

De Webcamtest



Een onderzoek naar de predictieve, incrementele en indrukvaliditeit van de webcamtest

Masterthesis

Augustus 2008

Door: Marit Op de Beek

Studentnummer: 288172

Begeleiding: Janneke Oostrom

2^o Beoordelaar: Marise Born

In opdracht van: GITP

Van der Maesen | Koch HRM-advies

Erasmus Universiteit Rotterdam

Faculteit der Sociale Wetenschappen, Psychologie

Samenvatting

De webcamtest is een nieuw selectie-instrument in videovorm gemaakt voor het meten van sociale competenties in werksituaties. Het doel van deze studie was het bepalen van de predictieve validiteit van de webcamtest. Daarnaast is de incrementele validiteit van de webcamtest-leidinggeven bepaald ten opzichte van een intelligentietest, een persoonlijkheidsvragenlijst en een *situational judgment test* voor leidinggevende capaciteiten in videovorm (video-SJT). Onderzoek onder 106 psychologie-studenten aan de Erasmus Universiteit te Rotterdam liet zien dat de webcamtest een significante correlatie heeft met het criterium professioneel gedrag ($r = .36, p < .01$), een vragenlijst met betrekking tot de inzet, het literatuurinzicht en het voorzitterschap van de studenten. Ook had de webcamtest een incrementele validiteit ten opzichte van een intelligentietest, een persoonlijkheidsvragenlijst en een video-SJT met betrekking tot het criterium professioneel gedrag. Deze resultaten impliceren dat de webcamtest een gedeelte van het professionele gedrag van studenten meet dat niet wordt verklaard door andere selectie-instrumenten. Ook zijn enkele percepties ten aanzien van dit nieuwe selectie-instrument onderzocht en is bekeken of deze percepties invloed hebben op de prestatie op de webcamtest. De prestatie op de webcamtest werd niet beïnvloed door testangst, computerangst en testmotivatie. Wel hadden studiejaar, *self-efficacy* en enkele dimensies van persoonlijkheid invloed op de prestatie van de kandidaten op de webcamtest. De kandidaten beschouwden de webcamtest als een indrukvalide test, maar beschouwden de video-SJT als een test met een hogere indrukvaliditeit. Ook vonden de kandidaten de video-SJT leuker om te maken dan de webcamtest. Tot slot zijn de limitaties van dit onderzoek en suggesties voor verder onderzoek naar de webcamtest besproken.

Inleiding

Nederlandse bedrijven hebben de laatste jaren moeite met het binnenhalen en behouden van goed personeel (ANP, 2007). Dit komt onder andere door een aantrekkende arbeidsmarkt (ANP, 2006), het stijgende aantal online vacatures (Novem, 2007) en het dalende aantal werkzoekenden (ANP, 2007). Als de arbeidsmarkt krap is, lijkt het minder zinvol om een uitgebreide selectieprocedure in te zetten (Cook, 2004). Veel organisaties zijn al blij als ze überhaupt een vacature kunnen vervullen en willen de sollicitanten niet afschrikken met een complexe selectieprocedure (Ards, 2006). Van Tilborg (2006) benadrukt echter dat het niet verstandig is om onzorgvuldig om te gaan met de selectieprocedure. Als er een ongeschikte kandidaat wordt aangenomen, moet er uiteindelijk toch weer een nieuwe procedure gestart worden. Ook kan het aannemen van een ongeschikte werknemer een bedrijf veel geld kosten. Voorbeelden hiervan zijn een nieuwe werknemer die zich na de proefperiode maandelijks ziek meldt, een werknemer die zijn of haar tijd met name besteedt aan *hyves* of een medewerker die niet achter de visie van het bedrijf staat en dat ook aan klanten verkondigt. Yoo en Muchinsky (1998) vonden grote utiliteitsverschillen tussen de meest productieve en de minst productieve werknemers in diverse beroepen. Hoe complexer het beroep hoe groter het verschil tussen de minst productieve en de meest productieve werknemer. Bijvoorbeeld het verschil tussen een zeer productieve en een niet zo productieve glazenwasser ligt rond de \$ 12.000 per jaar, maar bij het beroep arbeids- en organisatiepsycholoog kan dat verschil oplopen tot \$ 124.000 per jaar. Aan de andere kant wordt de werknemer zelf ook steeds mondiger en stelt hogere eisen aan een baan dan 15 jaar geleden (Ards, 2006). De werknemer heeft keus en het aanbod is heel toegankelijk. Denk hierbij maar aan de slogan van de online vacaturesite Monsterboard: “Er is altijd ergens een betere baan!”

Deze maatschappelijke ontwikkelingen reflecteren zich in de ontwikkeling van selectie-instrumenten en -methoden. Zo wordt bijvoorbeeld steeds meer aandacht besteed aan de mening van kandidaten over de tests in verband met de reputatie van de organisatie (Ryan & Ployhart, 2000). Ook het competentiegerichte testen, dat zorgt voor een holistischer beeld van de werknemer, is in opmars (Lievens & Sanchez, 2007).

De webcamtest (Van der Maesen, 2005) is een nieuwe test ontwikkeld om (sociale) werkgedragingen in een bepaalde functie of werksetting te meten. Bij afname van de webcamtest krijgt een kandidaat 10 á 12 situaties voorgelegd in videovorm, waarop hij of zij een adequate reactie moet geven in de webcam. De situaties zijn korte praktijksimulaties gerelateerd aan een bepaald vakgebied en dekken een breed scala aan competenties en contextfactoren. De reacties van de kandidaten worden opgenomen en achteraf op verbale en non-verbale aspecten beoordeeld door drie onafhankelijke experts (bijvoorbeeld psychologen).

In dit onderzoek wordt nagegaan of deze webcamtest een voorspellende waarde heeft voor professioneel sociaal gedrag van studenten tijdens hun studie. Ook wordt onderzocht of deze test een toegevoegde waarde heeft ten opzichte van een *situational judgment test* in videovorm (video-SJT) en

ten opzichte van veelgebruikte en onderzochte selectie-instrumenten, zoals een intelligentietest en een persoonlijkheidsvragenlijst. Daarna wordt bekeken of bepaalde percepties zoals motivatie, *self-efficacy*, computerangst en testangst invloed hebben op de prestatie op de webcamtest. Indien dat zo is worden de onderzoeksresultaten misschien negatief of positief beïnvloed. Tot slot gaan we na of de kandidaten de webcamtest indrukvalide en leuk vinden in vergelijking met de andere tests die in dit onderzoek zijn gebruikt.

In de hierop volgende paragraaf wordt beschreven welke trends zichtbaar zijn in de personeelsselectie en hoe deze trends de ontwikkeling van selectie-instrumenten en –methoden beïnvloeden. Ook de ontwikkeling van de webcamtest is gerelateerd aan deze trends. Omdat de webcamtest een vorm van video-assessment is, worden daarna de voor- en nadelen van video-assessment besproken. Aansluitend wordt dieper ingegaan op de theoretische achtergrond van de webcamtest en worden de veronderstelde kosten en baten die gepaard gaan met deze nieuwe vorm van assessment besproken. Tot slot wordt op basis van deze theoretisch achtergrond en voorgaand onderzoek naar video-assessment een aantal hypothesen opgesteld. Deze hypothesen worden getoetst bij een groep studenten van de Erasmus Universiteit te Rotterdam om antwoord te krijgen op de centrale vraag van dit onderzoek : “Wat is de voorspellende waarde van de webcamtest voor professioneel sociaal gedrag?”

Ontwikkelingen in personeelsselectie

Lievens, Van Dam en Anderson (2002) vatten de belangrijkste trends met betrekking tot personeelsselectie samen aan de hand van een literatuurstudie. De vier thema's die zij tegen kwamen zijn: (1) de krappe arbeidsmarkt, (2) dimensie-, construct- of competentiegericht testen, (3) de percepties van de sollicitanten op selectieprocedures en (4) technologische ontwikkelingen. Ook in andere literatuurbronnen vinden we deze, soortgelijke of gerelateerde factoren terug die alle invloed hebben op de ontwikkeling van selectie-instrumenten en -methoden (e.g., Bakker, 2006).

De eerste twee thema's die Lievens en zijn collega's (2002) noemen, namelijk de krappe arbeidsmarkt en het competentiegericht beoordelen, zijn gerelateerd aan elkaar. Bakker (2006) noemt bijvoorbeeld dat selectiecentra een verschuiving maken naar ontwikkelcentra als gevolg van de krapte op de arbeidsmarkt. Ook de selectie-instrumenten worden daardoor steeds meer ontwikkelgericht. Eigentijdse selectie-instrumenten meten eerder competenties die ontwikkeld kunnen worden, dan vaardigheden of kennis waarover een kandidaat wel of niet beschikt. Het meten van kwaliteiten op competentieniveau geeft namelijk een completer beeld van de werknemer (Lievens & Sanchez, 2007) en dat brengt enkele voordelen met zich mee.

Zo kunnen macroaspecten, zoals aansluiting bij de organisatiecultuur, ook meegenomen worden in de beoordeling. Niet alleen moet gezocht worden naar een kandidaat die de werkzaamheden kan (leren) en leuk vindt, ook moet de potentiële werknemer passen in de organisatie (Borman, Ilgen, & Klimoski, 2003). Het is bijvoorbeeld belangrijk dat de werknemer zich thuis voelt in de organisatie en dat de werknemer zich kan vinden in het organisatiebeleid en de organisatiecultuur. Wanneer gestreefd

wordt naar deze *person-organization fit* is de kans groter dat de nieuwe werknemer zich graag wil ontwikkelen binnen de organisatie en minder snel op zoek gaat naar groeikansen buiten de organisatie. Dit laatste is natuurlijk vooral gewenst bij een krappe arbeidsmarkt. Een andere belangrijke reden om sollicitanten te beoordelen en te selecteren op hun competenties is dat medewerkers tegenwoordig multi-inzetbaar moeten zijn. Banen zijn immers niet meer altijd duidelijk gedefinieerd en werknemers nemen diverse werkrollen aan in verschillende projecten binnen de gehele organisatie (Lievens et al., 2002).

Het derde thema dat Lievens en zijn collega's (2002) beschrijven als trend in personeelsselectie is de groeiende aandacht voor percepties die de sollicitanten hebben op sollicitatieprocedures. Ryan en Ployhart (2000) stellen dat ook deze trend gerelateerd is aan de krappe arbeidsmarkt. Omdat het lastiger is om werknemers aan te trekken wordt binnen organisaties als doel gesteld de organisatie aantrekkelijker te maken voor de arbeidsmarkt. Als het selectieproces deze aantrekkelijkheid positief al dan negatief kan beïnvloeden is het van belang om aandacht te besteden aan de perceptie van de kandidaat op de selectiemethoden en -instrumenten. Ook kan de mening van de kandidaat over de relevantie van de tests invloed hebben op de reputatie van de organisatie (Smither, Reilly, Millsap & Stoffey, 1993). Denk hierbij aan zaken als legaliteit en eerlijkheid bij selectieprocedures. Daarnaast kan het zijn dat percepties invloed hebben op de prestatie van de kandidaat (e.g., Weichmann & Ryan, 2003) waardoor er verkeerde selectiekeuzes gemaakt kunnen worden. Om dit te voorkomen is het belangrijk de percepties van kandidaten ten opzichte van selectiemethoden- en instrumenten te onderzoeken.

Bovenstaande trends hebben ervoor gezorgd dat personeelsselecteurs en -adviseurs nieuwe instrumenten zijn gaan ontwikkelen en onderzoeken (e.g., Lievens et al., 2002). Ook het feit dat veel selectiemethoden behoorlijk wat geld en tijd kosten is een vaak genoemd argument voor het zoeken naar nieuwe selectiemethoden (e.g., Dipboye Wooten & Halverson, 2004). Een organisatie zou bij voorkeur gebruik maken van instrumenten waarmee zo veel mogelijk resultaat behaald wordt met zo min mogelijk kosten.

Het ontwikkelen van nieuwe selectie-instrumenten die zowel een voorspellende waarde hebben, als efficiënt in gebruik zijn, die ontwikkelgericht zijn en daarbij ook nog laagdrempelig genoeg zijn voor vele bevolkingsgroepen is een behoorlijke uitdaging. Aan de andere kant biedt de vierde trend - de ontwikkelingen op technologisch gebied - weer veel mogelijkheden (Lievens et al., 2002). Een voorbeeld van deze mogelijkheden is dat de software de resultaten kan berekenen. Dit gaat veel sneller dan het handmatig berekenen van resultaten (Perkins, 1995) en het voorkomt slordigheidsfouten die kunnen ontstaan door vermoeidheid of onoplettendheid (Liu, Papathanasiou & Hao, 2001). Andere voorbeelden van deze mogelijkheden zijn het online afnemen en beoordelen van tests en het gebruik kunnen maken van videomateriaal. Chan en Schmitt (1997) adviseren bijvoorbeeld nader onderzoek te doen naar videotests omdat de uiteindelijke kosten van video-assessment veel lager zullen zijn dan de kosten voor *real-life* simulaties zoals rollenspellen.

De webcamtest sluit qua opzet aan bij deze vier trends. De webcamtest maakt gebruik van veel mogelijkheden die de technologie biedt bij testafname, beoordeling en terugkoppeling en sluit daarmee aan bij de vierde trend. De webcamtest meet competenties. Met de uitslag van de test kunnen ontwikkelpunten worden aangereikt en het opgenomen videomateriaal kan ingezet worden als trainingsmateriaal om de prestatie van de kandidaat te evalueren. Hiermee sluit de test aan op de eerste en tweede trend, waarin omschreven wordt dat de krappe arbeidsmarkt zorgt voor competentie- en ontwikkelgerichte selectiemethoden. Tot slot wordt bij de webcamtest getracht de kandidaat waarheidsgetrouwe praktijksimulaties aan te bieden. De test is gebaseerd op interviews met professionals uit het werkveld en de situaties worden gespeeld door professionele acteurs om de situaties zo realistisch mogelijk te maken. Het is de bedoeling dat de kandidaat de situaties ervaart als situaties die echt in de dagelijkse praktijk kunnen voorkomen en daarmee dus werkelijk de capaciteiten van de kandidaat meet om met deze situaties om te gaan. Hiermee vindt de webcamtest aansluiting bij de derde trend waarin de perceptie van de kandidaat centraal staat.

De webcamtest is een vorm van video-assessment en is niet eerder onderzocht. De centrale vraagstelling van deze scriptie is dan ook: ‘Wat is de voorspellende waarde van de webcamtest voor professioneel sociaal gedrag?’. Eerst beschrijven we de voordelen en nadelen van video-assessment op basis van een aantal onderzoeken (e.g., Lievens & Sackett, 2006). Vervolgens wordt de webcamtest besproken en wordt uitgelegd met welk soort video-assessment de webcamtest het beste vergeleken kan worden en wat de nieuwe aspecten van de webcamtest zijn. Daarna wordt toegelicht welke specifieke webcamtest voor dit onderzoek is gebruikt.

Video-assessment

Bij een video-assessment zoals de webcamtest, worden de opgaven gepresenteerd door middel van videobeelden. Het onderzoek naar video-assessment is schaars, maar de resultaten die er zijn, zijn veelbelovend. In dergelijk onderzoek (e.g., Chan & Schmitt, 1997; Lievens & Sackett, 2006) worden verschillende voordelen van video-assessment besproken en met onderzoek ondersteund.

Zo kan video-assessment *adverse impact* tegengaan (e.g., Weekley & Jones, 1997). *Adverse impact* is, in tegenstelling tot discriminatie, een ongewilde voorkeursbehandeling van meerderheidsgroepen bij personeelsselectie (Cook, 2004). Dipboye, Wooten en Halverson (2004) halen aan dat vooral bij het gebruik van cognitieve capaciteitentests minderheden in de samenleving buiten de boot vallen. Dit zou bijvoorbeeld kunnen komen doordat er bij veel selectie-instrumenten een groot beroep gedaan wordt op leesvaardigheid (Van der Maesen, 2005). Chan en Schmitt (1997) onderzochten de verschillen in prestatie bij studenten bij gebruik van een videotest en het gebruik van een schriftelijke versie van dezelfde test. Zij toonden aan dat het verschil in prestatie tussen blanken en zwarten op de videotests lager is dan bij tests waar gebruik gemaakt wordt van pen en papier.

Een ander voordeel van video-assessment is dat het een positieve invloed heeft op de mening van kandidaten over selectiemethoden en selectie-instrumenten. De waardering van video-assessment is

vaak hoog. O'Reilly, Hubbard, Lessler, Biemer en Turner (1994) bevestigden met hun studie dat proefpersonen ondervraging middels video of audio prettiger vonden dan met behulp van vragenlijsten. Een kanttekening bij hun studie is overigens wel dat zij maar een kleine steekproef hebben gebruikt ($N = 35$). Ook Van der Maesen (2005) stelt dat kandidaten video assessment als een eerlijke en relevante methode beschouwen. Indruksvaliditeit is een term die gebruikt wordt om aan te geven hoe valide de test er op het eerste oog uit ziet (Cook, 2004). Dat wil zeggen hoe relevant de test lijkt voor het meten van de vaardigheden die de test beoogt te meten. In veel onderzoek wordt onderzocht in hoeverre de kandidaat de test indruksvalide vindt. Chan en Schmitt (1997) vonden bij de vergelijking van dezelfde test in schriftelijke vorm en videovorm, dat kandidaten de videotest als meer indruksvalide beschouwden.

Indruksvaliditeit is voor de kandidaat en zijn of haar prestatie en motivatie wellicht belangrijk, maar het is voor onderzoekers en organisaties met name belangrijk om te weten of video-assessment een daadwerkelijk een hoge voorspellende waarde heeft (predictieve validiteit) en bovendien een toegevoegde waarde heeft (incrementele validiteit) ten opzichte van andere selectie-instrumenten (Ryan & Ployart, 2000). Uit een meta-analyse van Salgado en Lado (2000) blijkt dat video-assessment een incrementele validiteit heeft van 31.52% ten opzichte van algemene intelligentie bij het voorspellen van werkprestatie.

Lievens en Sackett (2006) deden onderzoek naar de predictieve en incrementele validiteit van *situational judgment tests* [SJT's]. Een SJT is een test waarbij verschillende situaties worden aangeboden (dat kan schriftelijk, mondeling of in videovorm zijn) en waarbij vervolgens de kandidaat meerdere antwoordopties krijgt. Afhankelijk van de soort SJT moeten de kandidaten of de verschillende antwoordopties individueel beoordelen op effectiviteit, of het beste antwoord kiezen uit de antwoordopties of het antwoord kiezen dat omschrijft hoe de kandidaat zelf zou handelen (McDaniel, Morgeson, Bruhn Finnegan & Campion, 2001; McDaniel, Hartman, Whetzel & Lee Grubb III, 2007). In vergelijking met een normgroep, een expertgroep of een theoretisch model wordt vervolgens iets over de (sociale) vaardigheden van de kandidaat gezegd.

In de studie van Lievens en Sackett (2006) voltooiden 1.159 studenten een SJT in videovorm en 1.750 studenten dezelfde SJT in schriftelijke vorm. Vervolgens werd gekeken wat de voorspellende waarde was van beide testvormen ten opzichte van twee criteria. Het eerste criterium was het gemiddelde cijfer van de student (*grade point average* [GPA]) en het tweede criterium was de prestatie van de student op interpersoonlijke georiënteerde vakken (interpersoonlijke criteria). Zij concludeerden ten eerste dat de video-SJT een grotere voorspellende waarde heeft voor de interpersoonlijke criteria dan de schriftelijke SJT. Ten tweede bleek uit hun stapsgewijze regressieanalyse, dat wanneer de SJT in videoversie was afgenomen, deze extra voorspellende waarde (incrementele validiteit) had ten opzichte van de schriftelijke SJT voor de interpersoonlijke criteria (11%, $p < .01$, $\beta = .34$), maar niet voor GPA. De schriftelijke SJT had daarentegen een significante incrementele validiteit ten opzichte van de video-SJT voor GPA, maar niet voor de interpersoonlijke criteria. Wanneer de SJT in

schriftelijke vorm afgenomen wordt, lijkt deze dus meer te zeggen over de cognitieve capaciteiten van kandidaten dan over hun interpersoonlijke capaciteiten. Wanneer interpersoonlijke capaciteiten gemeten worden, zoals bij SJT's vaak het geval is, is het dus aan te raden dit in videovorm te doen. Lievens en Sackett (2006) vonden overigens in hun onderzoek geen significant verschil in indrukvaliditeit tussen schriftelijke tests en de videotest.

