zelfde mate fakable waren als algemene persoonlijkheidsvragenlijsten. Hoewel alle non-cognitieve meetinstrumenten gevoelig zijn voor bedrog kunnen er wel onderlinge verschillen zijn (vergelijk ook het verschil tussen de fakability van de verschillende Big Five factoren). Naarmate een test meer 'bedekte' vragen bevat is faken moeilijker, omdat de sollicitant dan moeilijk kan achterhalen wat het item precies meet. Een bekend voorbeeld is een item uit de ABV dat ongeveer als volgt luidt: *Ik rijd nooit door rood licht*. Bijna alle respondenten denken dat dit item zoets als 'correct gedrag' meet, maar het item is een onderdeel van de rigiditeitschaal.

Wat betreft onderlinge verschillen in fakability vond Stanush (1997) de volgende effectgroottes: Integriteitstests: 1.11; biodata: .94; interessevragenlijsten: .66; persoonlijkheidsvragenlijsten: .45.

Alliger en Dwight maken een onderscheid tussen overte en coverte integriteitstests. Overte integriteitsvragenlijsten vragen naar gedachten, gevoelens en verwachte gedragingen met betrekking tot eerlijkheid en bestraffing, en ook naar vertoond wangedrag. Coverte integriteitsvragenlijsten vragen naar algemene (onderliggende) psychologische constructen die correleren met contra-productief gedrag, zoals consciëntieusheid en impulscontrole (Alliger en Dwight, *ibid.* p 59-60). Eerdere studies wijzen erop dat response distortion (liegen) voorkomt bij sollicitanten, de percentages variëren van 15-45% van de sollicitanten¹. Alliger en Dwight onderschrijven dat het niet duidelijk is of faking op persoonlijkheidsvragenlijsten de predictieve validiteit aantast. Zij wijzen op een studie van Pannone (1984) die in eerste instantie een predictieve validiteit van een biografische vragenlijst vond van .45 (N=221). Nadat Pannone de groep had gesplitst in fakers (N=77) en non-fakers (N=144) vond hij predictieve validiteiten van respectievelijk .26 en .55. Een algemene predictieve validiteitcoëfficiënt hoeft dus geen goede indicatie van het voorkomen van fakegedrag op te leveren, deze kan grote intragroepsverschillen in fakegedrag maskeren.

Zelf onderzochten Alliger en Dwight drie condities: fake good versus geen specifieke instructie; coaching versus geen specifieke instructie en coaching versus fake good instructies. Het grootste effectverschil werd gevonden in de coaching versus geen specifieke instructiegroep. Over het algemeen waren de effecten op de overte integriteitsvragenlijst groter dan op de coverte lijst. Over alle drie de condities was er voor de overte lijst een gemiddeld effect van 1 SD, voor de coverte lijst .33 SD. Op de overte integriteitsvragenlijst was het effect van coaching versus geen specifieke instructie ruim 1.5 SD. Fake good versus geen specifieke instructie bereikte hier een verschil van ruim 1 SD, coaching versus fake good .45 SD. Daar in dit onderzoek gebruikgemaakt werd van een between-subject design kunnen de effecten bij gebruik van een within-subject design nog hoger uitvallen.

Gezien het feit dat overte en coverte integriteitsvragenlijsten ongeveer dezelfde predictieve validiteiten hebben met betrekking tot contraproductief ge-

¹ NB: veel van dit onderzoek bestaat eruit dat men bijvoorbeeld vraagt of een sollicitant ervaring heeft met een niet-bestaande taak. De sollicitant die hierop 'ja' antwoordt hoeft niet per se de intentie te hebben om te liegen, men kan ook te volgzzaam zijn en niet durven tegenspreken.

drag en arbeidsprestaties, adviseren de auteurs om binnen organisaties vooral op persoonlijkheid gebaseerde Integriteitsvragenlijsten in te zetten.

Concluderend kunnen we stellen dat de effecten van coaching en fake good instructies grote effecten kunnen hebben op de scores van persoonlijkheidsvragenlijsten in het algemeen en integriteitsvragenlijsten in het bijzonder.

Vooral vrij doorzichtige constructen, zoals consciëntieusheid, zijn goed te beïnvloeden.

Mensen zijn dus redelijk in staat om de vragen van een vragenlijst te doorzien, een koppeling te leggen naar de functie-eisen en vervolgens de vragen te scoren in een richting die op verhoogde geschiktheid voor de functie-eisen wijst (zie ook het onderzoek van Iddekinge et al., 2005). Het simpelweg waarschuwen van sollicitanten om niet te faken kan fakegedrag iets corrigeren. Dwight en Donovan (2003) vonden een remmend effect van waarschuwingen tegen faking op vragenlijstcores van rond de .20 SD. Gezien de (grote) effecten van faking, vooral in coachingcondities, lijkt het remmende effect van waarschuwingen beperkt. Het lijkt bovendien alleen effect te hebben als het gepaard gaat met de sanctie om bij twijfel over een (te) grote hoeveelheid sociale wenselijkheid de betrokken sollicitanten te hertesten (Elligson & Heggstad, 2003).

