



Universiteit Twente
de ondernemende universiteit

**De invloed van afnamevorm en afnameconditie op de uitslag van een
intelligentietest**

H.C. van de Vis (s0047783)

Begeleiders:

dr. R.R. Meijer

drs. I.J.L. Egberink

Enschede, december 2007

Universiteit Twente

Faculteit Gedragwetenschappen

**DE INVLOED VAN AFNAMEVORM EN AFNAMECONDITIE OP DE UITSLAG
VAN EEN INTELLIGENTIETEST**

De invloed van afnamevorm en afnameconditie op de uitslag van een intelligentietest.

H.C. van de Vis

Universiteit Twente

Samenvatting

In dit onderzoek is onderzocht in hoeverre de afnamevorm en de afnameconditie van invloed zijn op de uitslag van een intelligentietest. Om dit te onderzoeken is een intelligentietest die in ontwikkeling was ingezet. Deze is zowel gecontroleerd als ongecontroleerd en zowel online als op papier afgenomen. Gecontroleerd wil zeggen dat de test gemaakt is onder supervisie, hetgeen niet het geval was in de ongecontroleerde conditie. In totaal zijn er 1815 respondenten geweest die de test volledig hebben gemaakt. Met behulp van deze gegevens zijn door middel van een t-toets voor onafhankelijke paren de gemiddelden van de verschillende groepen vergeleken en is de effectgrootte berekend met behulp van Cohen's d. Uit de resultaten blijkt dat de mensen in de gecontroleerde conditie gemiddeld hogere scores behaalden dan mensen in de ongecontroleerde conditie. Daarnaast waren de scores bij de papieren afnamevorm gemiddeld hoger dan bij de online afnamevorm. In de discussie worden zowel de implicaties van de resultaten als enkele kanttekeningen bij dit onderzoek beschreven.

Abstract

This study examined whether the medium used and the condition in which a test is made are of influence on the outcome of an intelligence test. In order to examine this, an intelligence test, which was under construction, was used.

This test was made in a proctored setting and in an unproctored setting and as well as on paper and online. The proctored test took place under supervision of a supervisor, who was not present in the unproctored setting. A total of 1815 respondents completed the whole test. Using these data a t-test for independent pairs was used to examine the means of these groups and the effect size was calculated using Cohen's *d*. The results showed that people in the proctored setting had mean scores that were higher than the scores of the people in the unproctored setting. In addition the scores of the paper test were on average higher than the scores on the online test. Implications of the results as well as the side notes are described in the discussion.

Achtergrond en mogelijkheden

De laatste jaren neemt het computergebruik en ook het gebruik van Internet toe. In bedrijven is de computer niet meer weg te denken en ook in huishoudens beschikken steeds meer mensen over een computer. In Nederland had in 2006 88% van de bevolking toegang tot een computer en 80% van de huishoudens had in 2006 toegang tot internet (Centraal Bureau voor de Statistiek, 2007). Ook binnen de psychologie zijn de computer en het medium Internet ontdekt vanwege de nieuwe mogelijkheden. Het Trimbos instituut presenteerde recentelijk in een persbericht de uitkomst van een onderzoek waaruit blijkt dat therapie via het internet even effectief kan zijn als reguliere therapie (Trimbos Instituut, 2007). Maar het gebruik van computers en het internet biedt niet alleen mogelijkheden voor therapie. Er verschijnen steeds meer psychologische tests die via de computer te maken zijn en sommigen daarvan zelfs online waardoor de test niet op de computer geïnstalleerd hoeft te zijn waarop de test wordt gemaakt. Dit maakt het mogelijk om mensen op afstand testen af te nemen. Het is niet langer noodzakelijk voor mensen om naar een centrale testlocatie te gaan, maar ze kunnen de test thuis maken. Dit betekent meer flexibiliteit, waardoor mensen niet meer genoodzaakt zijn testen te maken tijdens werktijd (Lievens, 2003) Op deze manier kan een selectieproces van personeel bijvoorbeeld efficiënter plaatsvinden. Mensen die niet voldoende scoren op een bepaalde test hoeven niet op gesprek te komen. Op deze wijze wordt er geen tijd en energie gestoken in kandidaten die niet over de juiste kwaliteiten beschikken.