Een nadeel van video-assessment is dat het lastig is om onervaren werknemers te testen met video-assessment omdat toch vaak vakinhoudelijke kennis nodig is om een adequate reactie te geven (Salgado & Lado, 2006). Ook de opstartkosten om een videotest te maken zijn hoog. Denk hierbij aan het inhuren van acteurs en apparatuur (Weekley & Jones, 1997), de kosten voor de softwareapplicaties en het uitgebreide voortraject zoals het maken van een functieanalyse en het schrijven van de scripts. Van der Maesen (2005) zet hier tegenover dat na deze initiële kosten, de operationele kosten gering zijn. De situaties hoeven slechts eenmaal door de acteurs gespeeld te worden. Na de eenmalige productiekosten kan een dergelijke test, op voorbehoud van veroudering, oneindig vaak ingezet worden. Daarnaast hoeven assessoren niet ter plaatse aanwezig te zijn. Tests kunnen op verschillende dagen gemaakt worden door vele kandidaten. Het materiaal wordt vervolgens verzameld en in één keer aangeboden aan de assessoren. Tot slot noemt Van der Maesen dat video-assessment de mogelijkheid biedt om verschillende soorten situaties in korte tijd aan te bieden. Deze diversiteit zorgt voor een grote variatiebreedte. Dat is een voordeel waar in de webcamtest gebruik van wordt gemaakt. De webcamtest kent nog enkele specifieke voordelen ten opzichte van andere selectie-instrumenten in videovorm en deze worden hieronder omschreven.

De webcamtest

De groei van de dienstverlenende sector zorgt ervoor dat er vraag is naar selectietests die sociale eigenschappen meten (Arbeits, 2006). Ondanks de aangetoonde voorspellende waarde van reeds bestaande tests beschrijven Dipboye en collega's (2004) dat organisaties blijven zoeken naar nieuwe methoden omdat veel van de gebruikte meetinstrumenten niet toetsen hoe de werknemer met collega's en klanten omgaat. Een voorbeeld van een test die sociale vaardigheden beoogt te meten is een SJT. Echter, de resultaten van onderzoek naar de voorspellende waarde van de SJT lopen nogal uiteen vanwege de vele verschillende vormen waarin SJT's worden aangeboden (e.g., McDaniel et al., 2001). Zoals hierboven omschreven, zegt een SJT in videovorm bijvoorbeeld meer over sociale vaardigheden dan wanneer de test in geschreven vorm wordt aangeboden (Lievens & Sackett, 2006). Ook correleert een SJT hoger met intelligentie naar mate de situaties minder gedetailleerd worden omschreven (McDaniel et al., 2001). Ook Weekley en Jones (1997) vonden een verband tussen prestatie op videovorm en cognitieve capaciteiten. Daarnaast blijkt dat het afhankelijk is van de instructies welk construct de SJT meet (Ployhart & Ehrhart, 2003; McDaniel et al., 2007). Er kan bijvoorbeeld gevraagd worden of de kandidaat aangeeft hoe hij of zij zelf zou reageren, wat meer zegt over de kandidaat zelf, of gevraagd kan worden wat de kandidaat denkt dat de beste reactie is. Dat laatste zegt

dan meer over kennis die een kandidaat heeft en het inzicht dat hij of zij heeft in sociale situaties. Kortom bij een SJT is het voor een organisatie nog maar de vraag welke eigenschappen er gemeten worden.

De webcamtest kan het beste vergeleken worden met een SJT in videovorm. Immers, net als bij een SJT in videovorm worden verschillende situaties aangeboden in de vorm van filmpjes. Het verschil schuilt hem in het responsegedeelte van de test. Bij de webcamtest moet de kandidaat zelf een reactie geven en kan hij of zij niet kiezen uit verschillende reacties. De SJT meet veelal op kennisniveau (McDaniel et al., 2007), terwijl de webcamtest op gedragsniveau meet. De webcamtest onderscheidt zich daarmee van een SJT. Bij de SJT wordt vooral gevraagd om kennis en inzicht en bij de webcamtest moet de kandidaat ook in staat zijn om de kennis en het inzicht actief toe te passen (Van der Maesen & Koch, 2007).

Toch blijft dan de vraag bestaan waarom niet gewoon gebruik wordt gemaakt van een open reactiemogelijkheid in geschreven vorm. Funke en Schuler (1998) onderzochten wat de toegevoegde waarde van videomateriaal is in een selectietest. De gedachtegang achter hun onderzoek was dat het presenteren van situaties in videovorm zorgt voor een waarheidsgetrouw beeld van de realiteit en dat zou zorgen voor betere prestatie op een test (Funke & Schuler). Om te onderzoeken of het aanbieden van stimuli in videovorm een positieve invloed heeft op de prestatie van de kandidaat, gaten Funke en Schuler dezelfde SJT-inhoud in verschillende testvormen. De stimulus werd ofwel middels video, ofwel gesproken aangeboden aan de kandidaat. De kandidaat kreeg bij beide vormen van presentatie, twee verschillende mogelijkheden om te reageren, namelijk (1) multiple choice en (2) geschreven reactie. Daarnaast werd een situationeel interview ingezet, waarbij de verschillende situaties tijdens een interview in gesproken vorm werden aangeboden en de kandidaat een open mondeling antwoord mocht geven. Als criterium werd een gemiddelde genomen van de beoordeling van een gesproken reactie (opgenomen en achteraf beoordeeld) op zowel de video als de gesproken stimulus.

Funke en Schuler (1998) vonden dat de manier van stimuluspresentatie (video of gesproken) geen effect heeft op de prestatie van een kandidaat. Het situationele interview had de hoogste voorspellende waarde voor het criterium. Funke en Schuler suggereren dat de hogere validiteit van het situationele interview te wijten kan zijn aan de waarheidsgetrouwheid van de responsevorm. Zij adviseren dat ontwikkelaars van videotests aandacht dienen te besteden aan de mate van waarheidsgetrouwheid van de responsemogelijkheden. De webcamtest vindt aansluiting bij dit advies door de kandidaat mondeling te laten reageren en de reactie op te nemen zodat zowel de verbale als non-verbale aspecten beoordeeld kunnen worden. Naar aanleiding van dit onderzoek rijst vervolgens de vraag of er dan niet beter gebruik gemaakt kan worden van een praktijksimulatie zoals een rollenspel of situationeel interview.

Van der Maesen (2005) benadrukt in dit kader het belang van de standaardisatie voor zowel de beoordelaar als de kandidaat. De beoordelaar heeft uitgebreide, gestructureerde en gekaderde scoringsinstructies. In de beoordelingsmodule staan de competenties met definities en algemene

indicatoren. Iedere situatie heeft twee competentiegerelateerde gespreksdoelen en meerdere positieve en negatieve specifieke situatie- en competentiegerelateerde indicatoren. Een beoordelaar van de webcamtest kan de reacties van de kandidaat meerdere malen terug zien en beoordelen aan de hand van de instructies zonder beïnvloed te worden door medebeoordelaars. Een rollenspel kan uiteraard ook gefilmd worden en achteraf beoordeeld worden, doch is het schrijven van beoordelingsinstructies voor een rollenspel een stuk complexer. Een rollenspel kan veel verschillende kanten opgaan, iets wat beïnvloed wordt door zowel de reacties van de kandidaat als de tegenreacties van de acteur. Bij de webcamtest hoeft enkel de gespreksaanzet van de kandidaat beoordeeld te worden, het schrijven van scoringsinstructies voor deze aanzet is een stuk minder complex.

Voor de kandidaat heeft de standaardisatie vooral als voordeel dat verkapte, onbewuste of ongewilde procedurele onrechtvaardigheid vermeden wordt. De webcamtest maakt bijvoorbeeld alleen gebruik van gesproken instructies en items die ondersteund zijn met videomateriaal. Eventuele nadelen die mensen met leesachterstand zouden kunnen ondervinden bij traditionele tekstuele tests, worden hiermee uitgeschakeld. Daarnaast kan de houding van de assessoren geen invloed hebben op de prestatie van de kandidaat. Ter illustratie van dit voordeel is het interessant om het wat oudere, maar nog steeds relevante onderzoek van Word, Zanna en Cooper uit 1974 (zoals beschreven in Kunda, 1999) aan te halen.

Word, Zanna en Cooper (1974, zoals beschreven in Kunda, 1999) voerden een onderzoek uit naar de invloed van de houding van de assessor op de daadwerkelijke prestatie van de kandidaat. Als eerste stap in hun onderzoek observeerden zij interviewers die zowel blanke als donkere mensen interviewde voor een functie. Zij schreven op basis van de observaties twee scripts, een *'black-treatment'* (minderheidsbehandeling) en een *'white-treatment'* (normale behandeling). Bij de minderheidsbehandeling waren de interviewers onder andere afstandelijker en maakten zij meer spreekfouten. De tweede stap was dat blanke studenten van Princeton willekeurig ofwel de minderheidsbehandeling ofwel de normale behandeling kregen van de interviewers tijdens een gefingeerd sollicitatiegesprek. Onafhankelijke beoordelaars beoordeelden de studenten op hun prestatie tijdens dit sollicitatiegesprek. Hieruit bleek dat de studenten die de minderheidsbehandeling kregen slechter presteerde dan de studenten die de normale behandeling kregen. Dit fenomeen wordt tegengegaan bij de webcamtest. Iedere kandidaat van de webcamtest krijgt namelijk precies dezelfde gesproken testinstructies, gesproken inleidingen en geacteerde situaties te zien. Het gedrag, de vooroordelen en de stemming van de acteur, kunnen de prestatie van de kandidaat niet beïnvloeden.

De webcamtest is inmiddels ontwikkeld voor verschillende vakgebieden. In dit onderzoek gaat de aandacht uit naar de webcamtest voor het meten van leidinggevende competenties. Hieronder wordt dieper ingegaan op de webcamtest-leidinggeven en de competenties die gemeten worden met deze webcamtest.

De webcamtest-leidinggeven

Wanneer de zoekopdracht “test voor leidinggeven” ingevoerd wordt op internet verschijnen rond de 288.000 hits. Het is lastig om uit het grote aanbod een goede keuze te maken en daarom heeft het Nederlands Instituut voor Psychologen [NIP] de Commissie Testaangelegenheden Nederland [COTAN] in het leven geroepen. De COTAN beoordeelt psychologische tests op onder andere betrouwbaarheid en validiteit (NIP, 2008). In hun overzicht van goedgekeurde tests is geen één test voor leiderschapskwaliteiten opgenomen (Smit, 2006). In 2003 is wel een test gekeurd, maar deze voldoet niet aan de betrouwbaarheid- en criteriumeisen (NIP, 2008). Het lijkt vreemd dat er geen COTAN-test is voor leiderschap, maar toch is hier wel een verklaring voor.

In de afgelopen decennia zijn al ontelbare voorstellen geweest om goed leiderschap uit te leggen en te identificeren aan de hand van competenties, leiderschapsstijlen of persoonlijkheidseigenschappen (Judge, Piccolo en Ilies 2004; Vroom & Jago, 2007). De vraag is of een goede leider over bepaalde karaktereigenschappen bezit, bepaalde geloofsovertuigingen heeft, binnen een bepaalde context moet passen of simpelweg voor het juiste gedrag in de juiste situatie kiest. Zolang er onduidelijkheid is over de indicatoren van goed leiderschap, is het lastig om een test te maken die een potentiële goede leider aanwijst.

Vroom en Jago (2007) geven een weergave van de geschiedenis van de definities en theorieën inzake leidinggeven. Deze theorieën zijn te verdelen in drie stromingen: (1) de persoonlijkheidstheorieën, (2) de gedragstheorieën en (3) de conditieafhankelijke (*contingency*) theorieën. De persoonlijkheidstheorieën stellen dat de leidinggevende over een set karakteristieken bezit, die kenmerkend zijn voor al dan niet goed leiderschap. Voorbeelden hiervan zijn intelligentie, zelfverzekerdheid, supervisiecapaciteiten, ambitie, behoefte aan vervulling, behoefte aan macht, effectieve interpersoonlijke interactie met ondergeschikten/ communicatieve vaardigheden, competitiedrang, assertiviteit, stressbestendigheid en tolerantie van onzekerheid. (e.g., Ghiselli, 1971; McClelland, 1967, 1975b; Miner, 1978; Bray & Campbell, 1974, zoals beschreven in Vroom & Jago, 2007 en McKenna, 2000).

De gedragstheorieën stellen als basis dat de leidinggevende kan kiezen voor een bepaalde leidinggevende stijl die volgens hem of haar het meest effectief is. De gedragsstijltheorieën beschrijven veelal één van de volgende twee dimensies: (1) directief/autocratisch leiderschap versus participatief/democratisch leiderschap en (2) productgerichte oriëntatie versus persoonsgerichte oriëntatie. De vier leidingstijlen van Likert (1967, zoals beschreven in McKenna, 2000) bijvoorbeeld, geven de eerste dimensie weer. Volgens Likert kiest de manager uit één van de volgende leiderschapsstijlen: (1) uitbuitend autocratisch, (2) welwillend autocratisch, (3) consultatief en (4) deelnemend. De theorieën uit de Ohio State University Studies (1940's and 1950's) omschrijven de tweede dimensie. In deze studies (e.g., Fleishman & Harris, 1962, zoals beschreven in Vroom & Jago, 2007) werd beschreven dat een leidinggevende met een persoonsgerichte houding zich met name richt op het creëren en onderhouden van goede relaties. Deze leidinggevende heeft aandacht voor de

persoonlijke wensen en de capaciteiten van de medewerkers. Een leidinggevende met een taakgerichte houding is meer geneigd om te werken aan de hand van regels en procedures. Naar aanleiding van deze conclusies werd duidelijk dat leiderschap uit twee componenten bestaat: *Initiation structure* (taakgericht leidinggeven) en *Consideration* (persoonsgericht leidinggeven).

Tot slot valt onder de theorieën over leidinggeven de stroming van de conditieafhankelijke (*contingency*) theorieën. In deze theorieën wordt het belang van de situatie benadrukt. De leidinggevende moet in staat zijn de juiste gedragsstijl te kiezen in iedere situatie. Een voorbeeld van zo'n theorie is de *Least Preferred Co-Worker [LPC] scale* van Fiedler (1967, zoals beschreven in McKenna, 2000). Fiedler benadrukt dat het niet gemakkelijk is om karaktereigenschappen te wijzigen, maar dat het wel realistisch is om de meest gepaste situatie voor de desbetreffende leider op te zoeken. Er moet bepaald worden of de leider taak of persoonsgeoriënteerd is. Hiervoor wordt gebruik gemaakt van de LPC-scale. Een hoge score betekent dat de leidinggevende persoonsgericht is en een lage score betekent dat hij of zij taakgericht is. Het idee is dat iemand die taak georiënteerd is effectief functioneert in zeer gunstige of zeer ongunstige situaties. En een manager die persoonsgeoriënteerd is, is effectief bij een gemiddelde situationele gunstigheid.

De conditieafhankelijke theorieën beschrijven dat het niet zo zeer gaat om de stijl van leidinggeven of de karaktereigenschappen van een leidinggevende, maar om de combinatie van de keuzes en het karakter van de manager met de context (Vroom & Jago, 2007). De ene situatie vergt niet dezelfde leidinggevende stijl als de andere situatie. Dit maakt het lastig te bepalen of een testkandidaat geschikt is aan de hand van slechts een vragenlijst voor leiderschapsstijlen, omdat de situationele factoren altijd in ogenschouw genomen moeten worden.

In de webcamtest-leidinggeven is een poging gedaan om door middel van de situatiespecifieke competentie-indicatoren, zowel rekening te houden met de kwaliteiten van de deelnemer als met de situatie. Iedere situatie lokt bepaalde gedragingen uit die de mate waarin een kandidaat de desbetreffende competentie beheerst weergeeft. Dit wordt gemeten aan de hand van de indicatoren die altijd passen binnen de definitie van een competentie, maar per situatie verschillen.

In de webcamtest-leidinggeven is, in lijn met bovengenoemde Ohio State Studies, gekozen voor de competenties persoonsgericht leidinggeven en taakgericht leidinggeven. De keuze voor deze twee competenties uit het enorme aanbod van theorieën is tweeledig. Allereerst is de webcamtest een gedragstest en geen persoonlijkheidstest of een test voor voorkeuren van leidinggeefstijlen. Daarom is gekozen voor een gedragstheorie. Het voordeel van deze specifieke gedragstheorie is dat verschillende onderzoekers zoals Fiedler (1967, zoals beschreven in McKenna, 2000) deze theorie gebruikt hebben in de conditieafhankelijke theorieën. Het principe dat de ene situatie een meer persoonsgerichte houding vraagt en de andere situatie een meer taakgerichte houding wordt in de webcamtest vertaald doordat gebruik gemaakt wordt van situatiespecifieke competentie-indicatoren. Ten tweede blijkt uit de meta-analyse van Judge, Piccolo en Ilies (2004) dat de componenten taakgericht leidinggeven en persoonsgericht leidinggeven lange tijd ten onrechte opzij zijn geschoven. In hun meta-analyse hebben

zij met behulp van statistische correcties 163 correlaties met *Consideration* en 159 correlaties met *Initiating Structure* bekeken. Hieruit bleek dat deze competenties beide hoog correleren met leiderschapsuitkomsten zoals de motivatie van de medewerker en de effectiviteit van de leider. Daarnaast blijken beide competenties duidelijk losse constructen te zijn. Judge en zijn collega's benadrukken de indrukwekkende consistentie van deze resultaten bij verschillende criteria en metingen, in verschillende bronnen en over een lange tijdsspanne.

Het huidige onderzoek

Dit onderzoek heeft ten doel om de predictieve en incrementale validiteit van de webcamtest te onderzoeken. Hiervoor wordt een groep psychologiestudenten aan de Erasmus Universiteit te Rotterdam benaderd. Deze studenten maken de webcamtest, een intelligentietest, een persoonlijkheidsvragenlijst, een video-SJT en vullen een perceptievragenlijst in. Als criteria worden de gemiddelden genomen van tentamencijfers (bloktoetscijfers) en van cijfers op professioneel gedrag. Het cijfer voor professioneel gedrag wordt iedere vijf weken bepaald door verschillende docenten die de student van dichtbij meemaken tijdens werkgroepen. De docent beoordeelt de student op zijn of haar sociaal professioneel gedrag in de groep. Dit wordt weergegeven in factoren als actieve deelname, vaardigheden als groepsvoorzitter, luistervaardigheden en betrokkenheid. Op basis van bovenstaande bevindingen is de verwachting dat de webcamtest-leidinggeven een goede voorspeller is van sociaal professioneel gedrag en beschouwd wordt door de kandidaat als een indrukswalide test. Daarnaast is de verwachting dat de webcamtest-leidinggeven een toegevoegde waarde heeft naast de andere selectie-instrumenten die ingezet worden. Om dit te onderzoeken wordt een aantal hypothesen getest. In de volgende paragrafen worden de hypothesen theoretisch onderbouwd.