Het vermogen van sommige sollicitanten om testitems (persoonlijkheidseigenschappen) bij de invulling bewust te scoren in de richting van de gewenste functie-eisen kan ook anders worden uitgelegd (Schmitt & Ryan, 1993). De gemotiveerde vragenlijstinvuller behandelt de persoonlijkheidsvragenlijst als een intelligentietest met als doel zo hoog mogelijk te scoren op het door de werkgever vereiste profiel. Een 'lage' score op een persoonlijkheidsvragenlijst kan dan indicatief zijn voor het gegeven dat een sollicitant een gebrekkige kennis heeft van de functie-eigenschappen die relevant zijn voor de functie in kwestie. Of de sollicitant bij een hoge score ook overeenkomstig gedrag in de werksituatie zal vertonen is een andere vraag. Schmitt en Ryan vonden in hun experiment dat de validiteit van persoonlijkheidsvragenlijsten lager was voor hooggemotiveerde testkandidaten in tegenstelling tot die bij laaggemotiveerde kandidaten. Hun verklaring is dat laaggemotiveerde kandidaten de persoonlijkheidsvragenlijst 'echt' invullen, hooggemotiveerde kandidaten scheppen een te 'succesvol' profiel. Bij intelligentietests werden precies tegengestelde effecten gevonden.

Nog een belangrijke factor bij het bepalen van de effecten van faking op de voorspellende validiteit van persoonlijkheidsvragenlijsten is de ecologische validiteit van de experimenten. In een natuurlijke setting blijken de effecten kleiner dan in laboratoriumexperimenten. Hough en Ones (2001) concluderen dat in een natuurlijke setting de voorspellende validiteit niet of nauwelijks wordt beïnvloed door faking.

Onderzoek naar de effecten van coaching en faking op de voorspellende validiteit bij persoonlijkheidsvragenlijsten blijft echter actueel, ook al vanwege het feit dat integriteitsvragenlijsten een behoorlijke toegevoegde waarde blijken te hebben ten opzichte van cognitieve capaciteiten (Tabel 1). Het loont om een zo

nauwkeurig mogelijk beeld te hebben van de betrouwbaarheid en integriteit van een sollicitant.

Voor de selecteur blijft het belangrijk om zich niet op geïsoleerde vragenlijstuitkomsten te baseren, maar deze te valideren aan de hand van een 'multi-trait multimethod'-benadering waarbij de uitkomsten van vragenlijstonderzoek, interviewbevindingen, assessmentsimulaties, biografische gegevens en referenties met elkaar worden vergeleken.

10 Trainingseffecten op overige selectie-instrumenten

Er resteren nog enkele selectie-instrumenten waarvan trainings- en of coachingseffecten (nog) niet bekend zijn. Zo zijn er bijvoorbeeld de aan populariteit winnende situational judgement tests (SJT). In Nederland is er wat dit betreft de SQ, een videotoets voor sociale vaardigheden (Maesen de Sombreff et al., 2003). Een SJT is in feite een soort 'geabstraheerde' test voor functiekennis. Hij bestaat uit relevante samples van kritiek functiegedrag. Dat kan in video-vorm zijn, maar ook in schriftelijke, 'multiple choice'-vorm (voor een overzicht, zie McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). De voorspellende validiteit van SJT's is heel behoorlijk, al is er een significante relatie met cognitieve capaciteiten zodat de toegevoegde validiteit beperkt is.

Nguyen, Biderman en McDaniel (2005) onderzochten of faking op SJT's mogelijk is en wat de effecten zijn. Zij onderscheiden twee vormen van SJT's: een gedragsmatige vorm, waarin de sollicitant moet aangeven hoe (on)waarschijnlijk het is dat hij bepaalde gedragingen vertoont, en een kennisvorm, waarbij de sollicitant moet aangeven in hoeverre bepaald gedrag goed dan wel fout is. Merk op dat de eerste variant veel weg heeft van een persoonlijkheidsvragenlijst (*typical behavior*) terwijl de tweede variant een voorbeeld is van een kennis-toets (*maximum behavior*). De resultaten wijzen uit dat een fakingeffect kan optreden als een gedragsmatig format gebruikt wordt, variërend tussen .15 en .34 SD. Dat het kennisformat geen fakingeffect vertoonde is niet zo verwonderlijk: om op dit format te kunnen faken moet de sollicitant voorkennis hebben van de vragen of getraind zijn in de (werking van) onderliggende domeinen.

De effecten van faking op de SJT-gedragsmatig waren echter minder groot dan fakingseffecten op Big Five dimensies (extraversie, consciëntieusheid en emotionele stabiliteit). Zoals eerder verondersteld hingen resultaten op het gedragsmatige format inderdaad minder samen met cognitieve capaciteiten dan het kennisformat.

Van coachingsresultaten met betrekking tot het vervaardigen van resumés zijn ons geen specifieke gegevens bekend. Gezien de mogelijkheden om op het *world wide web* een cv te laten screenen en te laten verbeteren mag worden aangenomen dat er in grote mate sprake is van IM bij het opstellen van cv's. Onderzoek van Cole, Feild en Stafford (2005) laat zien dat uit een cv in redelijke mate persoonsbeschrijvingen van de kandidaat zijn af te leiden. Onderzocht zou dan kunnen worden of dit ook bij gefakete cv's het geval is en wat hiervan de gevolgen zijn voor het verdere selectieproces. Leidt faking op cv's tot een verhoogde kans op doorstroming of slaat het op de kandidaat terug en leidt het eerder tot ontmaskering in het opvolgende interview?

Ook kan het bijvoorbeeld helpen om, als het gaat om biografische gegevens, kandidaten hun scores of antwoorden te laten toelichten, zodat duidelijk wordt dat de gegeven antwoorden gecontroleerd kunnen worden. Uit een onderzoek van Schmitt et al., (2003) blijkt dat deze wijze van scoring van biografische gegevens tot veel lagere scores leidt. De correlatie tussen sociale wenselijkheid en de biografische scores, verschillen tussen subgroepen en de criteriumvaliditeit bleken door deze methode echter niet te worden beïnvloed.