De voordelen van computertesten zijn onder andere de grote mate van standaardisatie (consistente instructie, precieze timing) en het feit dat de test snel en accuraat gescoord kan worden (Tippins. et. al., 2006). Nadelen zijn de mogelijke hardware problemen, bijvoorbeeld een computer die vastloopt of de internetverbinding die verbroken wordt (Tippins et. al., 2006; Mooney, 2002), maar ook het feit dat mensen die niet gewend zijn om met computers te werken, moeite hebben met deze wijze van testen (Potosky, 2004; Buchanan, 2002).

Gecontroleerd vs. ongecontroleerd

Op het gebied van online testen worden er verschillende vormen onderscheiden. The International Test Commission (ITC) onderscheidt vier vormen online testen (International Test Commission, 2001);

- 1) *Open mode*: geen interactie met supervisor, vaststellen identiteit niet mogelijk.
- 2) *Controlled mode*: test is alleen toegankelijk voor mensen die een gebruikersnaam en wachtwoord hebben.

- 3) *Supervised (Proctored) mode*: supervisor aanwezig die de identiteit van kandidaat kan vaststellen.
- 4) *Managed mode*: hoge mate van supervisie, grote mate van controle door supervisor.

Anderen spreken van 2 vormen, namelijk proctored (gecontroleerd) en unproctored (ongecontroleerd) (Ployhart, 2003; Tippins et. al., 2006). Proctored online testing houdt in dat de kandidaat de test maakt onder toezicht van een supervisor, meestal op een vaste locatie. Unproctored online testing houdt in dat de kandidaat de test maakt op een willekeurige plek met internetaansluiting en zonder toezicht van een supervisor (Ployhart, 2003).

Voor- en nadelen van testen via internet

Er is in de literatuur discussie over de voor- en nadelen en de haken en ogen die komen kijken bij het testen via Internet.

Voordelen van testen via het internet zijn de consistente instructie, de precieze timing en een snelle en accurate scoring. Andere voordelen zijn onder andere het feit dat men niet meer aan een locatie gebonden is (Chuah, 2006), dat veranderingen in de test gemakkelijk doorgevoerd kunnen worden (Lievens, 2003; Davis, 1999) en dat het complete en accurate databases voor onderzoeksdoeleinden op kan leveren (Davis, 1999). Tevens lijken mensen, wanneer ze een test via internet maken, eerlijker te zijn en minder sociaal wenselijk te antwoorden dan wanneer ze een paper & pencil test maken, omdat ze het gevoel van anonimiteit hebben (Joinson, 1999; Rosenfeld 1991).

Er is ook een aantal nadelen. Wanneer iemand de test thuis maakt kan de identiteit van deze persoon niet vastgesteld worden (Tippins et. al., 2006; Buchanan, 2002). Hoewel er een hoop maatregelen kunnen worden genomen, zoals het gebruik van een inlogcode, controleren van het IP-adres en het gebruiken van cookies, is er nog geen maatregel die 100% zekerheid geeft over de identiteit van de persoon die de test maakt (Buchanan, 2002). Tevens draagt de thuisomgeving bij aan een minder gestandaardiseerde testomgeving. Thuis kunnen mensen afgeleid worden door omgevingsgeluiden (Buchanan, 2002) of de test halverwege onderbreken om andere activiteiten te ondernemen. Allemaal zaken die de testuitslag kunnen beïnvloeden. Daarnaast heeft niet iedereen hetzelfde computersysteem en dezelfde bandbreedte, wat van invloed kan zijn op de testuitslag. In het geval van cognitieve testen kunnen mensen gebruik maken van hulpmiddelen of anderen om hulpvragen zonder dat dit gecontroleerd kan worden.