Interne consistentie

Voordat er überhaupt gesproken kan worden van validiteit, is het zaak om eerst de betrouwbaarheid van de webcamtest te bepalen. Hiervoor wordt gebruik gemaakt van Conbrach's Alfa (Drenth & Sijtsma, 2006). Smiderle, Perry en Conshaw (1994) spreken naar aanleiding van hun onderzoek naar de Metropolitan Seattle Video Test – een video SJT - twijfels uit over de betrouwbaarheid van video-assessment. De alfa die zij vonden was .47. Wel geven zij aan dat dit kan liggen aan de inhoudelijke aspecten van de test die zij gebruikt hebben. De resultaten uit het onderzoek van Stricker (1982) tonen een positiever beeld over de betrouwbaarheid van video-assessment.

Stricker (1982) heeft onderzoek gedaan naar een meetinstrument dat lijkt op de webcamtest. Zijn Interpersonal Competence Instrument (ICI) schetst ook verschillende videogepresenteerde situaties waarop de kandidaat moet reageren. De reactie van de kandidaat werd echter alleen met een microfoon opgenomen. Er was dus geen videomateriaal van de kandidaat beschikbaar zoals bij de webcamtest. De reacties van de kandidaten werden door twee onafhankelijke experts beoordeeld. Op basis van de

beoordelingen berekende Stricker per component de interne consistentie met behulp van Conbrach's Alfa (α). Deze betrouwbaarheids-coëfficiënten waren verbazingwekkend hoog voor de acht items, van $\alpha = .74$ tot $\alpha = .82$ met een gemiddelde van $\alpha = .79$ (Stricker). Omdat de webcamtest gebruik maakt van drie beoordelaars in plaats van twee, omdat de webcamtest-leidinggeven uit tien items bestaat in plaats van acht en omdat ook de non-verbale reactie van de kandidaat wordt meegenomen, is de verwachting dat de betrouwbaarheid ten minste even hoog is als de gemiddelde betrouwbaarheid van de ICI ($\alpha \geq .79$).

Hypothese 1 De webcamtest-leidinggeven heeft een hoge interne consistentie ($\alpha \geq .79$).

Beoordelaarsovereenstemming

De overeenstemming tussen beoordelaars met betrekking tot de prestatie van een deelnemer op een gedragstest (zoals een rollenspel, een interview of de webcamtest) is relevant om te onderzoeken in relatie tot de constructvaliditeit (Lievens, 2002) en de accuraatheid van beoordelingen (Lievens, 2001). Uiteraard is het voor zowel de kandidaat als voor de organisatie van belang dat de beoordelaars hetzelfde beeld hebben over de eigenschappen die gemeten worden en hierbij zo veel mogelijk consensus bereiken en zo min mogelijk in beoordelaarsvalkuilen trappen. Zo blijkt bijvoorbeeld uit een meta-analyse van Conway, Jako en Goodman (1995) dat de beoordelaarsovereenstemming van een ongestructureerd interview .34 is, voor een gedeeltelijk gestructureerd interview is de beoordelaarsovereenstemming .56 en voor een gestructureerd interview is de overeenstemming .67. De eerste twee beoordelaarsovereenstemmingen zijn niet bijzonder hoog, de derde is redelijk.

Een webcamtest-deelnemer wordt in het huidige onderzoek door drie onafhankelijke experts beoordeeld op relevante competenties. De beoordelaars ondergaan een training voordat zij de kandidaten mogen beoordelen. Hiervoor wordt gebruik gemaakt van een Frame-of-Reference (FOR) – training. Deze keuze is gebaseerd op de resultaten uit het onderzoek van Lievens (2001). Lievens vergelijkt de schemageoriënteerde FOR -training met de traditioneel gedragsgeoriënteerde assessor trainingen. Het verschil tussen deze vormen van training is, dat de gedragsgeoriënteerde training uitgaat van een bottom-up proces bij een evaluatie (dat wil zeggen: de observator wordt getraind om zonder voorkennis het gedrag te observeren en pas daarna een beoordeling te geven) en de schemageoriënteerde FOR-training gaat uit van een top-down proces, waarbij ervan uit wordt gegaan dat een beoordeling altijd wordt beïnvloed door reeds bestaande cognitieve schema's. Het primaire doel van een FOR-training is dan ook om deze schema's dusdanig te beïnvloeden dat een beoordelaar een neutraal referentiekader ontwikkelt.

De resultaten uit het onderzoek van Lievens (2001) toonden aan dat een training sowieso beter is dan geen training, dat de FOR-training effectiever is dan de gedrags-georiënteerde training en dat FOR-training de constructvaliditeit van een assessment aanzienlijk verhoogt. Lievens en Conway (2001) voegen daaraan toe dat de duur van de training geen invloed heeft op de kwaliteiten van een assessor.

Ook Woehr (1994) toonde aan dat FOR-training zorgt voor een meer accurate beoordeling. Uit zijn meta-analyse bleek dat FOR-training effecten als halo en *leniency* (te mild beoordelen), beter kon reduceren dan andere vormen van training.

Vanwege deze positieve onderzoeksresultaten zal bij de training van de beoordelaars met name gewerkt worden aan hun *frame-of-reference*. Naast het bespreken van de valkuilen van beoordelaars, zoals *leniency* en halo, wordt er in groepsverband geoefend en worden er waarden toegekend aan de schalen. Ook wordt duidelijk uitgelegd dat een kandidaat binnen één situatie heel verschillend kan scoren op de twee relevante competenties. Zoals hierboven omschreven hebben Judge, Piccolo en Iliès (2004) empirisch aangetoond dat leidinggeven twee componenten bevat, namelijk taakgericht leidinggeven en persoonsgericht leidinggeven. De leidinggevende kan in een bepaalde situatie wel goed scoren op de ene competentie en niet op de andere competentie. Zo zou een leidinggevende aandacht kunnen hebben voor de persoon, maar de taak uit het oog kunnen verliezen. Een algehele score zou naar het midden kunnen neigen, wat een slecht beeld zou geven van de capaciteiten van de kandidaat. Echter bij scheiding van de competenties kan een kandidaat bijvoorbeeld hoog beoordeeld worden op ‘persoongericht leidinggeven’ en laag scoren op ‘taakgericht leidinggeven’.

Stricker (1982) vond bij zijn studie naar de ICI redelijke tot hoge beoordelaarsovereenstemming (.53 tot .90: gemiddeld .72). Deze beoordelaars waren weliswaar experts, maar niet getraind in het beoordelen van de reacties. Daarnaast werd geen gebruik gemaakt van gestandaardiseerde scoringsinstructies. Omdat bij dit onderzoek naar de webcamtest de beoordelaars wel getraind worden en omdat de beoordelaars gebruik maken van gestandaardiseerde scoringsinstructies is de verwachting dat de beoordelaarsovereenstemming hoger is dan de gemiddelde beoordelaarsovereenstemming in het onderzoek van Stricker.

Hypothese 2 De Webcamtest-leidinggeven heeft een hoge beoordelaarsovereenstemming, (r ≥ .72).

Predictieve en incrementele validiteit van de webcamtest-leidinggeven

De voorspellende waarde van veelgebruikte selectie-instrumenten verschillen behoorlijk. Het blijkt dat één van de meest gebruikte selectie-instrumenten, het ongestructureerde interview, niet de beste voorspellende waarde heeft voor werkprestatie ($r = .38$, Schmidt & Hunter, 1998)¹. Cook (2004) geeft een overzicht van verschillende onderzoeken waarin de validiteit van het ongestructureerde interview varieert van $r = .20$ tot $r = .33$. Daarentegen heeft het gestructureerde interview een hoge voorspellende waarde ($r = .51$). Echter is het voor veel interviewers – vaak het hoofd van een afdeling

¹ De validiteiten weergegeven in deze alinea zijn de gecorrigeerde validiteiten. Gecorrigeerde validiteit houdt in dat de geobserveerde validiteit gecorrigeerd is voor meetfouten en range restriction. De gecorrigeerde validiteit ligt hoger dan de geobserveerde validiteit (Schmidt & Hunter, 1998).

of een afgevaardigde van P&O - lastig om consequent een gestructureerd interview toe te passen. Daarom is het raadzaam om naast het interview de sollicitanten te onderwerpen aan diverse tests.

Bijvoorbeeld persoonlijkheidsvragenlijsten met een gemiddelde voorspellende waarde van $r = .23$ (Cook, 2004) kunnen worden ingezet. Een noot hierbij is dat wanneer de persoonlijkheid in dimensies van de Big Five wordt weergegeven, de voorspellende waarde voor werkprestatie per dimensie verschilt. Zo zou Consciëntieusheid een hogere voorspellende waarde hebben voor werkprestatie dan Openheid (Cook, 2004). Ook tests voor cognitieve capaciteit, met een voorspellende waarde van ongeveer $r = .51$ (Cook, 2004), worden terecht vaak ingezet. Judge, Higgins, Thoresen en Barrick (1999) vonden in een longitudinale studie dat intelligentie en persoonlijkheid tezamen zowel het subjectieve als het objectieve carrièresucces goed konden voorspellen (respectievelijk $R^2 = .42, p < .01$ en $R^2 = .64, p < .01$).

Het uiteindelijke doel van dit onderzoek is om de voorspellende waarde oftewel de predictieve validiteit van de webcamtest te bepalen. Het is immers van belang om na te gaan of de webcamtest als voorspellend instrument gebruikt kan worden. De onderzoeksperiode is echter te kort om een later meetmoment voor de criteria in te passen. In plaats daarvan wordt gekeken naar de studieresultaten die tot het moment van de testafname behaald zijn. Daarom is het beter om te spreken van gelijktijdigheidvaliditeit, dat alleen verschilt van voorspellende validiteit wat betreft het meetmoment van het criterium (Drenth & Sijtsma, 2006). De tot nu toe behaalde studieresultaten worden weergegeven in een gemiddelde van alle tot nu toe behaalde bloктоetscijfer en een gemiddelde van alle tot nu toe behaalde cijfers op professioneel gedrag.

Het cijfer op professioneel gedrag is zoals hierboven omschreven een beschrijving van het sociaal professioneel gedrag van de student in de werkgroep. De bloктоets is daarentegen een kennistoets. Onderstaande hypothese is gebaseerd op het onderzoek van Lievens en Sackett uit 2006, zoals hierboven omschreven, waaruit bleek dat video-assessment beter de prestatie op sociale vakken voorspelde dan *GPA*.

Hypothese 3 *De voorspellende waarde van de webcamtest-leidinggeven is voor het criterium 'professioneel gedrag' hoger dan voor het criterium 'gemiddeld cijfer'.*

Behalve dat de webcamtest een goede voorspellende waarde moet hebben voor de criteria, is het ook van belang dat de test een toegevoegde waarde heeft ten opzichte van andere selectie-instrumenten. Immers, het heeft weinig zin om de webcamtest te ontwikkelen als een reeds bestaand instrument, zoals de veelgebruikte en goedkopere intelligentietests en persoonlijkheidsvragenlijsten, de voorspelling al geheel dekt. Zoals eerder omschreven hebben video-assessments een incrementele validiteit ten opzichte van schriftelijke assessments (Lievens & Sackett, 2006) en ten opzichte van algemene intelligentie (Saldago & Lado, 2000). Zeker wanneer het criterium een maat is voor sociaal

gedrag (Lievens & Sackett, 2006). De verwachting is daarom dat ook de webcamtest een incrementele validiteit heeft ten opzichte van de andere in dit onderzoek opgenomen meetinstrumenten, te weten de intelligentietest, de persoonlijkheidsvragenlijst en de video SJT.

Hypothese 4 De Webcamtest-leidinggeven heeft een incrementele validiteit ten opzichte van de video SJT, de intelligentietest en de persoonlijkheidsvragenlijst met betrekking tot het criterium 'professioneel gedrag'.

Kandidaat-percepties

Smither, Reilly, Millsap en Stoffey (1993) noemen drie redenen waarom de percepties van kandidaten op tests belangrijk worden bevonden. Zij beschreven de eerste reden als het spillover effect, waarbij de mening van sollicitanten over een organisatie beïnvloed worden door de selectiemethoden of selectie-instrumenten. Als tweede reden gaven zij dat de mening van de kandidaten over de relevantie van de tests invloed kan hebben op de reputatie van de organisatie wat betreft eerlijkheid en legaliteit. Tot slot beschreven zij dat de perceptie van de kandidaat indirect invloed kan hebben op de prestatie van de kandidaat op een test. Een aantal jaar later omschreven Ryan en Ployhart (2001) twee stromingen in het onderzoek naar percepties. Bij de ene stroming gaat de aandacht uit naar de ervaringspercepties van de kandidaat tijdens de procedure en de invloed die dat heeft op de prestatie. Deze stroming sluit aan op de derde reden die Smither en collega's gaven. Wanneer bijvoorbeeld kandidaten een negatieve perceptie hebben op de selectietest(s) zal de prestatie en de motivatie dalen (Lievens et al., 2002).

De andere stroming, die aansluit op de eerste twee redenen die Smither en collega's (1993) gaven, is gericht op het onderzoeken van de mening van de participanten over de procedures en de methoden wat betreft bijvoorbeeld eerlijkheid en welke invloed dat dat heeft op het beeld van de organisatie (Ryan & Ployhart, 2001). De sollicitanten associëren de - voor hun nog onbekende - organisatie met de selectietest(s). Op basis van een overzicht van de perceptieliteratuur van beide stromingen, beschrijven Ryan en Ployhart de determinanten en gevolgen van percepties. Dit overzicht laat zien dat percepties een groot onderzoeksgebied is dat niet gedekt kan worden in deze studie. Wel besteden we in deze studie aandacht aan enkele specifieke percepties met betrekking tot de webcamtest.

Voor deze studie is van belang in welke mate de perceptie van de deelnemer invloed heeft op zijn of haar prestatie. Het doel van deze studie is om te bepalen wat de voorspellende waarde is van de webcamtest. Als de prestatie van de deelnemers op wat voor manier dan ook wordt beïnvloed door de mening van de kandidaat is het zaak te controleren voor deze variabelen. Daarnaast willen we weten hoe de kandidaat denkt over de webcamtest in het kader van de hiervoor omschreven legaliteit en het spillover effect. Daarom onderzoeken we de relaties tussen testangst, computerangst, de voorafgemeten motivatie en *self-efficacy* voor de webcamtest, de achteraf gemeten waardering van de webcamtest (het leuk vinden van de test en de indrukvaliditeit) en de score op de webcamtest. Met betrekking tot deze relaties stellen we enkele specifieke hypothesen.

Testangst en computerangst

Allereerst wordt nagegaan of de kandidaat last heeft van testangst en/of computerangst. Testangst wordt al vanaf het begin van de vorige eeuw onderzocht (Ströber & Pekrun, 2004). De twee componenten van testangst zouden bestaan uit emotionaliteit en het zich zorgen maken (e.g., Cassady & Johnson, 2001), waarbij het cognitieve aspect ‘zorgen maken’ de belangrijkste component is bij testangst (Deffenbacher & Hazaleus, 1985). Daarentegen stelt Hembree (1988) dat een kandidaat zich pas zorgen gaat maken als hij of zij emotionaliteit ervaart. Ongeacht de volgorde van de testangstreactie blijkt een hoge mate van testangst negatief gecorreleerd te zijn met zaken als intelligentie, schoolprestaties en probleemoplossend vermogen (Deffenbacher & Hazaleus, 1985; Hembree, 1988). De invloed die testangst zou hebben op prestatie ligt zo rond de 8% verklaarde variantie (e.g., Cassady & Johnson, 2001). In een recent onderzoek vonden Berke en Nanda (2006) zelfs een verklaarde variantie van 21%. Uit bovenstaande onderzoeken blijkt dat testangst een aanzienlijke invloed heeft op prestatie. Daarom wordt in dit onderzoek aandacht besteed aan de invloed van testangst op de prestatie op de webcamtest.

Net als testangst zou computerangst de prestatie op een test negatief kunnen beïnvloeden (e.g., Heinssen, Glass & Knight, 1987). Het onderzoek van Heinssen en zijn collega's laat zien dat computerangst een fenomeen is dat losstaat van bijvoorbeeld testangst. Thorpe en Bosnan (2007) onderzochten of computerangst misschien een DSM-waardige fobie of angst is. In hun onderzoek kwam naar voren dat computerangst niet gelijk is aan testangst, maar eerder omschreven kan worden als een sociale fobie of angststoornis. Ook pleiten Thorpe en Bosnan om computerangst te beschouwen als een serieuze angst die niet zal verdwijnen naarmate de computer een groter aandeel in het leven van de mens gaat spelen, want zelfs bij sommige kinderen die opgegroeid zijn met computers wordt computerangst geconstateerd. Ook Beckers, Wicherts en Schmidt (2006) sluiten zich hierbij aan. In hun onderzoek probeerden zij de vraag te beantwoorden of computerangst een situatiegebonden toestand is waarin iemand verkeert (*state*) of dat dit een karaktereigenschap (*trait*) is. Beckers en collega's vonden in hun onderzoek een duidelijke aanwijzing dat computerangst een karaktereigenschap is aangezien computerangst meer gerelateerd was aan persoonlijke angst dan aan situationele angst. Computerangst is dus een losstaand fenomeen dat apart onderzocht moet worden naast testangst.

Onderzoek dat de relatie tussen testangst en/of computerangst met prestatie op video-assessment beschrijft is niet in deze concrete termen te vinden. Wel zijn verschillende onderzoeken te vinden over de invloed van testangst op testprestatie. Hembree (1988) bijvoorbeeld vond aan de hand van een meta-analyse van 562 studies dat testangst een negatief effect heeft op testprestatie. Ook Pintrich en De Groot (1990) vonden een direct effect van testangst op prestatie.

Ook is de invloed van computerangst op prestatie op computertests onderzocht. Perkins (1995) vond een significante negatieve correlatie tussen computerangst en testprestatie. Hij deed een onderzoek onder 83 studenten. Alle studenten kregen voorafgaand aan het onderzoek dezelfde schriftelijk test om

zeker te stellen dat de groepen gelijk waren. Vervolgens werden de studenten opgedeeld in twee groepen. De ene groep kreeg een test op de computer en de andere groep kreeg dezelfde test in schriftelijke vorm. Dit was inhoudelijk dezelfde test als de test die voorafgaand aan het onderzoek was afgenomen. De kandidaten die meer computerangst hadden presteerde slechter op de computertest. Ook Brosnan (1998) vond dat computerangst een negatief effect had op het aantal correcte antwoorden dat deelnemers gaven. Hij onderzocht of computerangst een indirect effect had op testprestatie. Hierbij stelde hij dat *self-efficacy* een mediator was tussen computerangst en prestatie. Dit was gedeeltelijk waar, maar computerangst zelf had ook een direct effect op de testprestatie.

Omdat wij willen nagaan in welke mate testangst en computerangst onze onderzoeksresultaten beïnvloeden, onderzoeken we de relatie tussen testangst, computerangst en testprestatie. Op basis van bovenstaande onderzoeksresultaten is de verwachting dat zowel testangst als computerangst een negatieve invloed hebben op de prestatie op de webcamtest.

Hypothese 5 De prestatie op de webcamtest wordt negatief beïnvloed door testangst en door computerangst.

Motivatie

Het meten van testmotivatie is van belang wanneer lage motivatie leidt tot lagere testcores (Wise & DeMars, 2003). Namelijk, als deze testcores zouden veranderen als de motivatie van alle kandidaten gelijk is, zijn de onderzoeksresultaten niet valide bij ongelijke motivatie onder testkandidaten. Deze niet valide testcores kunnen vervolgens onterecht leiden tot niet valide predictieve en incrementele validiteit. Sundre en Wise (2003) toonden aan dat correlaties tussen predictor en criterium stegen wanneer de scores van ongemotiveerde testkandidaten buiten de data werden gelaten.