11 Conclusie

Wat betreft de mogelijke verschillende interventies om de score op een selectie-instrument te verhogen kan worden aangenomen dat de meeste sollicitanten zich op enigerlei wijze voorbereiden op een psychologisch selectieonderzoek. Veelal zal dat gaan via gebruik van de oefen- en oriëntatiemogelijkheden van het *www* of via beschikbare literatuur. Het *www* heeft een veel sterker effect dan bijvoorbeeld *grapevinecoaching*. Er zijn tientallen sites om intelligentietests en big five vragenlijsten te oefenen, interviewvragen door te nemen, enz. Ook zijn er discussiefora waarop sollicitanten hun ervaringen delen en vaak om specifieke informatie vragen omtrent een bepaald instrument of een bepaalde procedure. In die zin is *grapevinecoaching* door de technologie achterhaald. Het zou een experiment waard zijn om de effecten van zelftraining via het *www* te vergelijken met de scores van een naïeve groep sollicitanten.

11.1 EFFECTEN: ALGEMEEN OVERZICHT

Wat zijn de effecten van oefening en training op psychologische selectiemiddelen? In tabel 3 is weergegeven wat de in dit hoofdstuk besproken effecten zijn. Daarbij is een onderscheid gemaakt tussen oefenings- en trainingseffecten op de predictor alleen, zoals bij een intelligentietest, en op zowel de predictor als het criteriumgedrag, zoals bij assessmentsimulaties, waarbij de kandidaat ander gedrag leert en dit ook in de functie in praktijk kan brengen. Deze vorm van oefening is niet schadelijk voor de predictieve validiteit en kan gezien worden als een soort 'job' training.

Verder is er, zoals eerder gesuggereerd, sprake van een soort voorbereidingscontinuüm, dat loopt van Impression Management (elke sollicitant doet dit, mag worden aangenomen) tot specifieke en intensieve coaching, iets wat zeer weinig zal voorkomen. Er zijn, analoog aan Kaplan's testing service in de VS, enkele bureaus in Nederland waar sollicitanten tegen betaling specifieke selectie-instrumenten kunnen oefenen. Waar het de echte instrumenten betreft is er sprake van fraudering en unfairness van de zijde van de sollicitant. NIP-psychologen mogen conform hun beroepscode geen authentieke selectie-instrumenten gebruiken bij coachingstrainingen (zie Bloemers, 1988 1989). Maar in sommige tijden is een bepaalde baan een zeer waardevol iets en bij 'high stakes testing' zijn er sollicitanten die er alles voor over hebben om een bepaalde baan te verwerven.

Tabel 3: Overzicht van globale effecten van IM en coaching op selectie-instrumenten

	Effecten van		Coaching	
	IM			
	<i>predictor</i>	<i>criterium</i>	<i>predictor</i>	<i>criterium</i>
IQ tests	-	-	+	-
Vragenlijsten	++	-	+++	-
Interview	+	-	++	+
ACM-opdracht	+	-	++	+

Uit tabel 3 valt af te leiden dat op cognitieve instrumenten oftewel 'maximum behavior'-tests (IQ tests) de effecten minder zijn dan op vragenlijsten. Bij een intelligentietest lijkt een predictoreffect van 1 SD het maximum haalbare (Van der Molen, Keen & Te Nijenhuis, 1995). Let wel: dit is een groot effect. Bij de vragenlijsten of 'typical behavior'-instrumenten is het effect groter. Effecten van rond de 2 SD op persoonlijkheidsvragenlijsten en gedragsgerichte integriteits-tests bij faking zijn geen uitzondering. Bij realistische gedragssimulaties (rolenspelen en het interview) lijken eveneens effecten van ongeveer 1 SD de grens aan te geven van hoeveel er aan gedrag kan worden verbeterd. Het grote verschil met de andere twee categorieën is dat hier predictor- en criteriumgedrag samenvallen. Training op dit soort simulaties zou dus in principe ook tot betere functieprestaties moeten leiden, mits de simulaties voldoende inhoudsva-
 lide ten opzichte van de functie zijn.

Zoals te verwachten zijn de effecten van coaching groter dan die van IM. Intensieve oefening/coaching, gekoppeld aan training (herhaling) en adequate feedback kan een gemiddeld tot groot effect op het vertoonde gedrag hebben. Effecten tussen de .5 en de 1 SD zijn haalbaar.

11.2 FAIRNESS

Vanuit het oogpunt van fairness zou coaching, waar dit effect kan hebben op daadwerkelijk gedrag, voor iedereen in gelijke mate beschikbaar moeten zijn. Wat betreft cognitieve maten is er voldoende oefenmateriaal op het internet en in boekvorm voorhanden.

Zoals uit verschillende onderzoeken is gebleken is er bij coaching op intelligentietests niet sprake van een algemeen effect. Kandidaten profiteren niet allen in gelijke mate van coaching. Hoog scorende kandidaten profiteren meer van training, en ook kandidaten die al herhaaldelijk getest zijn, zijn in het voordeel. Gezien de waarde van de intelligentiescore voor arbeidsprestaties, verdient het aanbeveling om vooral de 'zwakke' groepen in de samenleving op dit punt te ondersteunen, door hen in elk geval een acceptabele vorm van maximale voorbereiding te bieden. Over mogelijkheden om bijvoorbeeld minderheden die gemiddeld lager scoren op intelligentiematen bij selectieonderzoek toch een grotere acceptatiekans te bieden, zie Sackett et al., 2001.

Waargenomen fairness speelt ook een rol bij de wijze waarop sollicitanten tegen een organisatie aankijken (Hausknecht, Day & Thomas, 2004), en bijvoorbeeld bij de vraag of men een baanaanbod accepteert of een selectieprocedure verlaat.