Wanneer ervoor gekozen wordt om unproctored te gaan testen, dient er met een aantal zaken rekening gehouden te worden (Tippins et. al. e.a., 2006), nl.:

- *De aard van de test: cognitieve test óf persoonlijkheidstest.*
- *Het doeleinde van de test: selectie óf ontwikkeling.*

Wanneer een test gebruikt wordt voor iemand persoonlijke ontwikkeling, is het minder waarschijnlijk dat deze sociaal wenselijk wordt ingevuld (Naglieri e.a., 2004)

- *het belang van de testscore voor de kandidaat (high stakes vs. low stakes)*
- *het effect van cheating op de validiteit en rank order*
- *de middelen om cheating tegen te gaan of te elimineren en de daaraan verbonden kosten*

Het ITC heeft richtlijnen opgesteld met betrekking tot computer- en internettesten (International Test Commission, 2001).

Deze zijn onderverdeeld in 1) technische factoren, 2) kwaliteit, 3) controle en 4) veiligheid.

1. Technische factoren:

- Hardware en software vereisten
- De robuustheid van de test
- De presentatie van het materiaal
- De mogelijkheid tot aanpassingen voor gehandicapten
- Helpfunctie, oefenitems en informatievoorziening in de test

2. Kwaliteit:

- Juiste kennis en vaardigheden van het testen, en een juist gebruik van internettesten
- Goede psychometrische eigenschappen van de test
- Equivalentie, wanneer de test afgeleid is van een paper&pencil test
- Accurate scoring en analyse van de testresultaten
- Juiste interpretatie van resultaten en juiste feedback
- Gelijke toegang voor alle groepen

3. Controle:

- Omschrijf gedetailleerd hoeveel controle in de verschillende testcondities nodig is
- Omschrijf gedetailleerd wat de gewenste controle is wat betreft de supervisie van het testen
- Bekijk de mogelijkheden die er zijn om te controleren of er sprake is van oefening voorafgaande aan de test of dat er sprake is van 'item-exposure'

- Bekijk de mogelijkheden die er zijn om de identiteit te controleren en om te controleren op 'cheating'
4. Veiligheid:
- Veiligheid van testmaterialen
 - Veiligheid van data die via het Internet verstuurd worden
 - Waarborg de vertrouwelijkheid van de resultaten van de kandidaat

Hypothesen

Om een selectieproces efficiënter te laten verlopen kunnen mensen, zoals eerder genoemd, testen thuis maken via het Internet. Door het gebrek aan controle door een testleider is deze situatie ongecontroleerd. Hierdoor is de mogelijkheid aanwezig dat mensen de testuitslag proberen te manipuleren met het doel zich beter voor te doen dan ze zijn. Deze manipulatie kan plaatsvinden door hulpmiddelen te gebruiken, door andere mensen te raadplegen of, in het meest extreme geval, door iemand anders de test te laten maken. Er bestaat discussie over de mate waarin mensen vals spelen. In een artikel dat in 2006 verscheen in 'Personnel Psychology' zeggen Beaty en Shepherd dat er bijna geen sprake is van vals spelen onder topscoorders in een test (Tippins et. al., 2006). In datzelfde artikel spreekt Segall dit tegen. Volgens hem speelt 50% van de mensen die een test thuis maken vals. Tevens lijkt het verschil te maken of er veel van de uitslag van de test afhangt voor de testmaker. Hoe meer er vanaf hangt, hoe groter de kans dat mensen vals gaan spelen (Tippins et. al., 2006). Om vals spelen tegen te gaan wordt door verschillende auteurs (Tippins et. al., 2006, Nye et. al., 2007) de mogelijkheid genoemd om in twee stappen te testen. De kandidaat maakt eerst een ongecontroleerde test en wanneer hij een voldoende resultaat behaalt, maakt hij eenzelfde test in een gecontroleerde situatie. Nye et. al.(2007) presenteerden op het congres van het SIOP (Society for Industrial and Organizational Psychology) in 2007 in New York een onderzoek waarin onderzocht werd in welke mate er vals gespeeld werd in een ongecontroleerde situatie, door de uitkomsten te vergelijken met de uitkomsten van een gecontroleerde test die later door dezelfde kandidaten gemaakt was. Zij vonden geen aanwijzingen dat er sprake was van vals spelen in de ongecontroleerde situatie.