In dit onderzoek zijn er geen consequenties voor de studenten als zij wel of niet goed presteren op de webcamtest. Het zou kunnen zijn dat sommige studenten daardoor een lagere motivatie hebben om de webcamtest te maken en dit kan vervolgens de resultaten op de webcamtest en daarmee de resultaten van dit onderzoek beïnvloeden. Om deze reden is het van belang om na te gaan of motivatie en testprestatie aan elkaar gerelateerd zijn.

Hypothese 6 Er is een positief verband tussen testmotivatie en prestatie op de webcamtest.

Volgens Ployhart en Ryan (2000) worden in veel onderzoeken perceptievragen pas achteraf gesteld. Dit is vaak ook het geval bij het meten van testmotivatie. De eigen veronderstelde prestatie van de kandidaat op de test heeft dan invloed op de achteraf omschreven motivatie. Ook Arvey, Strickland, Drauden en Martin (1990) beschrijven het probleem met de interpretatie van de causaliteit van motivatie omdat vaak achteraf pas de motivatie wordt gemeten. In deze studie wordt daarom vooraf gemeten of de kandidaat gemotiveerd is om de webcamtest te maken.

Self-efficacy

Ook *self-efficacy*, iemands geloof in zijn of haar eigen vermogen om te kunnen presteren, (Bandura, 1977) wordt vooraf gemeten. In deze context spreken we van test *self-efficacy*. Dit is iemands geloof in zijn of haar eigen vermogen om goed te presteren in een testsituatie (Wiechmann & Ryan, 2003). Test *self-efficacy* zou positief gecorreleerd zijn aan motivatie en prestatie, maar negatief aan testangst (Ryan, Ployhart, Greguras & Schmit, 1998). Pintrich en De Groot (1990) vonden een direct en een indirect effect van *self-efficacy* op prestatie. Zij voerden een regressieanalyse uit waaruit bleek dat bij bepaalde taken studenten door *self-efficacy* meer cognitieve toewijding vertoonden en door die cognitieve toewijding beter presteerden. Brosnan (1998) vond dat *self-efficacy* zorgde voor het gebruik van betere technieken om vraagstukken op te lossen. Indirect zorgde *self-efficacy* hier voor een betere prestatie. Ondanks dat de relatie (direct of indirect) tussen *self-efficacy* en prestatie veelal is aangetoond, blijft een discussie bestaan over de causaliteit (Richard, Diefendorff & Martin, 2006).

Richard en collega's (2006) onderzochten de relatie tussen *self-efficacy* en prestatie. Hun onderzoek bestond uit twee studies. Bij de eerste studie kregen studenten (N =83) voorafgaand aan hun tentamens een vragenlijst over hun eigen *self-efficacy*. Zij werden gedurende vier tentamens gevolgd. In de tweede studie kregen 197 studenten herhaaldelijk een computertaak waarin zij door oefening steeds beter werden. Ook hier werd *self-efficacy* voorafgaand aan de taken gemeten. Uit beide studies bleek een significante relatie tussen *self-efficacy* en prestatie. Ook bleek dat de *self-efficacy* positief gecorreleerd was met voorgaande prestatie, maar niet met opvolgende prestatie.

In dit onderzoek wordt voorafgaand aan het maken van de webcamtest gevraagd hoe de kandidaten verwachten dat zij de test zullen gaan maken. Uitgaande van het onderzoek van Richard et al. (2006) waarin werd aangetoond dat *self-efficacy* een gevolg is van prestatie en het feit dat de webcamtest een nieuwe test is, is de verwachting dat *self-efficacy* en de score op de webcamtest niet gecorreleerd zijn.

Hypothese 7 Self-efficacy is niet gecorreleerd met prestatie op de webcamtest

Indruksvaliditeit

De webcamtest is nog niet bekend onder selectietest-kandidaten en kandidaten kunnen daarom nog geen algemene mening over de webcamtest hebben zoals zij bijvoorbeeld wel kunnen hebben over intelligentietests. Zoals hierboven omschreven besteden organisaties om diverse redenen meer en meer aandacht aan de perceptie van de kandidaat op de selectieprocedure en selectietest (e.g., Ryan & Ployhart, 2000). Zo zou het bijvoorbeeld kunnen zijn dat het verschil in prestatie op diverse tests tussen meerderheids- en minderheidsgroepen ligt aan het verschil in attitude die deze groepen hebben ten opzichte van de tests (Ryan & Ployhart). Als de kandidaat bij voorbaat al een negatief beeld zou hebben van de webcamtest, zou de motivatie en de prestatie kunnen dalen. Daarom is het interessant om voor potentiële gebruikers van de webcamtest te weten of de webcamtest wordt beschouwd als een indruksvalide test ten opzichte van andere (veelgebruikte) selectietests.

Uit het hierboven omschreven onderzoek van Chan en Schmitt (1997) bleek dat de kandidaten de video-SJT meer indrukvalide vonden dan de dezelfde SJT afgenomen met pen en papier. Cartney (2005) heeft een kwalitatief onderzoek gedaan naar de reacties van kandidaten op een nieuw meetinstrument binnen een sociaal werk programma. De methode om communicatieve vaardigheden te meten veranderde van het schrijven van een essay naar het afnemen van video-interviews. Haar conclusie was dat de meeste deelnemers een positief beeld hadden van deze nieuwe meetmethode en deze methode beschouwde als een betere manier om communicatieve vaardigheden te meten. Ook Richman-Hirsch, Olson-Buchanan en Drasgow (2000) onderzochten of de manier van testafname invloed heeft op de perceptie van de kandidaat op de test. Zij vergeleken de percepties van de kandidaten op een inhoudelijk gelijke test die of schriftelijk werd afgenomen, of met een computerprogramma of met een multimedia-instrument. Zij vonden dat er geen significant verschil was tussen de mening over de schriftelijke methode en de mening over de afname met de computer. Wel was er een significant verschil tussen deze twee afnamemethoden en het multimedia-instrument. De kandidaten, 131 managers van verschillende organisaties, vonden het multimedia-instrument meer indrukvalide ($F(2,127) = 2.74, p < .05$). Richman-Hirsch et collega's benadrukken dat het enkel het aanbieden van een test in computervorm niet genoeg is om verschil te maken, het integreren van multimedia heeft een toegevoegde waarde. De video-SJT en de webcamtest zijn beide multimedia instrumenten. Op basis van de hierboven omschreven onderzoeken is de verwachting dat de webcamtest als meer indrukvalide wordt beschouwd dan de intelligentietest en de persoonlijkheidsvragenlijst en tenminste zo indrukvalide als de video-SJT voor het meten van leidinggevende kwaliteiten.

Hypothese 8 De webcamtest wordt door de kandidaat als een meer indrukvalide test beschouwd dan de intelligentietest en de persoonlijkheidsvragenlijst en als een tenminste even indrukvalide test als de video-SJT voor het meten van leidinggevende kwaliteiten.

Het leuk vinden van de test

Liu, Papathanasiou en Hao (2001) onderzochten op kwantitatief en kwalitatief niveau wat studenten vonden van multimediatests en door welke factoren deze mening beïnvloed werd. Zij hadden twee groepen studenten binnen hun onderzoekspopulatie. De ene groep volgde onderwijs online (N = 97) en de andere groep volgde onderwijs in een schoolsetting (N = 100). Van de groep studenten die klassikaal les kregen vond 44% de multimediatest het leukste, 30% prefereerde de pen en papier versie en 26% had geen voorkeur. Bij de groep die online onderwijs volgde waren de percentages respectievelijk 65%, 24% en 11%. De belangrijkste redenen om de multimediatest leuker te vinden waren (1) de directe feedback die beschikbaar was en (2) dat er gebruik gemaakt werd van plaatjes, video, audio en animaties.

Ook uit het hierboven aangehaalde onderzoek van Richman-Hirsch en collega's (2000) bleek het resultaat dat kandidaten de multimediatest leuker vonden om te doen dan de schriftelijke test en de test op de computer ($F(2,127) = 5.25, p < .05$). Op basis van deze onderzoeken is de verwachting dat de kandidaten de webcamtest leuker vinden om te doen dan de intelligentietest en de persoonlijkheidsvragenlijst en tenminste even leuk vinden om te doen als de video-SJT.

Hypothese 9 De webcamtest wordt door de kandidaat als een leukere test beschouwd dan de intelligentietest en de persoonlijkheidsvragenlijst en als een tenminste even leuke test als de video-SJT.

Methode

Participanten en Procedure

Voor dit onderzoek zijn psychologiestudenten van de Erasmus Universiteit Rotterdam benaderd. De vierdejaars studenten ($N=35$) zijn in het kader van hun studierichting arbeids- en organisatiepsychologie, uitgenodigd voor een assessmentdag bij GITP. Op deze dag hebben zij onder andere alle tests voor dit onderzoek gemaakt. De eerste, tweede en derdejaars studenten ($N=71$) zijn benaderd via het intranetkanaal van de faculteit psychologie: Psyweb. Zij konden in ruil voor drie proefpersoonuren meedoen aan het onderzoek op de Erasmus Universiteit in Rotterdam. Proefpersoonuren zijn verplichte uren die studenten moeten besteden aan het vrijwillig meewerken aan het onderzoek van psychologen. Verder verloopt de afnameprocedure hetzelfde als bij de vierdejaars studenten. Uiteindelijk hebben 106 studenten meegedaan (37 mannen en 69 vrouwen, zie Tabel 1). De gemiddelde leeftijd was 21,94 jaar ($SD = 2.85$).

Vijftig studenten zaten in het eerste jaar, 20 studenten in het tweede jaar en 36 studenten in het derde jaar van de studie psychologie in de periode van testafname (zie Tabel 1). Bij de dataverzameling is onderscheid gemaakt tussen opleidingsniveau en studiejaar. Bij opleidingsniveau is nagegaan welk diploma de kandidaat in zijn of haar bezit heeft, een bachelordiploma of een VWO-diploma. In de praktijk komt het erop neer dat de meeste vierdejaarsstudenten een bachelordiploma hebben en de eerste- en tweedejaars studenten in de meeste gevallen alleen een VWO-diploma hebben, met uitzondering van de enkele studenten die reeds een HBO-opleiding hebben voltooid. De correlatie tussen opleidingsniveau en studiejaar is erg hoog ($r = .94, p < .01$) en er is voor gekozen om in de analyses enkel gebruik te maken van de variabele studiejaar. De onderzoeksgroep bevatte 98 studenten die in Nederland zijn geboren, 7 studenten zijn in een ander land geboren en 1 student heeft de vraag met betrekking tot geboorteland niet beantwoord (zie Tabel 1).

Tabel 1. Steekproef

	Geslacht		Studiejaar			Geboorteland		
	man	vrouw	1	2	4	Nederland	Elders**	Onbekend
Aantal (%)*	37 (35)	69 (65)	50 (47)	20 (19)	36 (34)	98 (92)	7 (7)	1 (1)

* Percentages zijn afgerond naar hele getallen.

** Geboorteland: Vietnam 1x, Afghanistan 1x, Engeland 1x, Suriname 1x, Rusland 1x, Amerika 1x, Turkije 1x.

In Tabel 2 staat hoeveel werkervaring en hoeveel leidinggevende ervaring de studenten hebben. Van de studenten heeft 62% tenminste één jaar werkervaring, 9% heeft tenminste een jaar leidinggevende ervaring. Zesendertig procent heeft minder dan één jaar werkervaring of geen werkervaring en 89% heeft minder dan één jaar of geen leidinggevende ervaring. Voor de verwerking in de resultaten zijn deze categorieën respectievelijk ingedeeld in een schaal van 1 tot 5.

Tabel 2. Werkervaring en leidinggevende ervaring onder studenten

	Aantal studenten (%)*	
	Werkervaring	Leidinggevende ervaring
Geen	22 (21)	86 (81)
Minder dan 1 jaar	16 (15)	7 (7)
1-5 jaar	47 (44)	9 (8)
6-10 jaar	18 (17)	1 (1)
Meer dan 10 jaar	1 (1)	0 (0)
Onbekend	2 (2)	3(3)

* Percentages zijn afgerond naar hele getallen.

Materiaal

Hieronder omschrijven we de tests die voor dit onderzoek zijn gebruikt. Van iedere test wordt informatie verstrekt inzake de inhoud, testgrootte, procedure, betrouwbaarheid en schaling. Ook wordt indien mogelijk een voorbeeld gegeven. Achtereenvolgens beschrijven we de webcamtest, de video-SJT, de intelligentietest, de persoonlijkheidsvragenlijst en de verschillende schalen van de perceptievragenlijst. De volgorde van de afname van de tests kon per kandidaat verschillen. Er werd bijgehouden in welke volgorde iedere kandidaat de tests maakte. De studenten kregen allen gelijke instructies voorafgaand aan de tests. Voorafgaand aan de deelname kregen de kandidaten enkele algemene vragen over onder andere computerangst en testangst. Daarnaast kregen de kandidaten voorafgaand aan iedere test en na afloop van iedere test enkele testspecifieke vragen die onder andere betrekking hadden op motivatie voor en waardering van de tests.

Webcamtest

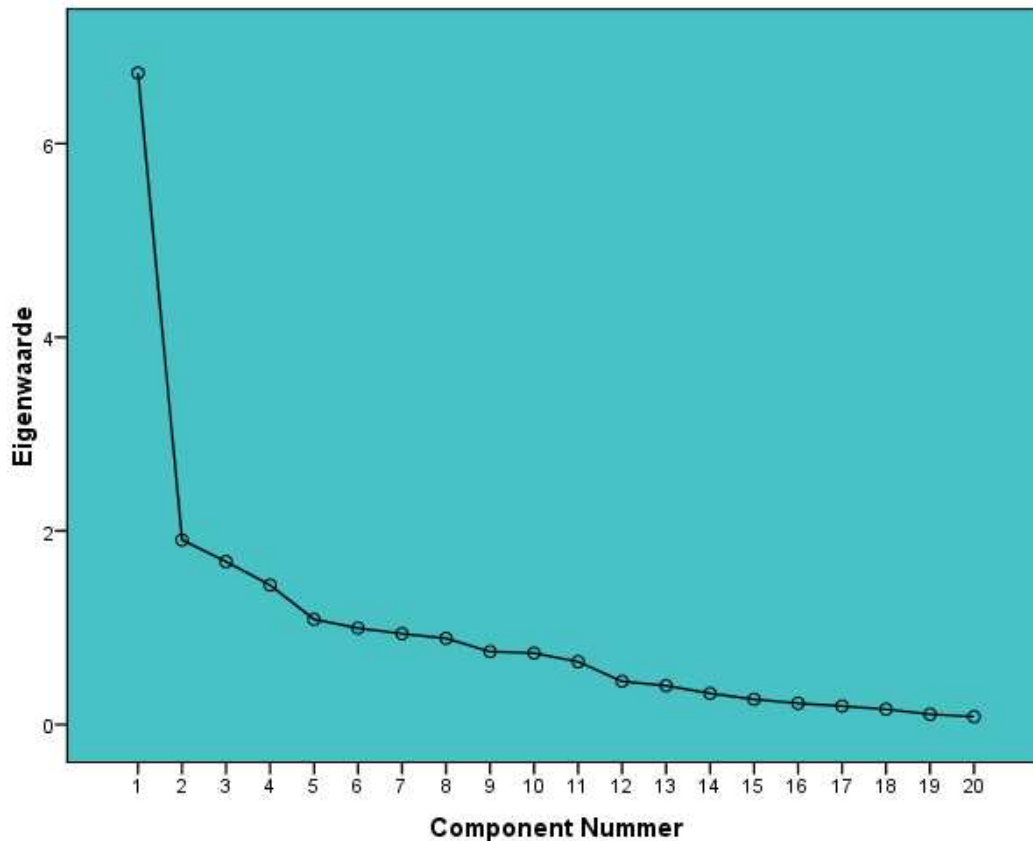
Voor dit onderzoek werd gebruik gemaakt van de webcamtest die speciaal is ontwikkeld voor het meten van leidinggevende kwaliteiten: *De webcamtest-leidinggeven*. De test bestaat uit 10 items. Op advies van vier adviseurs waren 10 situaties geselecteerd uit 20 situaties. De selectie was afhankelijk van de beoordeling van de situaties door de adviseurs. De adviseurs vulden per situatie een vijfpuntsschaal in op drie criteria: (1) persoonsgericht leiderschap, (2) taakgericht leiderschap en (3) praktische relevantie. De uiteindelijke selectie bevat situaties die hoog scoorde op alle drie deze criteria.

De webcamtest (Van der Maesen, 2005) is een meetinstrument dat gemaakt is om sociale competenties in realistische situaties te meten. De test vraagt om een mondelinge spontane reactie. De testkandidaat zit voor het computerscherm en krijgt eerst een uitgebreide uitleg van een acteur. Daarna krijgt de kandidaat de gelegenheid om te oefenen alvorens de test begint. Per situatie krijgt de kandidaat eerst een gesproken inleiding te horen. Vervolgens verschijnt er een acteur op het scherm die rechtstreeks tegen de kandidaat iets zegt. Daarna is het de bedoeling dat de kandidaat een reactie geeft in de webcam. Deze reactie wordt opgenomen en achteraf door drie onafhankelijke experts (e.g., psychologen) aan de hand van algemene en specifieke situatiegerelateerde indicatoren beoordeeld op twee voor de situatie relevante competenties.

In dit onderzoek werd de webcamtest achteraf beoordeeld door drie getrainde vierdejaars psychologie of sociologie studenten. Deze beoordelaars zijn getraind met een FOR-training. Deze training bestond uit twee bijeenkomsten. Tijdens de eerste bijeenkomst kregen de beoordelaars praktische instructies voor het gebruik van de beoordelingsmodule en gewerkt werd aan het referentiekader van de beoordelaars. De definities van competenties werden uitgebreid besproken, de indicatoren die bij de competenties en situaties hoorden werden doorgenomen en er is in groepsverband geoefend met het toebedelen van indicatoren aan en het waarderen van voorbeeldreacties. Tussen de eerste en de tweede bijeenkomst in kregen de beoordelaars proefbeoordelingen. Berekend werd hoe hoog de beoordelaarsovereenstemming was en hoe de individuele beoordelaartendentie (bijvoorbeeld mild, streng, vermijden van extremen) was. Tijdens de tweede bijeenkomst werden deze resultaten besproken en werden situaties behandeld waarvan de beoordelaarsovereenstemming laag was. Ook kregen de beoordelaars de gelegenheid om voorstellen te doen ter verbetering van de beoordelingsinstructies. Na deze training waren de beoordelaars bevoegd om te beoordelen. Zij kregen vanaf dat moment iedere week 10 kandidaten toegewezen die zij moesten beoordelen. Het beoordelen gaat op een schaal van één tot vijf, de score wordt omgerekend tot een score van 1 tot 100 door te vermenigvuldigen met 20.

De webcamtest beoogt de competenties persoonsgericht en taakgericht leidinggeven te meten. In de webcamtest correleren deze competenties hoog met elkaar ($r = .80, p < .01$). Naar aanleiding van een principale componenten analyse met varimax rotatie is ervoor gekozen om de situaties in de webcamtest niet op te splitsen in twee competenties. Zelfs bij het forceren van twee componenten

blijven de resultaten erop duiden dat er geen sprake is van twee competenties. Ondanks dat de eerste vijf componenten een eigenwaarde hebben van meer dan één (respectievelijk $EW1 = 6,73$; $EW2 = 1,91$; $EW3 = 1,68$; $EW4 = 1,44$; $EW5 = 1,09$), is zichtbaar in Figuur 1 dat de webcamtest slecht één competentie lijkt te meten. Deze competentie heeft een verklaarde variantie van 33.65%. In de discussie wordt dieper ingegaan op dit gegeven. Voor het onderzoek wordt de score van een kandidaat op beide competenties per situatie samen genomen als één score. Als totaalscore wordt een gemiddelde score over alle situaties op de webcamtest berekend.



Figuur 1. Componenten in de webcamtest.