11.3 PREDICTIEVE VALIDITEIT

De meeste studies naar de effecten van coaching laten zien, voor zover onderzocht, dat de effecten op de predictieve validiteit niet schadelijk zijn. Zolang coaching de rangorde op de predictor onaangetast laat is er niet veel aan de hand. Er zal echter meer onderzoek naar dit coachingseffect verricht moeten worden, bij voorkeur in een realistische A&O-setting. Alleen zo kunnen de 'ware' effecten van faking, testangst en motivatie om te presteren in een selectiesituatie in kaart gebracht worden. Ook de methodologische opzet (between subjects of within subjects) is daarbij van belang. Veel van de besproken onderzoeken werken met studenten als proefpersonen en/of met gespeelde situaties. Dit tast de ecologische validiteit van het onderzoek aan en laat 'ware' effecten in selectiesituaties verhold.

Voorlopig lijken de effecten van coaching en oefening echter niet al te bedreigend voor de predictieve validiteit. De gemiddelde selectieprocedure, bestaande uit tests, vragenlijsten en een interview, heeft een dusdanig grote 'power' dat het waarschijnlijk onmogelijk is om als 'true negative', via faking tot (false) positive bestempeld te worden.

11.4 COACHING EN FAKING

Coaching wijst kandidaten op de implicaties van verschillende antwoordstrategieën op selectie-instrumenten. De sollicitant die vervolgens kiest voor een bepaalde strategie bepaalt met deze keuze of hij bedrog pleegt. Deze vraag speelt niet alleen bij zelfbeoordelvragenlijsten. Beschouw bijvoorbeeld een situatie waarin een potentieel WAO'er een intelligentiedefect suggereert op een intelligentietest door bewust foutieve antwoorden te geven en zo een uitering in de wacht te slepen wegens verminderd cognitief functioneren. Is dit principieel anders dan de situatie van een sollicitant die oefent met een echte intelligentietest om zo zijn score ruim een standaarddeviatie op te hogen en zo een baan als managementtrainee te verwerven?

In feite kan de vraag of er gefaket is alleen door de sollicitant beantwoord worden: vulde hij, willens en wetens iets anders in dan hij 'normaal' zou hebben ingevuld of zou hebben *kunnen* invullen? Maar wat is normaal? Faking wordt uitgelokt door specifieke situaties waarin belangen op het spel staan. Dat mensen hun huid zo duur mogelijk verkopen (IM) wordt algemeen gewaardeerd, maar bij faking is er blijkbaar een principiële grens overschreden. In feite kan alleen op objectieve maten worden aangetoond dat de sollicitant gefaket heeft, zoals bij het CV en de IQ-test (en bij de IQ test kan de sollicitant dan nog vol houden dat hij plotseling het licht zag tijdens het maken van de test). Bij zelfbeoordelingsschalen ligt de bewijslast echter bij de selecterende psycholoog. Zoals eerder betoogd heeft die meer soorten gegevens ter beschikking. Uit een samenstel van indicatoren zal hij conclusies dienen te trekken over het persoonlijkheidsprofiel van de kandidaat. Een oplossing is om bij blijvende twijfel dit met de kandidaat te bespreken, of, als een tussenoplossing een herafname aan te bieden. Meestal zal de kandidaat dan eieren voor z'n geld kiezen en met een genuanceerdere versie komen. Men beschouwt de kandidaat dan nog steeds als een verantwoordelijk persoon die zichzelf als object in de selectieprocedure ter discussie wil stellen. Als de sollicitant stug volhoudt dat hij niet ge-

faket heeft en dat de scores op de vragenlijst toch echt wel kloppen, dan kan de psycholoog constateren dat hij met zoveel tegenstrijdige gegevens zit, dat hij geen eenduidig profiel kan opstellen omdat hij dat ten opzichte van de sollicitant, het wetenschappelijk forum en de opdrachtgever niet verantwoord vindt. In die zin slaat een te grote mate van faking op korte termijn (nog binnen het selectieonderzoek) of langere termijn (binnen het functioneren) altijd terug op de faker: hij of zij wordt ontmaskerd. Echt slimme kandidaten faken dan ook met verstand, zij blijven binnen de grenzen van wat geloofwaardig is, maar poetsen hun imago wel zo veel mogelijk op (IM). Het is aan de psycholoog om dit spel te doorgronden en te signaleren waar de grenzen van de menselijke natuur in het geding komen.

12 Tot besluit: een nieuw selectiescenario?

Algemene voorbereiding op een psychologisch onderzoek via boeken of via het *World Wide Web* is tegenwoordig gemeengoed. Daar is ook niets mis mee, het getuigt van een gemotiveerde en serieuze instelling van de kant van de sollicitant. Van de kant van de selecteur getuigt het van een streven naar transparantie en het willen faciliteren van de zelfselectiemogelijkheid die sollicitanten hebben. Door voldoende info over procedures en instrumenten vrij te geven wordt de acceptatie (en de *face validity*) van procedures vergroot. Ook kan dit proces van transparantie bijdragen tot verhoogde zelfselectie van de kant van de sollicitant: ‘gezien de procedure en de instrumenten, is dit wel iets voor mij?’ Sollicitanten kunnen zich een beeld vormen van de inhoud en de moeilijkheidsgraad van de verschillende selectiemiddelen. Bij algemeen gebruik van deze strategie kunnen de constructvaliditeit en de predictieve validiteit van de procedure zelf toenemen, omdat *error* variantie (het niet begrijpen van opdrachten, onzekerheid, stress en testangst) geminimaliseerd worden.

Wat betreft de effecten, deze zullen zoals eerder weergegeven beperkt blijven, grofweg in de orde van .5 SD. Belangrijker is bij deze strategie dat sollicitanten optimaal de kans geboden wordt om zo goed mogelijk te scoren op de verschillende predictoren. Dat verzacht de verticale relatie enigszins (de sollicitant heeft niet gevraagd om een psychologisch onderzoek of assessment) en maakt ook de positie van de testpsycholoog meer acceptabel. Het blijft echter binnen de parameters van een verticale opdrachtrelatie.