Wanneer het gaat over online testen dient verder rekening te worden gehouden met de equivalentie met een papieren versie van een test. Volgens Buchanan (2002) moet de equivalentie goed onderzocht worden. Er mag niet klakkeloos vanuit worden gegaan dat een Internet test die gebaseerd is op een papieren test daadwerkelijk gelijk is aan deze test.

Preckel et. al. (2003) toonden in hun onderzoek aan dat zowel een online als een papieren versie van een intelligentietest valide en betrouwbare data op kunnen leveren. In een meer recentelijk onderzoek deden Chuah et. al. (2006) vergelijkbare bevindingen. Zij stelden dat het niet uitmaakt welk medium je inzet voor een persoonlijkheidstest.

Ployhart et. al. (2003) vergeleken in hun onderzoek een gecontroleerde online test met een gecontroleerde papieren test in een selectie context. Daarbij werden er wel kleine verschillen gevonden die in het voordeel van de online test uitpakten. Deze had onder andere een meer normale verdeling en een hogere interne consistentie.

Er zitten dus zowel voor- als nadelen aan ongecontroleerd testen. De vraag is in hoeverre de uitslag van de test beïnvloed wordt door het feit dat dit gecontroleerd of ongecontroleerd gebeurt. Er wordt in de literatuur genoemd dat mensen vals kunnen spelen, maar het is onduidelijk óf en in welke mate dat gebeurt. In dit onderzoek wordt onderzocht of er inderdaad sprake is van verschillen tussen gecontroleerde en ongecontroleerde afname. Daarnaast wordt onderzocht of de afnamevorm invloed heeft op de uitslag van een intelligentietest. Op basis van de bevindingen van Nye et. al. (2007) dat er geen verschil bestaat tussen gecontroleerd en ongecontroleerd testen en de bevindingen van Chuah et. al. (2006) dat het medium niet uitmaakt worden in dit onderzoek de volgende hypothesen getoetst:

Hypothese 1: De conditie (gecontroleerd of ongecontroleerd) waarin een intelligentietest wordt afgenomen heeft geen invloed op de uitkomst van de testscore.

Hypothese 2: De afnamevorm van een intelligentietest (papier of online) heeft geen invloed op de uitkomst van de testscore.

Daarnaast wordt er van uitgegaan dat er duidelijk verschillende scores gevonden zullen worden voor de verschillende opleidingsniveaus. Van mensen met een WO opleiding wordt verwacht dat ze hoger scoren dan mensen met een HBO opleiding. Daarom is hypothese 3 opgesteld.

Hypothese 3: Respondenten met WO-niveau scoren gemiddeld hoger dan respondenten met HBO-niveau

Tevens wordt onderzocht of er verschillen bestaan tussen de scores van mannen en vrouwen. Er is geen reden om aan te nemen dat vrouwen en mannen significant andere scores zullen behalen (Halpern, 1997). Vandaar dat hypothese 4 is opgesteld.

Hypothese 4: Mannen en vrouwen scoren, in zowel de gecontroleerde als de ongecontroleerde conditie, gemiddeld even hoog

Aan het eind van de test konden de respondenten nog een aantal evaluatievragen beantwoorden. Hierbij werd onder andere gevraagd of ze werden gestoord tijdens het maken van de test en of ze de test geconcentreerd hadden gemaakt.

Daarnaast is het plausibel dat de mensen die de test geconcentreerd gemaakt hebben hogere scores behalen dan de mensen die de test niet geconcentreerd gemaakt hebben.

Hypothese 5: Mensen die geconcentreerd waren gedurende het maken van de test scoren hoger dan mensen die niet geconcentreerd waren.

Het is aannemelijk dat de mensen die gestoord werden tijdens de test een lagere score hebben behaald. Daarom is de volgende hypothese opgesteld:

Hypothese 6: Mensen die tijdens de test gestoord worden, scoren lager dan mensen die niet gestoord worden.