SQ-test

SQ-leidinggeven is een video-SJT die inzicht in de vaardigheden van een leidinggevende meet. De kandidaat zit voor de computer. Zowel de situaties als de reacties worden in videovorm aangeboden. Iedere kandidaat krijgt voorafgaand aan de test dezelfde uitleg gegeven door een acteur. Daarna krijgen de kandidaten de gelegenheid om te oefenen alvorens de test begint. De test bestaat uit 17 situaties met een alfa van .75 (Van der Maesen & Koch, 2007). Bij elke situatie worden vier reacties gegeven die onafhankelijk van elkaar beoordeeld moeten worden op effectiviteit door de kandidaat op basis van een vijfpuntsschaal. De beoordeling van de kandidaat wordt vergeleken met de beoordeling van een expertgroep. Hoe hoger de overeenstemming, hoe hoger de score van de kandidaat op de test. De score wordt weergegeven op een tienpuntsschaal.

Intelligentietest

De *Captain test* is een intelligentietest voor mensen met een HBO opleiding of hoger ontwikkeld door GITP. Met deze test wordt beoogt de cognitieve capaciteiten ofwel de algemene intelligentie van de kandidaat te meten (GITP, 2006). De test bestaat uit 3 delen: (1) verbale analogieën (39 items, $\alpha = .90$), (2) figuren (30 items, $\alpha = .78$) en (3) cijferreeksen (25 items, $\alpha = .59$). Voor ieder correct item krijgt de kandidaat één punt. De test wordt afgenomen onder een tijdslimiet. De kandidaat krijgt geen punten voor de niet gemaakte opgave. Ook krijgt de kandidaat tijdens de test geen feedback over de prestatie op de onafhankelijke items. Voor dit onderzoek gebruiken we de totaalscore op de drie onderdelen, dit is een maximale score van 94 punten.

Persoonlijkheidsvragenlijst

De GITP Big Five persoonlijkheidsvragenlijst: G5-R (Koch, 2003, zoals beschreven in GITP, 2006) bestaat uit vijf constructen: (1) extraversie (27 items, $\alpha = .92$), (2) vriendelijkheid (28 items, $\alpha = .83$), (3) consciëntieusheid (40 items, $\alpha = .92$), (4) stabiliteit (23 items, $\alpha = .88$), en (5) openzinnigheid (47 items, $\alpha = .90$). GITP (2006) verstrekt testinformatie waarin deze vijf constructen als volgt worden omschreven: (1) extraversie: “open persoonlijkheid, behoefte hebben aan gezelschap, gemakkelijk in de omgang zijn, joviaal, gemakkelijk contacten leggen, het prettig vinden om in de belangstelling te staan”; (2) vriendelijkheid: “behulpzaam, begaan zijn met andere mensen, goedaardig, vertrouwen hebben in anderen, verdraagzaam”; (3) consciëntieusheid: “zich aan regels houden, gedisciplineerd zijn, taken volbrengen, volhardend zijn, spullen opruimen, ordelijk en planmatig te werk gaan, van hard werken houden, zich verantwoordelijk voelen”; (4) stabiliteit: “stevige persoonlijkheid, zelfverzekerd zijn, tegen spanningen en tegen kritiek kunnen, evenwichtig, kalm en geduldig zijn”; (5) openzinnigheid: “open staan voor meningen, gevoelens en nieuwe ervaringen, een onderzoekend persoon zijn, verbeeldingskracht hebben, over dingen nadenken, zich een eigen mening kunnen vormen” (GITP, 2006, p. 2)

Na onderzoek op basis van afname bij meer dan 20.000 kandidaten, bleek dat er naast de Big Five een extra overkoepelend construct toegevoegd moest worden. Dit construct heeft GITP zelfpresentatie genoemd (35 items, $\alpha = .93$). Zelfpresentatie is “de gretigheid en de ambitie waarmee men zichzelf wil presenteren, zichzelf als krachtig, ambitieus, energiek en zelfverzekerd neerzetten” (GITP, 2006, p. 3). Ieder construct wordt beoordeeld op een vijfpuntsschaal.

Kandidaatpercepties

Testangst en computerangst werden gemeten voordat de kandidaat begint aan het onderzoek. Motivatie en *self-efficacy* werden gemeten na de uitleg over de te maken test, maar voordat de kandidaat de test daadwerkelijk maakte. Indruksvaliditeit en hoe leuk de kandidaten het maken van de test vonden werden gemeten nadat de desbetreffende test is gemaakt. De vragen worden door de kandidaat beantwoord op een vijfpunts Likertschaal (1 = helemaal niet akkoord, 2 = niet akkoord, 3 =

neutraal, 4 = akkoord, 5 = helemaal akkoord). Ook werd een aantal vragen in omgekeerde vorm gesteld. Hieronder volgt een omschrijving van de verschillende constructen die met deze perceptievragenlijst worden gemeten.

Testangst is een interfererende angst tijdens een evaluatiemoment die zorgt voor een combinatie van verhoogde fysiologische activiteit (emotionaliteit) en zelfverachtende gedachten of zorgen maken (Cassady & Johnson, 2001). Testangst wordt gemeten met een vragenlijst bestaande uit zeven items die zijn geselecteerd uit de vertaalde Cognitive Test Anxiety Scale (CTAS) van Cassady en Johnson (2001). De alfa van deze vragenlijst is .72. Een voorbeeld van een vraag is: “Tijdens een test word ik zo nerveus dat ik feiten vergeet die ik normaal gesproken wel weet”.

Computerangst is de angst of vrees om met een computer te werken of de anticipatieangst om met een computer te werken (Beckers et al., 2006). In dit onderzoek wordt computerangst gemeten met de Computer Anxiety Rating Scale (CARS) van Heinssen, Glass en Knight (1987). Hierbij moet gedacht worden aan stellingen als: “Ik gebruik liever geen computer omdat ik bang ben dat ik fouten maak die niet meer te herstellen zijn”. De vragenlijst bestaat uit vijf items en heeft een alfa van .73.

Testmotivatie wordt gemeten met een vertaalde en verkorte versie van de Test Attitude Survey (TAS; Arvey, Strickland, Drauden & Martin, 1990). De gebruikte vragenlijst bestaat uit vijf items en heeft een alfa van .74. Items zijn stellingen die de mate van motivatie voor de test moeten weergeven, zoals: “Ik zal mijn uiterste best doen op deze test”.

Self-efficacy is het geloof in en het bewust zijn van het eigen kunnen (Bandura, 1977). Dit construct wordt in dit onderzoek gemeten met een vertaalde en verkorte schaal van Pintrich en De Groot (1990). Deze vragenlijst bestaat uit drie stellingen ($\alpha = .73$). Een voorbeelditem is: “Ik ben ervan overtuigd dat ik de vragen in deze test correct kan beantwoorden”.

Indrukvaliditeit, wat weergeeft in hoeverre een kandidaat de test relevant vindt voor het bepalen van het niveau van werkrelevante kennis, vaardigheden en kwaliteiten (Cook, 2004), wordt gemeten door middel van een vertaalde en verkorte schaal van Smither et al. (1993). De schaal bestaat uit zes items en heeft een alfa van .73. Vragen als “deze test is voor een leidinggevende functie relevant” en “deze test voorspelt prestaties in een leidinggevende functie goed” komen hierbij aan bod.

Het leuk vinden van de test wordt gemeten met 3 stellingen ($\alpha = .88$) zoals: “Deze test spreekt mij erg aan”. Ook deze vragen zijn geselecteerd uit een vertaalde vragenlijst van Weichmann en Ryan (2003).

Criteria

De criteria zijn de gemiddelde score op professioneel gedrag en het gemiddelde cijfer van de bloктоetsen. Een studiejaar bestaat uit acht blokken die elk wordt afgesloten met een bloктоets. Het criterium gemiddeld bloктоetscijfer is berekend uit de cijfers van tenminste vijf bloктоetsen. De eerstejaars studenten hadden op het meetmoment van de criteria slechts vijf blokken erop zitten, voor de tweedejaars is het gemiddelde van 13 blokken genomen en voor de vierdejaars is het gemiddelde van 22 bloктоetsen genomen.

De score op professioneel gedrag wordt ook acht keer per jaar gegeven. Ook hier geldt dat de eerstejaars slechts vijf cijfers voor professioneel gedrag hebben en bij de tweede en vierdejaars kunnen 16 cijfers worden meegenomen. Deze 16 cijfers zijn de cijfers uit het eerste en het tweede jaar. De cijfers voor professioneel gedrag worden door verschillende tutores gegeven aan de hand van een gestandaardiseerd scoreformulier. Ieder blok heeft de tutor een andere werkgroepen en ieder blok beoordeelt de tutor alle studenten van zijn of haar werkgroep op professioneel gedrag. Dit gebeurt aan de hand van een gestandaardiseerd scoreformulier met 19 vragen. Dit scoreformulier bevat onderwerpen als 'vaardigheden als gespreksleider/voorzitter', 'vaardigheden als notulist' en 'vaardigheden als groepslid'. De beoordeling gebeurt op een vijfpuntsschaal en uit deze gegevens wordt automatisch een cijfer berekend door de computer. Wanneer een principale componenten analyse met varimax rotatie wordt uitgevoerd, laden de vragen op drie factoren (zie Bijlage 1). De verklaarde variantie is 75,81%. Slechts één vraag, die betrekking heeft op het al dan niet op tijd komen van de student, is buiten beschouwing gelaten. De factoren zijn literatuuurbegrip ($EW = 11,41$), voorzitterschap, ($EW = 1,83$) en betrokkenheid bij de groep ($EW = 1,16$). Literatuuurbegrip bestaat uit stellingen als "De bijdragen van de student aan de nabespreking waren inhoudelijk van hoog niveau". Deze factor heeft vijf items en heeft een alfa van .94. De factor voorzitterschap ($\alpha = .92$) bestaat ook uit vijf stellingen. Een voorbeeld van een dergelijke stelling is: "De student stimuleerde als gespreksleider alle groepsleden om bij te dragen". Tot slot bestaat de factor betrokkenheid uit 8 stellingen ($\alpha = .93$). een voorbeeld van een stelling die bij deze factor hoort is: "De student had een geïnteresseerde, betrokken houding ten opzichte van de groep". Deze afzonderlijke factoren worden onderscheiden in de verdere analyses. In Bijlage 1 staan alle schalen ingedeeld onder de drie factoren.

Resultaten

Hieronder worden de resultaten van het onderzoek weergegeven. Allereerst worden enkele beschrijvende resultaten gegeven, vervolgens komen de verschillende hypothesen met bijbehorende toetsing en resultaten aan bod.

Uit het correlatieonderzoek (zie Tabel 4) is een aantal verschillen tussen groepen naar voren gekomen, die vervolgens nader is onderzocht met een aantal onafhankelijke t-toetsen (zie Tabel 3). Verschillen tussen mannen en vrouwen komen met name naar voren bij de persoonlijkheidsvragenlijst en bij de

score op de SQ-test. Vrouwen scoren beter op de SQ-test ($t = -2.12, p < .05$), maar zijn minder extravert ($t = 2.34, p < .01$), minder emotioneel stabiel ($t = 2.97, p < .01$), minder openzinnig ($t = 2.36, p < .05$) en presenteren zichzelf minder graag ($t = 3.56, p < .01$).

Tabel 3. Verschillen tussen mannen en vrouwen op persoonlijkheid.

	Geslacht	N	M	SD	T
Extraversie	man	37	98.46	16.94	2.34*
	vrouw	69	91.52	13.09	
Vriendelijkheid	man	37	103.41	8.76	0.08
	vrouw	69	103.26	8.72	
Consciëntieusheid	man	37	143.89	17.94	0.13
	vrouw	69	143.48	14.22	
Stabiliteit	man	37	78.08	9.97	2.97**
	vrouw	69	71.80	10.60	
Openzinnigheid	man	37	180.38	14.84	2.36*
	vrouw	69	173.68	13.38	
Zelfpresentatie	man	37	133.22	16.80	3.56**
	vrouw	69	122.68	13.15	
Score op de SQ-test	man	37	48.68	10.09	-2.12*
	vrouw	69	52.67	8.81	

* $p < .05$, ** $p < .01$ (tweezijdig getoetst)

In Tabel 4 staat nog een aantal significante correlaties tussen persoonlijkheidsdimensies en andere factoren. Zo correleert leeftijd significant met consciëntieusheid ($r = .20, p < .05$) en met stabiliteit ($r = .22, p < .05$). Het jaar waarin de student zit is significant gecorreleerd met extraversie ($r = .22, p < .01$) en consciëntieusheid ($r = .34, p < .01$). Intelligentie heeft een negatieve correlatie met consciëntieusheid ($r = -.24, p < .05$).

Verder is er een verschil tussen de kandidaten die eerst de SQ-test hebben gedaan en de kandidaten die eerst de webcamtest hebben gedaan. Negenenveertig kandidaten maakten eerst de webcamtest en 56 kandidaten maakten eerst de SQ-test. De kandidaten die eerst de SQ-test hebben gemaakt scoren significant hoger op de WCT ($t = 2.35, p < .05$). Andersom was dit niet het geval. De kandidaten die eerst de webcamtest hadden gemaakt scoorden niet significant beter op de SQ-test ($t = -0.27, p > .05$).

In Tabel 4 staat dat zowel werkervaring als leidinggevende ervaring beide geen invloed hebben op de prestatie op de webcamtest (respectievelijk $r = .15, p > .05$; $r = .16, p > .05$). Werkervaring en leidinggevende ervaring zijn wel aan elkaar gecorreleerd ($r = .24, p < .05$). Werkervaring en leidinggevende ervaring zijn beide gecorreleerd aan studiejaar (respectievelijk $r = .36, p < .01$ en $r = .27, p < .01$). Werkervaring is ook gecorreleerd aan leeftijd ($r = .37, p < .01$) en aan extraversie ($r = .20, p < .05$) en leidinggevende ervaring is gecorreleerd aan de factor betrokkenheid van het criterium professioneel gedrag ($r = .22, p < .05$).

Leeftijd (zie Tabel 4) is wel een significante voorspeller van prestatie op de webcamtest ($r = .32, p < .01$). Studiejaar is nog een betere voorspeller van de prestatie op de webcamtest ($r = .48, p < .01$). Tot slot correleert leeftijd nog significant met studiejaar ($r = .67, p < .01$).

Betrouwbaarheid en beoordelaarsovereenstemming van de webcamtest

De Conbrach's Alfa van de webcamtest heeft een waarde van .83. Hiermee is de eerste hypothese, die stelde dat de webcamtest een hoge betrouwbaarheid heeft ($\alpha \geq .79$), bevestigd. Voor de tweede hypothese die voorspelde dat de beoordelaars een hoge overeenstemming hebben in hun beoordeling, is een intra-klasse correlatiecoëfficiënt berekend. Iedere kandidaat is beoordeeld door drie beoordelaars en hun overeenstemming is voldoende hoog om de tweede hypothese te bevestigen ($r = .82$).

De voorspellende waarde van de webcamtest

Voor de derde hypothese die stelde dat de webcamtest een hogere voorspellende waarde heeft voor het criterium professioneel gedrag dan voor het criterium bloктоetscijfer, zijn correlaties berekend. De correlatietabel (Tabel 4) laat zien dat de webcamtest hoog gecorreleerd is met het criterium professioneel gedrag ($r = .36, p < .01$) en niet significant gecorreleerd is ($r = .11, p > .05$) met het criterium bloктоetscijfer. Uit een Fisher z-transformatie bleek dat beide correlaties significant van elkaar verschillen ($z = -2.80, p < .01$). Met dit resultaat is de derde hypothese bevestigd.

De bloктоets heeft wel een aantal voorspellers, namelijk de intelligentietest ($r = .23, p < .05$), vriendelijkheid ($r = -.26, p < .01$), openzinnigheid ($r = -.20, p < .05$) en de SQ-test ($r = .29, p < .01$). Daarnaast hebben de SQ-test en consciëntieusheid nog een voorspellende waarde voor het criterium professioneel gedrag (respectievelijk $r = .27, p < .01$ en $r = .24, p < .05$).

Wanneer het criterium professioneel gedrag opgedeeld wordt in de drie componenten literatuu**rbegrip**, voorzitterschap en betrokkenheid, blijkt dat de component literatuu**rbegrip** significant correleert met intelligentie ($r = .28, p < .01$), met SQ ($r = .28, p < .01$), met de webcamtest ($r = .40, p < .01$), met extraversie ($r = .20, p < .01$) en met studiejaar ($r = .21, p < .05$). De webcamtest is de enige vaardigheidstest die significant correleert met de component voorzitterschap ($r = .22, p < .05$). Alleen consciëntieusheid van de persoonlijkheidsvragenlijst correleert ook significant met de component voorzitterschap ($r = .27, p < .05$). De SQ-test, consciëntieusheid en studiejaar hebben een positieve voorspellende waarde voor de component betrokkenheid bij de groep (respectievelijk $r = .25, p < .05$; $r = .25, p < .05$; $r = .24, p < .05$).

Tabel 4. Correlaties

	M	SD	G	L	SJ	WE	LE	IQ	SQ	WCT	Extr	Vrie	Cons	Stbl	Open	Zelfp	Blok	Prof	LB	VZ	BT	
Geslacht	1.65	0.48	-																			
Leeftijd	21.94	2.85	.00	-																		
Studiejaar	2.21	1.34	.04	.67**	-																	
Werkervaring	2,65	0,99	-.04	.37**	.36**	-																
Leidinggevende ervaring	1,37	1,16	-.01	.18	.27**	.24*	-															
IQ	43.65	10.71	-.15	-.05	.13	-.17	-.14	(.76)														
SQ	51.28	9.43	.20*	.08	.16	.10	.01	.13	(.75)													
WCT	46.34	10.07	.13	.32**	.48**	.15	.16	.04	.21*	(.83)												
Extraversie	3.48	0.55	-.22*	.08	.22**	.20*	.19	.02	-.07	.24*	(.92)											
Vriendelijkheid	3.69	0.31	-.01	-.03	-.06	.13	-.03	-.03	-.13	-.06	.24*	(.83)										
Consciëntieusheid	3.59	0.39	-.01	.20*	.34**	.16	.16	-.24*	.03	.17	.09	.20*	(.92)									
Stabiliteit	3.22	0.47	-.28**	.22*	.18	.17	.03	.14	-.01	.20*	.30**	.16	.07	(.88)								
Openzinnigheid	3.75	0.30	-.23*	.15	.06	.12	.07	.03	.01	.12	.52**	.43**	.16	.35**	(.90)							
Zelfpresentatie	3.61	0.44	-.33**	.11	.17	.07	.16	-.09	-.10	.24*	.67**	.14	.38**	.41**	.59**	(.93)						
Bloktoets	6.34	0.81	.09	.07	.09	-.18	-.04	.23*	.29**	.11	-.17	-.26**	.07	-.18	-.20*	-.15	- ^a					
Professioneel gedrag	7.53	0.74	.04	.13	.13	.07	.14	.14	.27**	.36**	.12	.01	.24*	-.01	.04	.05	.56**	(.88)				
Literatuurbegrip	3.62	0.47	-.07	.16	.21*	-.01	.15	.28**	.28**	.40**	.20*	-.03	.16	.11	.13	.13	.57**	.90**	(.94)			
Voorzitterschap	3.84	0.41	.03	.05	.04	.05	.17	.09	.12	.22*	.13	.05	.27**	.01	.02	.09	.42**	.82**	.71**	(.92)		
Betrokkenheid	3.77	0.45	.13	.22	.24*	.14	.22*	.07	.25*	.10	.12	.00	.25*	-.06	-.02	.00	.52**	.93**	.82**	.77**	(.93)	

* $p < .05$, ** $p < .01$ (tweezijdig getoetst). M = gemiddelde; SD = standaarddeviatie; G = geslacht (1 = man, 2 = vrouw); L = leeftijd (min = 19, max = 35); SJ = studiejaar (jaar 1 t/m 4); WE = werkervaring (schaal 1-5); LE = leidinggevende ervaring (schaal 1-5); IQ = intelligentietest (min = 1, max = 94); SQ = SQ-test (situatiele beoordelingstest, min = 1, max = 100); WCT = webcamtest (min = 1, max = 100); Extr = extraversie (vijfpuntsschaal); Vrie = vriendelijkheid (vijfpuntsschaal); Cons = consciëntieusheid (vijfpuntsschaal); Stbl = emotionele stabiliteit (vijfpuntsschaal); Open = openzinnigheid (vijfpuntsschaal); Zelfp = zelfpresentatie (vijfpuntsschaal); Blok = bloktoets (min = 1, max = 10); Prof = professioneel gedrag (min = 1, max = 10); LB = literatuurbegrip (vijfpuntsschaal); VZ = voorzitterschap (vijfpuntsschaal); BT = betrokkenheid (vijfpuntsschaal). Op de diagonaal zijn de betrouwbaarheden weergegeven; ^a: betrouwbaarheid van de bloktoets is irrelevant omdat ieder blok andere onderwerpen getoetst worden.