Er is ook een geheel ander coachingscenario denkbaar, waarvan de aanzetten zijn gegeven door Hofstee (2001). De kern van dit nieuwe paradigma op het gebied van personeelsselectie bestaat uit het eenzijdig doorbreken van de verticale opdrachtrelatie. Uitgangspunt is het zelfinzicht en het zelfselecterende vermogen van de sollicitant, waarbij de sollicitant bepaalt welke selectiemiddelen hij wil ondergaan. De rol van de selectiepsycholoog verandert hierbij van beoordelaar naar adviseur. Organisaties faciliteren deze nieuwe rol door de sollicitant in de gelegenheid te stellen een keuze uit het arsenaal aan selectiemiddelen te ondergaan. Dit scenario heeft als grote winstpunt dat de verticale situatie, waarbij de sollicitant zich gereduceerd ziet tot object waarover geoordeeld wordt, verdwijnt. De sollicitant wordt nu zelf opdrachtgever en huurt via de organisatie de selectiepsycholoog in als adviseur. Hiermee is de doortrapte

sollicitant echter niet buitenspel gezet, want deze laat zich coachen op intelligentietests, en zegt vervolgens tegen de testpsycholoog dat hij twijfels heeft over zijn cognitieve capaciteiten en dus uitsluitend een intelligentietest wil ondergaan....Ook hier blijft dus alertheid van de selectiepsycholoog op zijn plaats, zij het nu in de rol van adviseur in plaats van selecteur.

Groot voordeel van de horizontale strategie is dat mensen als volwassen en autonome subjecten gezien worden die zelf in staat zijn beslissingen over hun eigen functioneren te nemen en keuzes te maken die aansluiten bij hun capaciteiten. Daarbij laten zij zich adviseren door een psycholoog, die, eventueel op aanvraag van de sollicitant, met behulp van bepaalde instrumenten meer licht kan werpen op onzekerheden en onduidelijkheden in gedrag en capaciteiten van de sollicitant. Dit scenario is aantrekkelijk voor sollicitanten en psychologen. Het woord is vooral aan het bedrijfsleven om dit idee te faciliteren. Coaching en voorbereidingsvraagstukken verschuiven zo naar de periferie, daar niet wordt geperverteerd op het leveren van de optimale prestatie, maar op realistische beeldvorming van de eigenschappen en vaardigheden van mensen om vanuit eigen verantwoordelijkheid een plek in het arbeidsbestel te verwerven waar zij optimaal tot hun recht komen. Zowel in menselijk als vanuit economisch perspectief.

Literatuur

- Anastasi, A. (1981). Diverse effects of training on tests of academic intelligence. In: B. F. Green (ed.). *New Directions for testing and measurement: issues in testing-coaching, disclosure and ethnic bias*, no. 11. San Francisco, Jossey-Bass.
- Allalouf, A., & G. Ben-Shakar (1998). The effect of Coaching on the Predictive Validity of Scholastic Aptitude Tests. *Journal of Educational Measurement*, pp. 31-47.
- Alliger, G. M. en S.A. Dwight (2000). A Meta-Analytic Investigation of the Susceptibility of integrity tests to Faking and Coaching. *Educational and Psychological Measurement*, 60, No. 1, pp. 59-72.
- Anderson, N., M. Born & N. Cunningham-Snell (2001). Recruitment and Selection: Applicant Perspectives and Outcomes. In: Anderson, N., Ones, D.S., Sinangil, H.K. & Viswesvaran, C.: *Handbook of Industrial, Work & Organizational Psychology*, vol. 1, pp. 200-218. London, Sage.
- Anderson, N. (2004). Editorial- the dark Side of the Moon: Applicant Perspectives, negative psychological Effects (NPEs), and Candidate Decision Making in Selection. In: *International Journal of selection and Assessment*, 12, no. 1-2, pp. 1-8.
- Bangert-Drowns, R. L., J.A. Kulik, & C.L.C. Kulik (1983). Effects of coaching programs on achievement test performance. *Review of Educational Research*, 4, 53, pp. 571-585.
- Bargh, J. A. (1994). Automatic and conscious processing of social information. In: R. Wyer & T. Srull (Eds.). *Handbook of social cognition*, vol. 3, 1-43. Hillsdale, NJ: Erlbaum.
- Barrick, M. R., M. K. Mount & T. A. Judge (2001). Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next? *International Journal of Selection and Assessment* 9 (1&2), pp. 9-30.
- Bloemers, W. (1988). Psychologische tests: If you can't beat them, buy them. Hoe de testpsychologie haar eigen ruiten in laat gooien. *Psychologie en Maatschappij*, 12, 4, pp. 361-375.
- Bloemers, W. (1989). De markt van angst en geluk. *Intermediair*, 25, 48, pp. 57-61.
- Bloemers, W. (1997). Het selectieproces: Een voorstel voor een optimaal rendement. In: Dam, K. van, & Bloemers, W., (red.): *Individu, Arbeid & Organisatie*. Utrecht, De Tijdstroom.
- Bloemers, W. (2001). De kleine assessmentgids. Amsterdam, Ambo.
- Bloemers, W. (2005). *Het psychologisch Onderzoek*, een oefenboek Baarn, Ambo.
- Bloemers, W. (2006). De gewapende sollicitant. Effecten van oefening, voorbereiding en coaching op psychologische selectiemiddelen. In: *Cursusboek Selectie en assessment*. Open Universiteit Nederland.
- Caldwell, D. F. & J. M. Burger (1998). Personality Characteristics of job applicants and success in screening interviews. *Personnel Psychology*, 51, pp. 119-136.