Procedure

Voor dit onderzoek is gebruik gemaakt van een online intelligentietest en een papieren versie van deze intelligentietest. De test is ontwikkeld door een Nederlands adviesbureau dat onder andere psychologische tests ontwikkelt. Op het moment van dit onderzoek was deze test in de eindfase van de ontwikkeling.

De test bestond uit 4 onderdelen, namelijk cijferreeksen, figurenreeksen, diagrammen en ravenitems. Elk onderdeel bestond uit 15 items. In totaal waren er dus 60 items. Deze intelligentietest is zowel gecontroleerd als ongecontroleerd afgenomen en zowel online als op papier om te controleren of het gebruikte medium verschil maakt. Bij de gecontroleerde conditie was er sprake van twee afname vormen, namelijk een papieren versie en een online versie. De gecontroleerde afnamen vonden plaats in een ruimte waar een supervisor aanwezig was die toezicht hield om er zeker van te zijn dat mensen geen hulpmiddelen gebruikten. In het geval van de ongecontroleerde afnamen konden mensen de intelligentietest op elke computer met een internetverbinding maken. In de papieren versie van de test kwamen dezelfde items voor als in de online versie. Het verschil was dat er meerdere items op één pagina stonden, terwijl bij de online versie de items één voor één werden aangeboden. Bij de online conditie konden mensen na het maken van de test een evaluatieformulier invullen waarin ze onder andere aan konden geven of ze gestoord waren gedurende de test en of ze de test serieus hadden ingevuld.

Deelnemers

In totaal zijn er 2115 respondenten die de intelligentietest hebben ingevuld. Van deze 2115 respondenten hebben 1886 mensen deze test volledig ingevuld. Het grootste deel van deze respondenten werd via Intomart geworven. Dit is een bedrijf dat zich gespecialiseerd heeft in marktonderzoek en over een database beschikt met mensen die zich beschikbaar stellen als deelnemer in een onderzoek. De overige respondenten waren voornamelijk studenten van verschillende hogescholen en universiteiten in Nederland. De data in dit onderzoek zijn sterk scheef verdeeld. Dit is het gevolg van het feit dat het hier een calibratiestudie betrof voor een computer adaptieve test, waarbij sprake was van een grote steekproef. Daarnaast bleek het in de praktijk lastig om voldoende gecontroleerde afnamen te verzamelen, met als gevolg de volgende verdeling van data. De respondenten waren verdeeld over twee condities, namelijk gecontroleerd (9,9%) en ongecontroleerd (90,1%). Daarnaast waren respondenten verdeeld over twee afname vormen, namelijk de online versie (93,4%) en een papieren versie (6,6%). De papieren versie werd alleen in een gecontroleerde conditie

afgenomen. Hierdoor ontstonden er drie groepen, namelijk gecontroleerd/online, ongecontroleerd/online en gecontroleerd/papier. Van de respondenten was 42,2% man en 57,6% vrouw. Alle respondenten waren Nederlandse autochtonen. Het opleidingsniveau van de respondenten was HBO (46,4%) of WO (53,6%). De leeftijd van de kandidaten is onderverdeeld in drie groepen, namelijk 'tot 30 jaar' (41,4%), '30 tot 45 jaar' (32,5%) en 'ouder dan 45 jaar' (26,1%).

Analyses

De test was zo ontwikkeld dat er na het invullen van de test direct een dataset beschikbaar was die in SPSS geïmporteerd kon worden. Omdat de te ontwikkelen test een adaptieve test wordt, is Item Response Theorie (IRT) gebruikt. Bij IRT worden item karakteristieken (item parameters) en karakteristieken van een individu (latente trek) gekoppeld aan de kans op een positieve response. Er zijn verschillende IRT modellen ontwikkeld voor zowel dichotome als polytome data. De kans op een goed antwoord kan grafisch worden weergegeven in een item response functie (IRF). Deze functies geven de non-lineaire regressie weer van de antwoordkans op een latente trek, zoals bijvoorbeeld consciëntieusheid of verbale capaciteit.