Incrementele validiteit van de webcamtest

De vierde hypothese voorspelde dat de webcamtest een extra voorspellende waarde heeft voor het criterium professioneel gedrag ten opzichte van intelligentie, persoonlijkheid en de SQ-test. De webcamtest (zie Tabel 4) heeft een positieve correlatie met de SQ-test ($r = .21, p < .05$). Daarnaast correleert de webcamtest ook significant met een aantal componenten van de persoonlijkheidsvragenlijst, namelijk met extraversie ($r = .24, p < .05$), zelfpresentatie ($r = .24, p < .05$) en met emotionele stabiliteit ($r = .20, p < .05$).

Om te onderzoeken of de webcamtest een toegevoegde waarde heeft ten opzichte van de andere selectie-instrumenten, is een stapsgewijze lineaire regressie analyse uitgevoerd, waarbij de webcamtest als laatste component is toegevoegd. In de onderstaande tabel (Tabel 5) staan de resultaten van deze analyse. Demografische gegevens en de intelligentietest voorspellen tezamen niet meer dan 5% van het criterium professioneel gedrag, dit verschilt niet significant van toeval. Het toevoegen van de persoonlijkheidsvragenlijst, de SQ-test en de webcamtest zorgt wel voor een significante toename in de verklaarde variantie. Als de persoonlijkheidsvragenlijst wordt toegevoegd aan het model is de verklaarde variantie 17% en dit is een toename van 12% ($F = 2.11, p < .05$) bovenop de demografische gegevens en de IQ-test (β extraversie = .37; β vriendelijkheid = .02; β stabiliteit = -.11; β consciëntieusheid = .33; β openheid = -.05; β zelfpresentatie = -.29). Als de SQ-test ($\beta = .31$) wordt toegevoegd aan het model is de verklaarde variantie 26%, dit is een toename van 9% ($F = 10.36, p < .01$) bovenop de demografische gegevens, de IQ-test en de persoonlijkheidsvragenlijst. De verklaarde variantie van het volledige model, inclusief de webcamtest ($\beta = .36$) is $r^2 = .34$ ($\Delta R^2 = .09$), wat een sterk effect suggereert ($F = 11.26, p < .01$). De vierde hypothese kan daarmee bevestigd worden.

Tabel 5. Incrementele validiteit van de webcamtest op het criterium professioneel gedrag

	Predictor	β	R^2	ΔR^2	F
Stap 1	Studiejaar	-.02			
	Geslacht	-.13			
	Opleiding (VWO/Bachelor)	.27	.04	.04	.95
	Leeftijd	.13			
Stap 2	IQ	.13	.05	.01	1.19
Stap 3	Extraversie	.37			
	Vriendelijkheid	.02			
	Stabiliteit	-.11			
	Consciëntieusheid	.33	.17	.12	2.11*
	Openheid	-.05			
Stap 4	Zelfpresentatie	-.29			
	SQ	.31	.26	.09	10.36**
Stap 5	WCT	.36	.34	.09	11.26**

* $p < .05$, ** $p < .01$ (eenzijdig getoetst)

Het meest interessante facet van het criterium professioneel gedrag in relatie tot de voorspeller de webcamtest-leidinggeven is voorzitterschap. Daarom is met deze factor als afhankelijke variabele nog een stapsgewijze lineaire regressieanalyse gedaan.

Wanneer alleen de factor voorzitterschap (Tabel 6) wordt voorspeld door de verschillende predictoren, is een significante toevoeging van de persoonlijkheidsvragenlijst zichtbaar ($R^2 = .17$; $\Delta R^2 = .14$; $F = 2.45$, $p < .05$; β extraversie = .32; β vriendelijkheid = -.03; β stabiliteit = -.02; β consciëntieusheid = .40; β openheid = -.19; β zelfpresentatie = -.19), alsmede van de webcamtest ($\Delta R^2 = .08$; $F = 9.64$, $p < .01$; $\beta = .35$). De SQ-test voegt bij de voorspelling van voorzitterschap niets toe. Dit model verklaart 27% van de variantie, wat een sterk effect suggereert.

Tabel 6. Incrementele validiteit van de webcamtest op het criterium voorzitterschap van professioneel gedrag

	Predictor	β	R^2	ΔR^2	F
Stap 1	Studiejaar	-.09			
	Geslacht	-.09	.02	.02	.54
	Opleiding	.29			
	Leeftijd	.18			
Stap 2	IQ	.10	.02	.00	.13
Stap 3	Extraversie	.32			
	Vriendelijkheid	-.03			
	Stabiliteit	-.02	.17	.14	2.45*
	Consciëntieusheid	.40			
	Openheid	-.19			
Stap 4	Zelfpresentatie	-.19			
	SQ	.13	.19	.02	2.01
Stap 5	WCT	.35	.27	.08	9.64**

* $p < .05$, ** $p < .01$ (eenzijdig getoetst)

Kandidaat-percepties

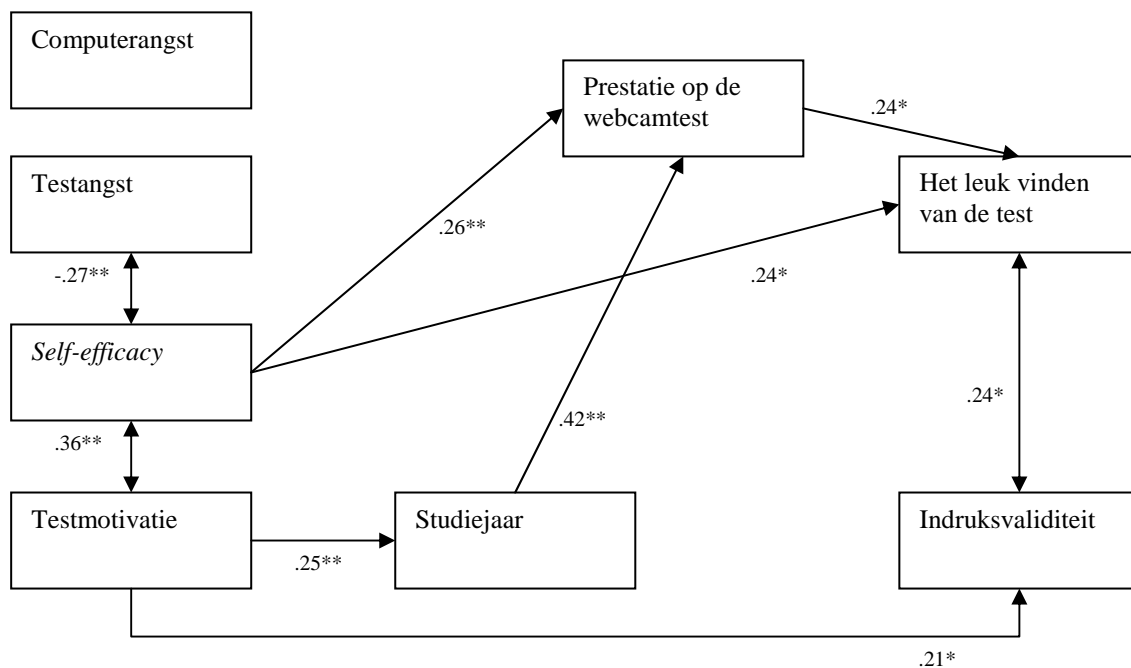
In Tabel 7 staan de correlaties tussen testangst, computerangst, de voorafgemeten motivatie en *self-efficacy* voor de webcamtest, de achteraf gemeten waardering van de webcamtest (het leuk vinden van de test en de indrukvaliditeit) en de daadwerkelijke score op de webcamtest. In Figuur 2 staan de relaties tussen de percepties die als significant uit de analyses kwamen.

Hypothese vijf voorspelde dat testangst en computerangst de prestatie op de webcamtest negatief zou beïnvloeden. Deze hypothese moet worden verworpen. Zoals zichtbaar is in Tabel 7, is er geen significante relatie tussen de score op de webcamtest en testangst ($r = -.08$, $p > .05$) en de score op de webcamtest en computerangst ($r = .03$, $p > .05$). Testangst heeft wel een significante negatieve correlatie met *self-efficacy* ($r = -.27$, $p < .01$).

Tabel 7. Correlaties tussen de percepties

	<i>M</i>	<i>SD</i>	TA	CA	M	SE	IV	L	WS
Testangst (TA)	2.45	0.70	(.72)						
Computerangst (CA)	1.36	0.44	.14	(.73)					
WCTmotivatie (M)	3.69	0.50	-.13	.01	(.74)				
WCT <i>self-efficacy</i> (SE)	3.03	0.54	-.27**	-.15	.36**	(.85)			
WCTindruksvaliditeit (IV)	3.90	0.58	.00	-.12	.21*	.04	(.73)		
Het leuk vinden van de WCT (L)	3.25	0.98	.01	.13	.17	.24*	.24*	(.76)	
Score op de webcamtest (WS)	48.14	10.07	-.08	.03	.26*	.26**	.08	.24*	(.83)

* $p < .05$, ** $p < .01$ (eenzijdig getoetst). De score op de webcamtest wordt weergegeven op een schaal van 0 -100; de overige variabelen worden weergegeven op een schaal van 1-5. Op de diagonaal zijn de betrouwbaarheden weergegeven.



* $p < .05$, ** $p < .01$ (eenzijdig getoetst).

Figuur 2. De significante relaties tussen de kandidaat-percepties en de prestatie op de webcamtest

Motivatie correleert significant met een aantal percepties. Motivatie is positief gecorreleerd met *self-efficacy* ($r = .36, p < .01$). Daarnaast zijn motivatie en indrukvaliditeit significant aan elkaar gecorreleerd ($r = .21, p < .05$). Ook is een relatie te zien tussen motivatie voor de webcamtest en de score op de webcamtest ($r = .26, p < .05$). Hiermee wordt in eerste instantie de zesde hypothese, waarin voorspeld werd dat de score op de webcamtest beïnvloed wordt door de motivatie van de kandidaat, bevestigd. Bij verdergaande analyse moeten we deze conclusie echter herzien.

Zoals eerder besproken heeft studiejaar een significante relatie met de score op de webcamtest ($r = .48, p < .05$). Er is daarnaast een significant verschil tussen de motivatie van de kandidaten van de verschillende jaren ($F(2,98) = 3.44, p < .05$). De vierdejaars zijn het meest gemotiveerd ($M = 3.87, SD = 0.60$), gevolgd door de tweedejaars studenten ($M = 3.61, SD = 0.31$) en tot slot de minst gemotiveerde groep de eerstejaars studenten ($M = 3.60, SD = 0.46$). Het verschil tussen de eerste en de tweedejaars is niet significant. Vanwege deze drie significante relaties werd een mediatio-analyse gedaan, die liet zien dat het effect van motivatie (zie Tabel 8) op de score op de webcamtest volledig wordt gemedieerd door studiejaar.

In de eerste twee stappen is motivatie een significante voorspeller. Motivatie voorspelt de score op de webcamtest ($\beta = .26, p < .05$) en studiejaar ($\beta = .25, p < .05$). Wanneer in de derde stap studiejaar als mediator wordt toegevoegd aan de regressieanalyse, is zichtbaar dat motivatie veel voorspellende waarde verliest ($\beta = .13, p > .05$) en dat studiejaar een hoge voorspellende waarde heeft voor prestatie op de webcamtest ($\beta = .42, p < .01$).

Tabel 8. Test mediatiemodel: Motivatie > studiejaar > score op de webcamtest

Test stappen	<i>B</i>	<i>SE B</i>	95% <i>CI</i>	β
Stap 1				
Uitkomst: WCT score				
Voorspeller: Motivatie	105.19	40.57	24.65, 185.73	.26*
Stap 2				
Uitkomst: Studiejaar				
Voorspeller: Motivatie	0.66	0.26	0.15, 1.17	.25*
Stap 3				
Uitkomst: WCT score				
Mediator: Studiejaar	64.57	14.62	35.54, 93.59	.42**
Voorspeller: Motivatie	53.17	38.90	-24.09, 130.42	.13

* $p < .05$, ** $p < .01$ (eenzijdig getoetst)

De zevende hypothese stelde dat *self-efficacy* niet gecorreleerd is met prestatie op de webcamtest. Deze hypothese moet verworpen worden. *Self-efficacy* is significant gerelateerd aan de score op de webcamtest ($r = .26, p < .01$).

De achtste hypothese moet gedeeltelijk verworpen worden. In deze hypothese werd gesteld dat de webcamtest door de kandidaat als een meer indrukvalide test beschouwd wordt in vergelijking met de intelligentietest en de persoonlijkheidsvragenlijst en als tenminste even indrukvalide beschouwd werd als de video-SJT voor het meten van leidinggevende kwaliteiten. Voor een vergelijking met de SQ-test, de persoonlijkheidsvragenlijst en de intelligentietest is een gepaarde t-toets uitgevoerd (zie Tabel 9).

Tabel 9. Indrukvaliditeit: vergelijking tests

	<i>M</i>	<i>SD</i>	<i>M-verschil met WCT</i>	<i>t</i>
WCT indrukvaliditeit	3.90	.58	-	-
IQ indrukvaliditeit	3.18	.76	0.73	7.36**
PV indrukvaliditeit	3.69	.57	0.22	3.09**
SQ indrukvaliditeit	4.02	.48	-0.11	1.90*

* $p < .05$, ** $p < .01$ (eenzijdig getoetst). Indrukvaliditeit wordt weergegeven op een schaal van 1-5. WCT = webcamtest; IQ = intelligentietest; SQ = SQ-test (video-SJT); M-verschil met WCT = het verschil tussen het gemiddelde van de desbetreffende test en het gemiddelde van de webcamtest.

Hieruit blijkt dat de kandidaten de webcamtest ($M = 3.90$, $SD = 0.58$) significant meer indrukvalide vinden voor het meten van leidinggevende kwaliteiten dan de intelligentietest ($M = 3.18$, $SD = 0.76$) en de persoonlijkheidsvragenlijst ($M = 3.69$, $SD = 0.57$; respectievelijk: $t = 7.36$, $p < .01$ en $t = 3.09$, $p < .01$), maar niet dan de SQ-test ($M = 4.02$, $SD = 0.48$; $t = 1.90$, $p < .05$).

In de laatste hypothese werd gesteld dat de webcamtest door de kandidaat als een leukere test wordt beschouwd dan de intelligentietest en de persoonlijkheidsvragenlijst en als een tenminste even leuke test wordt beschouwd als de video-SJT. Ook voor deze hypothese is een gepaarde t-toets uitgevoerd (zie Tabel 10). Hieruit blijkt dat de kandidaten de SQ-test ($M = 3.61$, $SD = 0.75$) significant leuker vonden om te doen dan de webcamtest ($M = 3.25$, $SD = 0.97$; $t = 3.01$, $p < .01$). De kandidaten vonden ook de persoonlijkheidsvragenlijst leuker om te maken ($M = 3.36$, $SD = 0.74$), maar vonden de intelligentietest minder leuk om te maken ($M = 3.10$, $SD = 0.98$). De verschillen tussen de webcamtest en de persoonlijkheidsvragenlijst en tussen de webcamtest en de intelligentietest waren echter niet significant (respectievelijk: $t = 1.08$, $p > .05$ en $t = 0.95$, $p > .05$). De negende hypothese moet dus verworpen worden. De score op de webcamtest heeft overigens wel een positieve relatie met het leuk vinden van de test ($r = .24$, $p < .05$). Ook *self-efficacy* en indrukvaliditeit zijn positief gecorreleerd met het leuk vinden van de test (respectievelijk $r = .24$, $p < .05$ en $r = .24$, $p < .05$).

Tabel 10. Het leuk vinden van de test: vergelijking tests

	<i>M</i>	<i>SD</i>	<i>M-verschil met WCT</i>	<i>t</i>
WCT leuk vinden	3.25	.97	-	-
IQ leuk vinden	3.10	.98	-0.14	0.95
PV leuk vinden	3.36	.74	0.12	1.08
SQ leuk vinden	3.61	.75	0.35	3.01**

* $p < .05$, ** $p < .01$ (eenzijdig getoetst). Het leuk vinden van de test wordt weergegeven op een schaal van 1-5. WCT = webcamtest; IQ = intelligentietest; SQ = SQ-test (video-SJT); M-verschil met WCT = het verschil tussen het gemiddelde van de desbetreffende test en het gemiddelde van de webcamtest.

Discussie

Allerlei veranderingen in de maatschappij zoals krapte op de arbeidsmarkt en technologische ontwikkelingen (e.g., Lievens et al., 2002), zorgen voor verschuivingen in de personeelsselectiemethoden. Het gevolg hiervan is dat er nieuwe selectiemethoden en –instrumenten worden ontwikkeld en onderzocht. Dit onderzoek had als doel de voorspellende waarde te bepalen van een nieuw selectie-instrument, namelijk de webcamtest. Daarnaast is nagegaan of deze test een toegevoegde waarde heeft ten opzichte van gangbare selectie-instrumenten, namelijk een intelligentietest en een persoonlijkheidsvragenlijst. Ook is onderzocht of de webcamtest een toegevoegde waarde heeft ten opzichte van een video-SJT, namelijk de SQ-test. Tot slot is gekeken naar (de invloed van) enkele percepties van kandidaten over de webcamtest. Zoals de eerste en de tweede hypothese voorspelden heeft de webcamtest een hoge betrouwbaarheid en hebben de beoordelaars een hoge overeenstemming. Beide resultaten zijn naar verwachting hoger dan de gemiddelde resultaten die Stricker (1982) vond in zijn onderzoek.

De derde hypothese voorspelde dat de webcamtest een betere voorspellende waarde zou hebben voor het criterium professioneel gedrag, dan voor het criterium bloктоetscijfer. Deze hypothese werd bevestigd: de webcamtest blijkt een goede voorspeller voor professioneel gedrag te zijn en heeft zelfs helemaal geen voorspellende waarde voor het bloктоetscijfer. Deze correlaties verschillen ook significant van elkaar. Dit sluit aan bij de onderzoeksresultaten van Lievens en Sackett (2006) die aantoonde dat video-assessment beter sociale gedragscriteria voorspellen dan cognitieve capaciteiten. De vierde hypothese voorspelde dat de webcamtest een incrementele validiteit zou hebben ten opzichte van een intelligentietest, een persoonlijkheidsvragenlijst en een video-SJT met betrekking tot het criterium professioneel gedrag. Bij een stapsgewijze regressieanalyse is de webcamtest zowel bij

het totale criterium professioneel gedrag als bij de factor voorzitterschap de variabele die zorgt voor een model met een aanzienlijke verklaarde variantie. In beide analyses werd de webcamtest als laatste voorspeller toegevoegd na enkele demografische gegevens, intelligentie, persoonlijkheid en de SQ-test. Hiermee werd de vierde hypothese bevestigd die voorspelde dat de webcamtest een incrementele validiteit heeft ten opzichte van intelligentie, persoonlijkheid en de SQ-test bij het criterium professioneel gedrag.