- Carrol, J.B. (1993). *Human Cognitive Abilities*. A Survey of Factor-analysis studies. Cambridge University Press, Cambridge.
- Chan, D., N. Schmitt, J. M. Sacco & R.P. DeShon (1998). Understanding Pretest and Posttest Reactions to Cognitive Ability and Personality Tests. *Journal of Applied Psychology*, 83, pp. 471-485.
- Clause, C. S., K. Delbridge, N. Schmitt, D. Chan & D. Jennings (2001). Test preparation activities and employment test performance. *Human Performance*, 14(2), pp. 149-167.
- Cohen, S. L. (1978). Standardization of assessment center technology: Some critical concerns. *Journal of Assessment Center technology*, 1, pp. 1-10.
- Cole, M. S., H.S. Field & J.O. Stafford (2005). Validity of Resume Reviewers' Inferences Concerning Applicant personality based on resume Evaluation. *International Journal of Selection and Assessment*, 13, pp. 321-324.
- Conlon, T.J. (2004). A review of informal learning literature, theory and implications for practice in developing global professional competence. *Journal of European Industrial Training*, 28, pp. 283-295.
- Costa, P. T. & R.R. McCrae (1992). *NEO PI-R professional manual*. Florida, Psychological Assessment Resources Inc.
- Dany, F. & V. Torchy (1994). Recruitment and selection in Europe: policies, practices and methods. In: C. Brewster & A. Hegewisch (eds.). *Policy and practice in European human resource management: The Price Waterhouse Cranfield Survey* (pp. 68-88). London: Routledge.
- Dodd, W. E. (1977). Attitudes toward assessment center programs. In: J.L. Moses & W.C. Byham (eds): *Applying the Assessment Center Method*. New York: Pergamon Press.
- Denning, D. L. & D.L. Grant (1979). Knowledge of the assessment center process: Does it affect candidate ratings? *Journal of Assessment Center Technology*, 2, pp. 7-12.
- Drasgow, F. (2003). Intelligence and the Workplace. In: Borman, W. C., Ilgen, D. R. & Klimoski, R. J. *Handbook of Psychology, volume 12, Industrial and Organizational Psychology*. Hoboken, NJ, John Wiley & Sons.
- Drenth, P. J. D. & G. Hoolwerf (1970). *Numerieke Aanleg Test*. Lisse, Swets & Zeitlinger.
- Drenth, P. J. D. & P.C.W. van Wieringen (1969). *Verbale Aanleg Testserie; Subtest Verbale Analogieën*. Lisse, Swets & Zeitlinger.
- Dulewicz, V. & C. Fletcher (1982). The relationship between previous experience, intelligence and background characteristics of participants and their performance in an assessment centre. *Journal of occupational Psychology*, 55, pp. 197-207.
- Dulewicz, V., C. Fletcher & P. Wood (1983). A study of the internal validity of an assessment centre and of participants' background characteristics and attitudes: A comparison between British and American findings. *Journal of Assessment Center Technology*, 6, pp. 15-24.
- Dwight, S.A. & Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 10-23.
- Eder, R.W. & H.H. Harris (eds.) (1999). *The Employment Interview Handbook*. Thousand Oaks (Cal.), Sage Publications.
- Elligson, J.E. & E.D. Heggstad (2003). The Viability of retesting and warnings for Managing Faking Tendencies. In: *Practical Considerations for Implementing personality testing in Organizations*. SIOP Symposium, Orlando, Florida.
- Ellis, A. P. J., B.J. West, A.M. Ryan & R.P. DeShon (2002). The Use of Impression Management Tactics in Structured Interviews: A Function of Question Type? *Journal of Applied Psychology*, 2002, 87, pp. 1200-1208.
- Gaugler, B.B., D.B. Rosenthal, G.C. Thornton III & C.B. Bentson (1987). Meta-Analysis of assessment center validity. *Journal of Applied Psychology*, 72, pp. 493-511.
- Gordon, R. A. (1997). Everyday Life as an Intelligence Test: Effects of Intelligence and Intelligence Context. *Intelligence*, 24 (1), pp. 203-320.
- Gottfredson L.S. (1997). Why g matters: the complexity of everyday life. *Intelligence*, 24, pp. 79-132.
- Gottfredson, L. S. & J.J. Deary (2003). Intelligence Predicts Health and Longevity, But Why? *Current Directions In Psychological Science*, 13, pp. 1-4.
- Griffin, B. & B. Hesketh (2004). Why Openness to Experience is not a Good Predictor of Job Performance. *International Journal of Selection and Assessment*, 12, (3), pp. 243-251.
- Griffin, B., B. Hesketh & D. Grayson (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. *Personality and Individual Differences*. 36, pp. 1545-1558.
- Hausknecht, J.P., C.O. Trevor & J.L.Farr (2002). Retaking Ability Tests in a Selection Setting: Implications for Practice Effects, Training Performance, and Turnover. *Journal of Applied Psychology*, 87, 2, pp. 243-254.

- Hausknecht, J.P., D.V. Day & S.C. Thomas (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, pp. 639-683.
- Hoff Macan, T., M.J. Avedon, M. Paese & D.E. Smith (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, pp. 715-738.
- Hofstee, W.K.B. (2001). Persoonsontwikkeling voor wie? In: *Human Resource Development, Thema Persoonlijke Ontwikkeling*. Alphen aan de Rijn, Kluwer.
- Hough, L. M. & D.S. Ones (2001). The structure, Measurement, Validity, and use of Personality Variables in Industrial Work, and Organizational Psychology. In: Anderson, N., Ones, D.S., Sinangil, H.K. & Viswesvaran, C.: *Handbook of Industrial, Work & Organizational Psychology*, vol. 1, pp. 233-267. London, Sage.
- Huffcut, A. I., J. M. Conway, P. L. Roth & N. J. Stone (2001). Identification and meta-analytic assessment of psychological construct measured in employment interviews. *Journal of Applied Psychology*, 86, pp. 897-913
- Kolk, N. J., M. Ph. Born, H. van der Flier (2003). The Transparent Assessment Centre: The Effects of Revealing Dimensions to Candidates. *Applied Psychology: An International Review*, 52, pp. 648-668.
- Kok, B. & F. de Jongh, F. (2004). *Assessment doen. Hoe werkt het voor jou?* Utrecht, Het Spectrum.
- Kulik, J.A., R.L. Bangert-Drowns, & C.C. Kulik (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin*, 95, pp. 179-188.
- Kulik, J.A., C.C. Kulik & R.L. Bangert (1984). Effectiveness of coaching for aptitude and achievement test scores. *American Educational research Journal*, 21, pp. 435-447.
- Kureca, P. M., J.M. Austin, W. Johnson & J.L. Mendoza (1982). Full and errant coaching effects on assigned role leaderless group discussion performance. *Personnel Psychology*, 35, pp. 805-812.
- Lance, C. E., W.H. Newbolt, R.D. Gatewood, M.R. Foster, N.R. French. & D.E. Smith (2000). Assessment Center Exercise Factors Represent Cross-Situational Specificity, Not Method Bias. *Human Performance*, 13 (4), pp. 323-353.
- Lance, C. E., T.A. Lambert, A.G. Gewin, F. Lievens. & J.M. Conway (2004). Revised Estimates of Dimension and Exercise Variance Components in Assessment Center Postexercise Dimension Ratings. *Journal of Applied Psychology* 89, No. 2, pp. 377-385.
- Lievens, F. (2002). Trying to Understand the Different Pieces of the Construct Validity Puzzle of assessment Center: An Examination of Assessor and Assessee Effects. *Journal of Applied Psychology*, 87, pp. 675-686.
- Lievens, F. & R.J. Klimoski (2001). Understanding the assessment center process: Where are we now? In C. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 12, pp. 245-286). New York: Wiley Publishers.
- Lievens, F., W. de Corte & K. Brysse (2003). Applicant Perceptions of Selection procedures: The Role of selection Information, Belief in tests, and Comparative Anxiety. *International Journal of Selection and Assessment*, 11, pp. 67-77.
- Lievens, F., T. Buyse & P.R. Sackett (2005). Retest Effects in Operational Selection Settings: Development And Test Of a Framework. *Personnel Psychology*, 58, pp. 981-1007.
- Maesen de Sombreff, P. van der, (1997). Intelligentietests en beroepssucces. In: W. Tomic & H. T. van der Molen (Ed.). *Intelligentie en Sociale competentie*. Lisse, Swets & Zeitlinger.
- Maesen de Sombreff, P. van der, M. Born, K. van Oudenhoven-van der Zee & D. Ruhe (2003). Situatie-nale beoordelingstests in de schijnwerpers, *De Psycholoog*, 38, 2, pp. 58-67.
- Maurer, T., J. Solamon & D. Troxtel (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology*, 83, pp. 128-136.
- Maurer, T. J., J.M. Solamon, K.D. Andrews & D.D. Troxtel (2001). Interviewee coaching, preparation strategies, and response strategies in relation to performance in situational employment interviews: An extension of Maurer, Solamon, and Troxtel (1998). *Journal of Applied Psychology*. 86, pp. 709-717
- McCarthy, J. M & R. D. Goffin, (2005). Selection test Anxiety: Exploring tension and Fear of failure Across the Sexes in Simulated Selection Scenarios. *International Journal of Selection and assessment*, 13, pp. 282-295.
- McDaniel, M. A., D.L. Whetzel, F.L. Schmidt & S.D. Maurer (1994). The validity of employment interviews: a comprehensive review and meta-analysis. *Journal of applied psychology*, 79, pp. 599-616.
- McDaniel, M.A., F.P. Morgeson, E.B. Finnegan, M.A. Campion & E.P. Braverman (2001). Predicting job performance using situational judgement tests: A clarification of the literature. *Journal of Applied psychology*, 86, pp. 730-740.