De webcamtest is, naast de persoonlijkheidsfactor consciëntieusheid, de enige test die een goede voorspeller blijkt voor het gedrag als voorzitter in een werkgroep bij het criterium professioneel gedrag. Dit is een belangrijke implicatie voor de constructvaliditeit van de webcamtest, namelijk de webcamtest beoogt leidinggevende vaardigheden te meten en correleert in dit onderzoek als enige vaardigheidstoets met het onderdeel van professioneel gedrag dat de voorzitterskwaliteiten van een student weergeeft. Een kanttekening hierbij is dat de webcamtest ook hoog correleert met de factor literatuurbegrip van het criterium professioneel gedrag. Het lijkt wellicht onlogisch dat de webcamtest, die in dit geval leidinggevende kwaliteiten beoogt te meten, ook significant gecorreleerd is met literatuurbegrip. Een verklaring hiervoor zou kunnen zijn dat een goede leidinggevende niet zonder inhoudelijke kennis kan. Een andere verklaring zou kunnen zijn dat de factor literatuurbegrip een ruimer begrip meet, namelijk het vermogen om grote hoeveelheden informatie te begrijpen en te integreren. Literatuurbegrip correleert namelijk ook hoog met de intelligentietest en de SQ-test. Met deze nieuwe interpretatie van de component literatuurbegrip, zou dit kunnen betekenen dat om een hoge score te behalen op een test, de kandidaat inzicht in en begrip van informatie nodig heeft.

De laatste factor van professioneel gedrag, betrokkenheid, correleert niet met de webcamtest, maar wel met studiejaar, consciëntieusheid en met de SQ-test. Wanneer we kijken naar de stellingen die vallen onder de factor betrokkenheid van professioneel gedrag (zie Bijlage 1), zien we dat er enkele stellingen in staan die betrekking hebben op actieve deelname en inzet tijdens de werkgroep. Dit zou de hoge correlatie tussen consciëntieusheid en deze factor van professioneel gedrag kunnen verklaren.

De significante correlatie tussen de SQ-test en de factor betrokkenheid van het criterium professioneel gedrag is lastiger te verklaren. De SQ-test die gebruikt is in dit onderzoek beoogt net als de webcamtest leidinggevende kwaliteiten te meten. Bij een opsplitsing van het criterium professioneel gedrag correleert deze SQ-test echter niet met de factor voorzitterschap, maar met de factoren literatuurinzicht en betrokkenheid. Een mogelijke verklaring voor het feit dat de SQ-test-leidinggeven eerder met deze schalen correleert dan met de schalen die voorzitterschap weergeven, zou kunnen zijn dat in de SQ-test-leidinggeven toch meer algemene sociale vaardigheden gemeten worden dan leiderschap. Echter is dit een vraag naar de constructvaliditeit van de SQ-test-leidinggeven en deze vraag ligt buiten het bereik van deze studie.

Verder heeft in dit onderzoek de SQ-test een voorspellende waarde voor het bloктоetscijfer. Dit is een opvallend resultaat omdat de SQ-test een test is die sociale vaardigheden - en in dit geval leidinggevende vaardigheden - beoogt te meten. Het verband tussen deze twee maten is wellicht

praktische intelligentie. Praktische intelligentie is het vermogen om op een effectieve manier om te gaan met diverse praktische problemen of situaties op basis van contextuele kennis (Sternberg, Forsythe, Hedlund, Horvath, Wagner & Williams, 2000, zoals beschreven in Chan en Schmitt, 2005). Om deze verklaring begrijpbaar te maken is het van belang om de onderwijsmethode van de studie psychologie aan de Erasmus Universiteit Rotterdam kort toe te lichten. De psychologiestudenten moeten aan de hand van probleem gestuurd onderwijs zelf literatuur selecteren, bestuderen, integreren en samenvatten. Dit probleem gestuurd onderwijs beoogt het probleemoplossend vermogen van de studenten te ontwikkelen (Dochy, Segers, Van den Bossche & Gijbels, 2003). De bloктоets toetst de verworven kennis over de voorgaande vijf weken bij de studenten. De bloктоets meet vanwege dit onderwijssysteem wellicht eerder de capaciteiten van de student om diverse informatiebronnen toe te passen om complexe vraagstukken op te lossen, dan kennis over de informatiebronnen op zich. Een gedeelte van de prestatie op de bloктоets zou dan verklaard kunnen worden door praktische intelligentie. Chan en Schmitt (2005) opperen het idee dat SJT's praktische intelligentie meten. Een kandidaat met een hoge praktische intelligentie zou bij het maken van een SJT onbewust zijn of haar contextuele kennis en de daarbij behorende procedurele regels toepassen om een goede reactie te geven of om de verschillende antwoordopties te beoordelen. De hoge correlatie tussen de SQ-test en het gemiddelde bloктоetscijfer zou verklaard kunnen worden als zowel de SQ-test als de bloктоets een vorm van praktische intelligentie meten.

Kandidaat-percepties

Weichmann en Ryan (2003) onderzochten de relaties tussen individuele verschillen en percepties bij kandidaten die of een schriftelijke test maakten of dezelfde test op een computer. Omdat wij een aantal van hun vragenlijsten hebben gebruikt zullen wij indien mogelijk de resultaten van dit onderzoek vergelijken met hun resultaten. In dit onderzoek werd voorafgaand aan de test gevraagd naar testangst, computerangst, testmotivatie en *self-efficacy*. Achteraf werd gevraagd hoe leuk de kandidaten de tests vonden en hoe indrukvalide de kandidaten de tests vonden.

De vijfde hypothese kon volledig verworpen worden: zowel testangst als computerangst hadden geen invloed op de prestatie van de kandidaten op de webcamtest. Allereerst werd geen effect gevonden van testangst op de prestatie op de webcamtest. Dit terwijl testangst wel aanwezig lijkt bij de kandidaten ($M = 2.45$, $SD = 0.70$). Dit is een positief teken en bevestigt de resultaten van O'Reilly et al. (1994) die vonden dat mensen liever video-assessment ondergaan dan een andere vorm van ondervraging. Testangst heeft verder wel een negatief verband met *self-efficacy*. Dat betekent dat mensen met testangst bij voorbaat minder zeker zijn over hun capaciteiten om goed te scoren op de webcamtest. Dit sluit aan bij eerder onderzoek. Zo vonden Wiechmann en Ryan (2003) een significante relatie tussen testangst en test *self-efficacy* ($r = -.19$, $p < .01$). Pintrich en De Groot (1990) vonden een nog sterkere significante relatie tussen deze twee factoren ($r = -.34$, $p < .01$).

Er werd ook geen relatie gevonden tussen computerangst en de score op de webcamtest. Als gekeken wordt naar de gemiddelde computerangst, kan geconcludeerd worden dat er vrijwel geen computerangst in deze groep is ($M = 1.36$, $SD = 0.44$). Computerangst correleert waarschijnlijk daarom met geen één andere factor. Berk en Nanda (2006) zaten met hun onderzoek in dezelfde situatie. Zij deden onderzoek naar in welke mate humor een positieve invloed had op testangst, echter vonden zij geen significante resultaten vanwege de lage mate van testangst bij deze steekproef.

Een verklaring voor het gebrek aan computerangst in deze onderzoeksgroep is dat studenten veel ervaring hebben met computers en daardoor minder computerangst hebben. Liu et al. (2001) concludeerden dat er een lage maar significante relatie was tussen computerangst en computerervaring ($r = -.16$, $p < .05$). Zij vonden dat de groep die online onderwijs volgde meer ervaring had met computers en minder last had van computerangst. Dit gaat in tegen het onderzoeksresultaat van Thorpe en Bosnan (2007) die aangeven dat computerangst een serieuze angst is die niet zal verdwijnen door ervaring. Belangrijk in relatie tot de generaliseerbaarheid van deze onderzoeksresultaten is dat computerangst wel degelijk in andere groepen kan voorkomen (e.g., Heinssen et al., 1987) en ook invloed kan hebben op verschillende percepties van de kandidaat. Wiechmann en Ryan (2003) vonden bijvoorbeeld wel computerangst in hun onderzoekspopulatie en vonden ook enkele negatieve effecten van computerangst. Computerangst was namelijk negatief gerelateerd aan het leuk vinden van de test ($r = -.45$, $p < .01$) en aan de indrukvaliditeit van de kandidaat over de test ($r = -.43$, $p < .01$). Het feit dat studenten in dit onderzoek geen computerangst hebben is dus gunstig voor onze onderzoeksresultaten, maar daarbij mag niet vergeten worden dat deze resultaten niet zonder meer gegeneraliseerd kunnen worden naar andere populaties.

Daarnaast is computerangst niet hetzelfde als webcam-angst. Het feit dat de studenten gefilmd worden kan voor een bepaalde prestatieangst zorgen die niets te maken heeft met computerangst. Cartney (2005) vond in haar onderzoek dat veel deelnemers prestatieangst ervoeren vanwege het feit dat ze opgenomen werden op video. In het huidige onderzoek is niet gevraagd naar deze eventuele prestatieangst. Wel zijn enkele significante correlaties gevonden tussen de score op de webcamtest en schalen van de persoonlijkheidsvragenlijst die in relatie tot deze verklaring interessant zijn. De score op de webcamtest is bijvoorbeeld positief gecorreleerd emotionele stabiliteit. Wanneer een kandidaat een lage emotionele stabiliteit heeft is hij of zij angstiger, zelfbewuster en gevoeliger (Judge et al., 1999). Deze eigenschappen kunnen het presteren voor de webcamtest beïnvloeden. Daarnaast is zelfpresentatie positief gecorreleerd aan de score op de webcamtest. De drang om goed te willen presteren en het willen tonen van de eigen vaardigheden kunnen bijdragen aan een goede prestatie op de webcamtest. Ook extraversie is positief gecorreleerd met de score op de webcamtest. Om goed te kunnen presteren voor een webcam zal een kandidaat gebaat zijn bij een open en sociabele houding.

De zesde hypothese stelde dat testmotivatie een positieve invloed zou hebben op de prestatie op de webcamtest. Toen bleek dat de vierdejaars studenten significant beter scoren op de webcamtest dan de eerste- en tweedejaars, was de verwachting dat dit gemedieerd zou worden door testmotivatie. Echter,

het omgekeerde was het geval. Bij het nagaan van de invloed van testmotivatie op de prestatie van de webcamtest werd ontdekt dat motivatie volledig gemedieerd wordt door studiejaar. Een verklaring hiervoor zou kunnen zijn dat alle vierdejaars studenten arbeids- en organisatiepsychologie studenten zijn. Wellicht zit er een verschil in interesse in, ervaring met en kennis van selectie-instrumenten in het algemeen wat de betere prestatie verklaard. Doch wordt door de vierdejaars niet significant beter gepresteerd op de intelligentietest of de SQ-test.

Een hoge *self-efficacy* zorgde in dit onderzoek voor een betere score op de webcamtest. Daarmee is de zevende hypothese, waarin gesteld werd dat *self-efficacy* niet gerelateerd zou zijn aan de score op de webcamtest, verworpen. Op basis van het onderzoek van Richard en collega's (2006) waarin gevonden werd dat *self-efficacy* een gevolg is van prestatie en niet andersom, was de verwachting dat met een onbekende test de *self-efficacy* van studenten geen effect zou hebben op de prestatie, omdat de studenten hun prestatie niet goed konden inschatten. De aanzienlijke correlatie tussen *self-efficacy* en de score op de webcamtest haalt deze redenatie onderuit. Een mogelijke verklaring voor het gevonden verband zou kunnen zijn dat de kandidaten in dit onderzoek dusdanig veel testervaring hebben dat het feit dat de webcamtest een onbekende test is geen rol speelt bij de *self-efficacy*. Ryan en collega's (1998) vonden immers een positieve relatie tussen test *self-efficacy* en testprestatie. Ook zou het kunnen zijn dat er sprake is van een indirect verband net zoals Brosnan (1998) vond. Test *self-efficacy* zou bijvoorbeeld invloed kunnen hebben op prestatieangst en zo een indirect effect kunnen hebben op de prestatie op de webcamtest.

Indruksvaliditeit wordt als steeds belangrijker beschouwd vanwege verschillende redenen, zoals het commerciële aspect dat de organisatie aantrekkelijk moet zijn voor de werkzoekende bij een krappe arbeidsmarkt. Wij hebben de kandidaten gevraagd hoe indruksvalide zij de verschillende tests vonden voor het meten van leidinggevende kwaliteiten. De verwachting zoals gesteld in de achtste hypothese was dat de kandidaten de webcamtest meer indruksvalide vonden voor het meten van leidinggevende kwaliteiten dan de persoonlijkheidsvragenlijst en de intelligentie test en tenminste even indruksvalide als de SQ-test voor leidinggeven. Het bleek dat de kandidaten de webcamtest inderdaad meer indruksvalide vonden dan de persoonlijkheidsvragenlijst en de intelligentietest, echter vonden zij de SQ-test meer indruksvalide dan de webcamtest. Een mogelijke verklaring hiervoor is dat de kandidaten meer bekend zijn met SJT's, meer van de validiteit van SJT's afweten en daarom de SQ-test als meer indruksvalide beoordelen. Een andere verklaring kan zijn dat kandidaten de SQ-test leuker vinden dan de webcamtest en dat dat gedeeltelijk de hogere indruksvaliditeit van de webcamtest verklaard. De resultaten van dit onderzoek laten zien dat het leuk vinden van de test een positief significant verband heeft met de indruksvaliditeit. Dus hoe leuker de kandidaat de test vindt hoe meer indruksvalide hij of zij de test vindt.

De laatste hypothese kan daarmee ook verworpen worden. De kandidaten vonden de SQ-test het leukste om te doen, daarna de persoonlijkheidsvragenlijst, daarna volgde pas de webcamtest en de kandidaten beschouwden de intelligentietest als de minst leuke test om te doen. Hypothese negen

stelde dat de kandidaten de webcamtest tenminste even leuk zouden vinden als de SQ-test en leuker dan de persoonlijkheidsvragenlijst en de intelligentietest. Wel moet hierbij vermeld worden dat de verschillen tussen de webcamtest en de persoonlijkheidsvragenlijst en de webcamtest en de intelligentietest niet significant waren. De SQ-test verschilde wel significant van de webcamtest. Een mogelijke verklaring hiervoor zou kunnen zijn dat de webcamtest toch de nodige prestatieangst opwekte. Dit is echter speculatief omdat wij dat niet hebben onderzocht. We hebben in dit onderzoek geen informatie over de redenen waarom kandidaten de tests wel of niet indrukvalide of leuk vonden. In dit onderzoek had *self-efficacy* een positieve invloed op hoe leuk kandidaten de test vinden. Dus hoe zelfverzekerder de kandidaten bij voorbaat waren over hun capaciteiten om de test goed te maken, hoe leuker zij de test achteraf vonden. Dit is in overeenstemming met de resultaten van Weichmann en Ryan (2003) die een verband vonden van $r = .20$ ($p < .01$) tussen *self-efficacy* en het leuk vinden van de test. In tegenstelling tot Wiechmann en Ryan vonden wij geen relatie tussen *self-efficacy* en indrukvaliditeit. Zij vonden deze relatie wel ($r = .19$, $p < .01$). Verder kan nog vermeld worden dat indrukvaliditeit gecorreleerd is met de testmotivatie. Wellicht vindt een kandidaat de webcamtest meer indrukvalide indien hij of zij vooraf meer gemotiveerd was om de test goed te maken. Dit zou aansluiten bij de cognitieve dissonantie theorie (Festinger, 1957, zoals omschreven in Kenrich, Neuberg & Cialdini, 2005) waarin gesteld wordt dat meer waardering gegeven wordt aan zaken waarvoor iemand veel inspanning levert. Verder heeft geen enkele perceptie invloed op de indrukvaliditeit. Dit is positief te noemen, de mening van de kandidaten over het niveau van de indrukvaliditeit van de webcamtest staat dus los van de eigen prestatie en daarbij wordt de test als relevanter beschouwd voor leidinggevende vaardigheden dan de persoonlijkheidsvragenlijst en de intelligentietest.

Limitaties van dit onderzoek en suggesties voor verder onderzoek

Dit onderzoek kent enkele limitaties die in overweging genomen moeten worden bij de waardering van de onderzoeksresultaten. Aansluitend hierop worden enkele suggesties gedaan voor verder onderzoek.

Allereerst is dit onderzoek een oriënterende studie met betrekking tot een nieuw selectie-instrument. Dat betekent dat het onderzoeksgebied in deze scriptie erg breed is en een initieel beeld geeft over de webcamtest, maar hier en daar verder uitgediept moet worden. Een eerste onderwerp dat van belang is om verder te onderzoeken is het feit dat de twee competenties, namelijk taakgericht en persoonsgericht leidinggeven, niet uit de factoranalyse zijn gekomen. De hoge correlatie tussen deze twee competenties is te verwachten, immers een goede reactie in bepaalde situatie vraagt om zowel een persoonsgerichte als een taakgerichte aanpak (Judge et al., 2004). Judge en collega's beschreven dat de correlatie tussen persoonsgericht leidinggeven en taakgericht leidinggeven erg verschilt per meetinstrument. Bij de ene vragenlijst is de gemiddelde geobserveerde correlatie tussen de twee competenties $r = .36$ en bij een andere vragenlijst $r = -.07$. Dit is lang niet zo hoog als de correlatie die

in dit onderzoek naar voren kwam ($r = .80$), maar het kan zijn dat het bij het beoordelen van geobserveerd gedrag lastiger is om de competenties te scheiden dan bij het beoordelen van een schriftelijke test of vragenlijst. Ook is mogelijk dat de competenties niet goed zijn gekozen of dat de scoringsinstructies niet voldoende onderscheid maken tussen de competenties. Echter zijn in de training van de beoordelaars de competenties en scoringsinstructies zeer uitgebreid aan bod gekomen en waar nodig aangepast en duidelijker gescheiden van elkaar.

Een andere mogelijke verklaring is dat de webcamtest leidinggevende kwaliteiten meet per situatie. Wellicht is het mogelijk om de theorie inzake taakgericht leidinggeven en persoonsgericht leidinggeven te combineren met een andere theorie. Zo stellen Judge en collega's voor om de theorie inzake *organizational justice* (Cropanzano, Byrne, Bobocel & Rupp, 2001) te betrekken in de theorie over persoonsgericht leidinggeven en taakgericht leidinggeven. Deze rechtvaardigheidstheorie beschrijft drie soorten rechtvaardigheid binnen een organisatie (1) verdelingsrechtvaardigheid "krijgt iedereen wat hij of zij verdient?", (2) procedurele rechtvaardigheid "verlopen processen binnen de organisatie op een rechtvaardige manier?" en (3) interactie rechtvaardigheid "wordt op interpersoonlijk gebied iedereen rechtvaardig behandeld (en dan met name door de beslissingnemers)?" Judge en collega's omschrijven dat een combinatie tussen taakgericht leidinggeven en verdelingsrechtvaardigheid betrekking heeft op situaties waarin een leidinggevende moet letten op een eerlijke verdeling naar inzet. Vervolgens benoemen zij dat persoonsgericht leidinggeven betrekking heeft op situaties waarin interactie rechtvaardigheid van belang is. Tot slot heeft procedurele rechtvaardigheid twee kanten. Een combinatie van persoonsgericht leidinggeven met procedurele rechtvaardigheid heeft betrekking op situaties waarin de leidinggevende de mening van de medewerker serieus meeneemt in beslissingen en een combinatie van taakgericht leidinggeven met procedurele rechtvaardigheid heeft betrekking op situaties waarin de leidinggevende duidelijke en heldere verwachtingen en regels stelt. Wellicht kunnen de situaties van de webcamtest-leidinggeven ingedeeld worden in deze categorieën en kunnen de competenties daarop aangepast worden.