- McFarland L.A., A.M. Ryan & S.D. Kriska (2002). Field study investigation of applicant use of influence tactics in a selection interview. *Journal of Psychology*, 136 (4), pp. 383-98.
- McFarland, L.A., A.M. Ryan & S.D. Kriska (2003). Impression management use and effectiveness across assessment methods. *Journal of Management*, 29, pp. 641-661.
- McFarland, L.A., G. (J) Yun, C.M. Harold, L. Viera Jr. & L.G. Moore (2005). An examination of impression management use and effectiveness across assessment center exercises: The role of competency demands. *Personnel Psychology*, 58, (4), 2005. pp. 949-980.
- McIntyre, F. (1980). The reliability of assessment center results after feedback. *Journal of assessment center Technology*, 3, pp. 10-14.
- Mehrens, W.A. (1989). Preparing student to take Standardized Achievement Tests (Washington DC. American Institutes for Research; ERIC Document Reproduction Service No. ED 314 427).
- Messick, S. (1981). The controversy over coaching: issues of effectiveness and equity. In: B. F. Green (ed.). *New Directions for testing and measurement: issues in testing-coaching, disclosure and ethnic bias*, no. 11. San Francisco, Jossey-Bass.
- Molen, H.T. van der, G. Keen & J. Te Nijenhuis (1995). The effects of intelligence test preparation. *European Journal of Personality*, 9, pp. 43-56.
- Molen, H.T. van der, G.N. Smit, M.A. Hommes & G. Lang (1995). Two decades of microtraining in the Netherlands; a meta-analysis. *Educational Research and Evaluation*, 1, pp. 347-378.
- Moscoseo, S. (2000). Selection Interview: A Review of validity Evidence, Adverse Impact and Applicant Reactions. *International Journal of selection and assessment*, 8, pp. 237-247.
- Nguyen, N. T., M.D. Biderman & M.A. McDaniel (2005). Effects of Response Instructions on Faking a Situational Judgment Test. *International Journal of Selection and Assessment*, 13 (4), pp. 250-260.
- Pannone, R. D. (1984). Predicting test performance; A content valid approach to screening applicants. *Personnel Psychology*, 37, pp. 507-514.
- Petty, M. M. (1974). A Multivariate Analysis of the Effects of Experience and Training upon Performance in a Leaderless Group Discussion. *Personnel Psychology*, 27, pp. 271-282.
- Powers, D. E. (1986). Relation of Test Item Characteristics to Test Preparation/Test Practice Effects: A Quantitative Summary. *Psychological Bulletin*, 1, pp. 67-77.
- Powers, D.E. & D.A. Rock (1999). Effects of coaching on SAT I: reasoning test scores. *Journal of Educational Measurement*, 36, pp. 93-118.
- Rosenfeld, P., R.A. Giacalone & C.A. Riordan (1995a). Impression Management. In: Nicholson, N. (Ed.). *Blackwell Dictionary of Organizational Behavior*. Oxford, UK, Blackwell Publishers.
- Rosenfeld, P., R.A. Giacalone & C.A. Riordan. (1995b) . *Impression Management in Organisations*. London, Routledge.
- Roth, P.L., C.H. van Iddekinge, A.I. Huffcut, C.E. Eidson jr, & M.J. Schmit (2005). Personality Saturation in Structured Interviews. *International Journal of Selection and Assessment*, 13, pp. 261-273.
- Ryan, A. M., R.E. Ployhart, G.J. Greguras & M.J. Schmitt (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, 51, pp. 599-621.
- Ryan, A.M., L. McFarland, H. Baron & R. Page (1999). An International Look at Selection Practices: Nation and Culture as Explanations for Variability in Practice. *Personnel Psychology*, 52, pp. 359-391.
- Sackett, P.R., L.R. Burreis & A.M. Ryan (1989). Coaching and practice effects in personnel selection. In: C. Cooper & I. Robertson (Eds.). *International review of industrial and organizational psychology*. New York: Wiley.
- Sackett, P.R., N. Schmitt, J.E. Ellingson & M.B. Kabin (2001). High-Stakes testing in Employment, Credentialing and Higher Education. *American Psychologist*, 56, pp. 302-318.
- Sackett, P.R., S. Zedeck & L. Fogli (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, 73, pp. 482-486.
- Salgado, J. F. & S. Moscoso. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, pp. 299-324.
- Salomon, G.E. (1985). Effects of training in test-taking skills on Achievement Test Performance: A Quantitative Synthesis. *Journal of Educational research*, 78, pp. 261-266.
- Schmidt, F.L & J.E. Hunter (1983). Individual differences in productivity: an empirical test of estimates derived from studies of selection procedure utility . In: *Journal of Applied Psychology*, 68, pp. 407-414.
- Schmidt, F.L & J.E. Hunter (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. In: *Psychological Bulletin*, 124, pp. 262-274.

- Schmitt, N., F.L. Oswald, B.H. Kim, L.J. Ramsay & M.A. Gillespie (2003). Impact of elaboration on Socially Desirable Responding and the Validity of Biodata Measures. *Journal of Applied Psychology* 88, 2, pp. 979-988.
- Schmitt, N. & A.M. Ryan (1993). The big five in personnel selection: factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, pp. 966-974.
- Schönrock-Adema, J. (2002). De ontwikkeling en evaluatie van een zelfinstructieprogramma voor een training in basisgespreksvaardigheden. Groningen, Rijksuniversiteit Groningen, academisch proefschrift.
- Stankov, L. (1986). Kvachev's Experiment: Can we Boost Intelligence? *Intelligence*, 10, pp. 209-230.
- Stanush, P.L. (1997). Factors that influence the susceptibility of self-report inventories to distortion: A meta-analytic investigation. Ongepubliceerde dissertatie, Texas, A&M University.
- Stevens, C.K. en A.L. Kristof (1995). Making the Right Impression: A field study of Applicant Impression Management During Job Interviews. *Journal of Applied Psychology*, 80, pp. 587-606.
- Te Nijenhuis, J., O.F. Voskuil & N.B. Schijve (2001). Practice and Coaching on IQ tests: Quite a Lot of g. *International Journal of Selection and Assessment*, 9, pp. 302-308.
- Thornton, G.C. & Byham, W.C. (1982). *Assessment center and managerial performance*. New York, Academic press.
- Van Dam, K. (2003). Trait perception in the Employment Interview: A Five-Factor Model perspective. *International Journal of Selection and Assessment*, 11, pp. 43-55.
- Van Iddekinge, C.H., P.H. Raymark & P.L. Roth (2005). Assessing personality with a structured employment interview: construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, pp. 536-552.
- Van Minden, J.J.R. (2006). *Assessment Centers*. Amsterdam, Business Contact.
- Viswesvaran, C. & D.S. Ones (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational & Psychological Measurement*, 59, pp. 197-211.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.
- Zickar, M. J. & C. Robie (1999). Modeling faking good on personality items : An item-level analysis. *Journal of Applied Psychology*, 84, pp. 551-563.