Een andere limitatie van dit onderzoek is dat er enkel studenten onderzocht zijn die vrijwel allemaal verplicht zijn om mee te doen aan onderzoek op de universiteit. De inspanning, motivatie en prestatie kan hierdoor beïnvloed zijn. In hetzelfde kader is te stellen dat er voor de studenten niets afhangt van de tests wat op verschillende manieren invloed kan hebben op de resultaten. Zo kan het zijn dat studenten die normaal testangst hebben, dit nu niet hadden en daardoor beter scoorden op de tests dan wanneer zij de tests zouden maken in een selectiecontext. Bijvoorbeeld de conclusie dat er geen verband is tussen testangst en de score op de webcamtest wordt hiermee in twijfel getrokken.

Verder kwam in dit onderzoek naar voren dat het aantal jaren werkervaring of het aantal jaren leidinggevende ervaring geen invloed heeft op de prestatie op de webcamtest. Slecht 17% van de steekproef heeft geringe leidinggevende ervaring en 79% heeft werkervaring. Wel moet hierbij gesteld worden dat met werkervaring onder studenten hoogstwaarschijnlijk bijbaantjes op MBO-niveau bedoeld wordt (bijvoorbeeld bediening in de horeca of werken in een kledingwinkel). Ook de

informatie over het niveau van leidinggevende ervaring is onbekend. Omdat het niveau van het werk onbekend is en erg kan verschillen onder de studenten, zou het kunnen zijn dat er wel degelijk een effect is van werk- en leidinggevende ervaring, maar dat dat in dit onderzoek niet naar voren komt.

In relatie tot bovenstaande limitaties is het raadzaam om onderzoek te doen naar andere groepen in bijvoorbeeld een selectiecontext. Zo kan onderzoek gedaan worden naar prestatie op de webcamtest en invloed van testangst of computerangst bij een groep waarin prestatie op de webcamtest wel consequenties heeft voor de loopbaan. Ook kan het interessant zijn om werknemers uit verschillende branches te testen of mensen met en zonder leidinggevende ervaring.

Daarnaast zou in vervolgonderzoek aandacht besteed kunnen worden aan het prestatieverschil tussen allochtonen en autochtonen. Ondanks de theoretische onderbouwing dat de webcamtest *adverse impact* kan tegengaan, is daar in dit onderzoek geen aandacht aan besteed. Slechts 7% van de onderzoeksgroep is geboren in een ander land dan Nederland en dit zijn allemaal verschillende landen. Verder moeten we in dit onderzoek spreken van een gelijktijdigheidsvaliditeit in plaats van een predictieve validiteit. Het tegelijkertijd meten van voorspeller en criterium kan andere resultaten opleveren dan wanneer de voorspeller eerder gemeten wordt dan het criterium (McDaniel et al., 2001). Cook (2004) noemt de vaak veronderstelde nadelen van gelijktijdigheidsvaliditeit. Zo is causaliteit een verondersteld probleem, maar bijvoorbeeld ook *range of restriction*. Barrett, Phillips en Alexander (1981) weerleggen deze argumenten en geven aan dat de meerwaarde van predictieve validiteit over gelijktijdigheidsvaliditeit overschat wordt. Niet alleen in het kader van onderzoek naar de webcamtest, maar ook in het kader van verschillen tussen gelijktijdigheidsvaliditeit en predictieve validiteit is het interessant om dezelfde groep studenten over enkele jaren nogmaals te benaderen en een nieuw criterium te gebruiken om de voorspellende waarde van de webcamtest te bepalen. Hierbij kan gedacht worden aan een 360°-feedback-instrument.

Tot slot zijn de kandidaat-percepties interessant om verder uit te diepen. Zo is testmotivatie een onderwerp dat veel breder is dan dat hier besproken is. De vele factoren die invloed kunnen hebben op motivatie en de relatie tussen andere percepties en motivatie kunnen leiden tot interessante onderzoeksvragen die buiten het bereik van deze studie liggen. Een advies voor verdergaand onderzoek is daarom om met iedere hypothese de diepte in te gaan. Hierbij kan gedacht worden aan een vergelijking van meerdere selectie-instrumenten met betrekking tot bijvoorbeeld testmotivatie of testangst en computerangst. Daarnaast zou prestatieangst in verband met de hierboven genoemde redenen een onderwerp zijn dat interessant is om te onderzoeken. Bovendien gaat dit onderzoek niet in op de redenen waarom kandidaten de webcamtest minder leuk en minder indrukvalide vinden dan de SQ-test. Het zou met betrekking tot de verdere testontwikkeling interessant kunnen zijn om een kwalitatief onderzoek te doen naar de meningen van de kandidaten over de webcamtest.

Conclusie

De resultaten uit dit onderzoek laten zien dat de webcamtest een betrouwbaar instrument is dat een voorspellende en toegevoegde waarde heeft bij het meten van professioneel sociaal gedrag. De webcamtest heeft een hoge betrouwbaarheid, een hoge beoordelaarsovereenstemming, een voorspellende waarde voor het criterium professioneel gedrag en voor de factor voorzitterschap van het criterium professioneel gedrag. Daarnaast heeft de webcamtest een incrementele validiteit ten opzichte van een intelligentietest, een persoonlijkheidsvragenlijst en een video-SJT met betrekking tot het criterium professioneel gedrag en met betrekking tot de factor voorzitterschap van het criterium professioneel gedrag. De webcamtest is verder niet gecorreleerd aan intelligentie of de bloктоets wat suggereert dat de test echt een ander construct meet dan cognitieve capaciteiten en kennis. Met deze resultaten kan gesteld worden dat de webcamtest een interessant instrument is om te gebruiken bij een selectieprocedure.

Een mooi bijkomend resultaat is dat testangst en computerangst beide geen invloed hebben op de prestatie van de kandidaat op de webcamtest. Een kanttekening hierbij is dat de test niet is afgenomen in een selectiecontext en dat er wellicht sprake zou kunnen zijn van een andere angst zoals prestatieangst. Studiejaar en self-efficacy zijn beide, in tegenstelling tot testmotivatie, significant gerelateerd aan de prestatie van de kandidaat op de webcamtest. Verdere verbanden tussen de verschillende kandidaat-percepties zouden in vervolgonderzoek uitgediept kunnen worden. Tot slot is het resultaat dat de kandidaten de video-SJT zowel leuker als meer indrukvalide vonden interessant om verder te onderzoeken met betrekking tot testontwikkeling. Kwalitatief onderzoek kan wellicht antwoord geven op de vraag waarom de video-SJT een hogere waardering krijgt van de kandidaten dan de webcamtest.

Onderzoek naar de webcamtest staat in de kinderschoenen, doch zijn de resultaten tot nu toe veelbelovend. De vele aspecten die gedekt moeten worden bij onderzoek naar de webcamtest liggen niet allemaal binnen het bereik van deze studie, echter is hier een aanzet gegeven om de webcamtest verder te ontwikkelen en te onderzoeken.

Referenties

ANP (2007): *Middenbedrijf in de verdrinking*. Verkregen op 30 oktober 2007 van:

http://www.nu.nl/news/1264979/31/rss/%27Middenbedrijf_in_de_verdrinking%27.html

ANP (2006): *Banengroei zet door*. Verkregen op 30 oktober 2007 van:

http://www.nu.nl/news/836726/36/rss/Banengroei_zet_door.html

ANP (2007): *Een op de drie vacatures moeilijk vervulbaar*. Verkregen op 30 oktober 2007 van:

http://www.nu.nl/news/1166368/36/rss/Een_op_drie_vacatures_moeilijk_vervulbaar.html

ANP (2007): *Medisch personeel veelgevraagd*. Verkregen op 30 oktober 2007 van:

http://www.nu.nl/news/962127/36/rss/Medisch_uitzendpersoneel_veelgevraagd.html

ANP (2007): *Aantal werkzoekende fors gedaald*. Verkregen op 30 oktober 2007 van:

http://www.nu.nl/news/956183/36/rss/Aantal_werkzoekenden_fors_gedaald.html

Ards, J. (2006). De nieuwe context. In G. Smit, H. Verhoeven & A. Driessen (Eds.)

Personeelsselectie en assessment: Wetenschap in de praktijk (p. 2-10). Assen, Nederland: Koninklijke Van Gorcum BV.

Arvey, R.D., Strickland, W., Drauden, G. & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695-715.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 84(2), 191-215.

Bakker, P. (2006). Selectiebureaus worden ontwikkelcentra. *Volkscrant Banen* van dinsdag 28 november 2006.

Barrett, G.V., Phillips, J.S. & Alexander, R.A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology* 66(1), 1-6.

Beckers, J.J., Wicherts, J.M. & Schmidt, H.G. (2006). Computer anxiety: "trait" or "state"? *Computers in Human Behaviour*, 23, 2851-2862.

Berk, E.A. & Nanda, J. (2006). A randomised trial of humor effects on test anxiety and test performance. *Humor*, 19(4), 425-454.

- Borman, W.C., Klimoski, R.J. & Ilgen D.R. (2003) Stability and Change in Industrial and Organizational Psychology. In W.C. Borman, R.J. Klimoski & D.R. Ilgen (Eds.) *Handbook of psychology Vol 12 Industrial and Organizational Psychology* (p. 1-20). Hoboken, New Jersey: John Wiley & Sons Inc.
- Brosnan, M.J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of Computer Assisted Learning* 14, 223-234.
- Cartney, P. (2006). Using video interviewing in the assessment of social work communication skills. *Britisch Journal of Social Work*, 36, 827-844.
- Cassady, J.C. & Johnson, R.E. (2001). Cognitive test anxiety and academic performance. *Elsevier Science (USA)* 27, 270-295.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology* 82(1), 143-159.
- Chan, D. & Schmitt, N. (2005) Situational judgment tests. In A. Evers, N. Anderson & O. Voskuijl (Eds.) *Handbook of personnel selection* (p. 219–246). Oxford: Blackwell.
- Conway, J.M., Jako, R.A. & Goodman, D.F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology* 80(5), 565-579.
- Cook, M. (2004). *Personnel selection: Adding value through people* (4th ed.). Chichester: John Wiley & Sons Ltd.
- Cropanzano, R., Byrne, Z.S., Bobocel, D.R. & Rupp, D.E. (2001). Moral virtues, fairness heuristics, social entities, and other denizens of organizational justice. *Journal of Vocational Behavior* 58, 164–209
- Deffenbacher, J.L. & Hazaleus, S.L. (1985). Cognitive, emotional, and physiological components of test anxiety. *Cognitive Therapy and Research*, 9(2), 169-180.
- Dipboye, R.L., Wooten, K. & Halverson, S.F. (2004) Behavioral and situational interviews. In M. Hersen (Ed.) *Comprehensive Handbook of Psychological Assessment Vol 4*. Chapter 17 (p. 297-318). Hoboken, New Jersey: John Wiley & Sons Inc.

- Dochy, F., Segers, M., Van den Bossche, P. & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*, 533-568.
- Drenth, P.J.D. & Sijtsma, K. (2006). *Testtheorie; Inleiding in de theorie van de psychologische test en zijn toepassingen* (4th ed.). Houten: Bohn Stafleu van Loghum.
- Funke, U. & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment, 6*(2), 115-123.
- GITP INTERNATIONAL BV (2006) *Test informatie Captain analogieën*. Versie 1.0.doc.
- GITP INTERNATIONAL BV (2006) *Test informatie Captain cijferreeksen en rekenvraagstukken*. Versie 1.0.doc.
- GITP INTERNATIONAL BV (2006) *Test informatie G5R*. Versie 1.0.doc.
- GITP INTERNATIONAL BV (2006) *Test informatie figuurklassen*. Versie 1.0.doc.
- Heinssen, R.K., Glass, C.R. & Knight (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behaviour, 3*, 49-59.
- Hembree (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research, 58*(1), 47-77.
- Judge, T.A., Higgins, C.A, Thoresen, C.J. & Barrick, M.R. (1999). The big five personality traits, general mental ability, and careersuccess across the lifespan. *Personnel Psychology, 52*, 621-652.
- Judge, T.A, Piccolo, R.F. & Ilies, R. (2004) The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89*(1), 36-51.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, Mass: MIT Press.
- Lievens, F. (2001). Assessor training strategies and their effects on accuracy, Interrater reliability, and discriminant validity. *Journal of Applied Psychology, 86*(2), 255-264.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology, 87*(4), 675-686.
- Lievens, F. & Conway, J.M. (2001). Dimension and Exercise Variance in Assessment Center Scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*(6), 1202-1222.

- Lievens, F., Dam, K. van, & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review*, 31(5), 580-601.
- Lievens, F. & Sackett, P.F. (2006). Video-Based versus written situational judgement tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91(5), 1181-1188.
- Liu, M., Papathanasiou, E. & Hao, Y (2001). Exploring the use of multimedia examination formats in undergraduate teaching: results from the fielding testing. *Computers in Human Behavior*, 17, 225-248.
- McDaniel, M.A., Hartman, N.S., Whetzel, D.L. & Lee Grubb III, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 6. 63-91.
- McDaniel, M.A., Morgeson, F.P., Bruhn Finnegan, E. & Campion, M.A. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology* 86(4), 730-740.
- McKenna, E.F. (2000). *Business psychology & organizational behaviour: A students handbook* (3rd ed.). Hove, East Sussex: Psychological Press Ltd.
- NIP (2008). *Documentatie van tests en testresearch in Nederland, aanvullingen 2008/01*. Amsterdam, Nederland: Boom test uitgevers.
- NIP (2008). *COTAN*. Verkregen op 17 april 2008 van: http://www.psynip.nl/intern_4level.asp?deurl=/nipalg/subpage.asp&topmenuid=2&submenuid=11
- NOVEM (9 oktober 2007): *Stijging online vacatures in september*. Verkregen op 30 oktober 2007 van: http://www.nu.nl/news/1266275/50/rss/%27Stijging_online_vacatures_in_september%27.html
- Van der Maesen Advies (2005). *Webcamtest voor actieve sociale vaardigheden*. Verkregen op 30 oktober 2007 van: <http://www.vandermaesen.nl/upload/files/Webcamtest%20algemeen.pdf>
- Van der Maesen | Koch HRM-advies (2007). *Achtergrond & onderzoek*. Verkregen op 3 maart 2008 van: <http://www.vandermaesenkoch.nl/webcamtest/index.php?id=2>
- O'Reilly, J.M., Hubbard, M.L., Lessler, J.T., Biemer, P.P. & Turner, C.F. (1994). Audio and video computer assisted self-interviewing: Preliminary tests of new technologies for data collection. *Journal of Official Statistics* 1(2), 197-214.

- Perkins, R.F. (1995). Using hypermedia programs to administer tests: effects on anxiety and performance. *Journal of Research on Computing in Education*, 28(2), 209-220.
- Pintrich, P.R. & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* 82(1), 33-40.
- Ployhart, R.E. & Ehrhart, M.G. (2003). Be carefull what you ask for: Effects of response instructions on the construct validity and reliability of situational judgement tests. *International Journal of Selection and Assessment* 11(1), 1-16.
- Richard, E.M., Diefendorff, J.M. & Martin, J.H. (2006). Revisiting the within-person self-efficacy and performance relation. *Human Performance*, 19(1), 67-87.
- Richman-Hirsch, W.L., Olson-Buchanan, J.B. & Drasgow, F. (2000). Examining the impact of administartion medium on examinee perceptions and attitudes. *Journal of Applied Psychology* 85(6), 880-887.
- Ryan, A.M. & Ployhart, R.E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565-606.
- Ryan, A.M., Ployhart, R.E., Greguras, G.J. & Schmit, M.J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology*, 51, 599-621.
- Salgado, J.F. & Lado, M. (2000). Validity generalization of video tests for predicting job performance ratings. Paper presented at the 15th Annual Conference of the *Society of Industrial and Organisational Psychology*, April, New Orleans.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Smiderle, D., Perry, B.A. & Cronshaw, S.F. (1994). Evaluation of video-based assessment in transit operator selection. *Journal Business and Psychology* 9(1), 3-22.
- Smither, J.W., Reilly, R.R., Millsap, R.E. & Stoffey, R.W. (1993). Applicants reactions to selection procedures. *Personnel Psychology*, 46, 49-76.
- Stöber, J. & Pekrun, R. (2004). Editorial: advances in test anxiety research. *Anxiety, Stress and Coping* 17(3), 205-211.

- Stricker, L.J. (1982). Interpersonal competence instrument : Development and preliminary findings. *Applied Psychology Measurement* 6(1), 69-81.
- Sundre, D.L. & Wise, S.L. (2003). *Motivation filtering: an exploration of the impact of low examinee motivation on the psychometric quality of tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Thorpe, S.J. & Brosnan, M.J. (2004). Does computer anxiety reach levels which conform to DSM IV criteria for specific fobia? *Computers in Human Behaviour*, 23, 1258-1272.
- Van Tilborg, J. (2006). Een passende selectieprocedure. In G. Smit, H. Verhoeven & A. Driessen (Eds.) *Personeelselectie en assessment: Wetenschap in de praktijk* (p. 2-10). Assen, Nederland: Koninklijke Van Gorcum BV.
- Vroom, V.H. & Jago, A.G. (2007). The role of the situation in leadership. *American Psychologist*, 62(1), 17-24.
- Weekley, J.A. & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25-49.
- Wiechmann, D. & Ryan, A.M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, 11(2/3), 215-229.
- Wise, S.L. & DeMars, C.E. (2003). *Examinee motivation in low-stakes assessment: Problems and potential solutions*. Paper presented at the annual meeting of the American Association of Higher Education Assessment Conference, Seattle, June, 2003.
- Woehr, D.J. (1994). Understanding frame-of-reference training: The impact of training on the recall of performance information. *Journal of Applied Psychology*, 79(4), 525-534.
- Yoo, T. & Muchinsky, P.M. (1998). Utility estimates of job performance as related to the data, people, and things parameters of work. *Journal of Organizational Behavior*, 19(4), 353-370.

Bijlage 1. Factoren van professioneel gedrag

Naam factor	Vragen behorende bij de factor
	1) Als gespreksleider had de student zich goed voorbereid en had hij/zij een duidelijk overzicht van wat er besproken moest worden.
	2) Als gespreksleider zorgde de student voor duidelijke structuur in de bespreking van het probleem.
Voorzitterschap	3) De timing en de kwaliteit van de samenvattingen van de student als gespreksleider was goed.
	4) De student stimuleerde als gespreksleider alle groepsleden om bij te dragen.
	5) De student stelde als gespreksleider concrete vragen aan de groep om diepgang in de discussie te stimuleren.
	6) De student slaagde erin als notulist zelfstandig bijdragen van groepsleden kort en helder samen te vatten op het bord.
	7) De student gaf blijk van goede zelfstudie.
	10) De student had extra, relevante literatuur bestudeerd bovenop de minimale eis van twee bronnen per probleem.
	13) De student nam actief deel aan de brainstorm en probleemanalyse.
Betrokkenheid	14) De student nam actief deel aan de nabespreking van de problemen.
	15) De student was tijdens de voor- en nabespreking gemotiveerd om de problemen tot in de details uit te werken.
	18) De student luisterde actief naar de bijdragen van andere groepsleden.
	19) De student had een geïnteresseerde, betrokken houding ten opzichte van de groep.
	8) De bijdragen van de student aan de nabespreking waren inhoudelijk van hoog niveau.
	9) De student was in staat onderscheid te maken tussen hoofd- en bijzaken in de literatuur.
Literatuurbegrip	11) De student was in staat de bestudeerde stof in eigen woorden uit te leggen.
	16) De student creëerde diepgang in de discussie door bijvoorbeeld kritische vragen te stellen.
	17) De student zocht uit zichzelf naar verbanden tussen theorieën en/of bijdragen van groepsleden.