Ten opzichte van klassieke testtheorie biedt IRT enkele voordelen. In tegenstelling tot klassieke testtheorie zijn bij IRT de item- en persoonsparameters onveranderlijk (invariant). Dit biedt de mogelijkheid om de bijdrage van individuele items te onderzoeken door ze toe te voegen of weg te laten uit een test. Daarnaast stelt IRT onderzoekers in de gelegenheid om items te selecteren die een bepaalde capaciteit/trek binnen een range met veel precisie kunnen meten. IRT biedt ook de mogelijkheid tot computer adaptief testen (CAT). Hierbij kunnen items gekozen worden die de meest informatie geven over elke kandidaat. Hierdoor kan de tijdsduur van een test verkort worden (University of Illinois IRT laboratory). In dit onderzoek werden de ruwe scores van de kandidaten omgezet in theta waarden zodat voor elke subtest van iedere respondent een theta waarde beschikbaar was. De theta konden waarden van -4 tot 4 aannemen.

Om te toetsen of er verschillen waren is de t-toets voor onafhankelijke paren gebruikt. Gezien het grote aantal respondenten zijn de verschillen al snel significant. Daarom is ervoor gekozen de effectgrootte te berekenen. Deze wordt bepaald aan de hand van de waarde van Cohen's d (Cohen, 1988). Als vuistregel geldt dat wanneer Cohen's d bijvoorbeeld 0,2 is, de effectgrootte klein is, bij een waarde van 0,5 is de effectgrootte normaal en bij een waarde van 0,8 is de effectgrootte groot te noemen. Hoe groter dus Cohen's d, hoe groter de effectgrootte.

Op deze manier kan bepaald worden wat de omvang van een effect is. De effectgrootte geeft dus aan hoe groot de invloed van een bepaalde variabele is op de uitslag van de test. Cohen's d is gebruikt om de effectgrootte van de verschillen tussen de variabelen conditie, afnamevorm, geslacht, opleidingsniveau en leeftijd te berekenen. Daarnaast is gebruik gemaakt van variantieanalyse om de verschillen tussen de verschillende groepen te bestuderen.

Resultaten

Cijferreeksen. In Tabel 1 staan de gemiddelden, de standaardafwijkingen en Cohen's d van de thetawwaarden van de respondenten voor de subtest cijferreeksen. Wanneer de t -toets voor onafhankelijke paren wordt gebruikt om te controleren of er verschillen optreden tussen de gecontroleerde en de ongecontroleerde conditie, wordt er een significant verschil gevonden ($t=-5.937$; $p<0.005$). De effectgrootte van het gevonden verschil is -0.37 . Dit houdt in dat er voor de ongecontroleerde afname een klein negatief effect gevonden wordt. De afnamevorm levert ook significante verschillen op ($t=4.911$; $p<0.005$) waarbij de effectgrootte 0.40 is. Afname via papier heeft dus een klein positief effect op de score. Ook voor opleidingsniveau zijn significante verschillen gevonden ($t=-7.465$; $p<0.005$). De effectgrootte is 0.33 waardoor er een klein positief effect bestaat voor het WO-niveau. Mannen scoren gemiddeld net iets hoger dan vrouwen, maar dit effect is klein. Wat leeftijd betreft zijn er ook verschillen gevonden. De oudere groepen scoren gemiddeld lager, maar ook dit effect is klein. Deze verschillen lijken in alle vier de subtests hetzelfde te zijn.

Tabel 1 Cijferreeksen: gemiddelden, standaardafwijkingen en Cohen's d

| Variabele | | M | SD | Leeftijds groep | Cohen's d |
|------------------|-----------------|-------|------|-----------------|-------------|
| Conditie | Gecontroleerd | 0.29 | 2.08 | | |
| | Ongecontroleerd | -0.38 | 1.50 | | -0.37 |
| Afnamevorm | Online | -0.36 | 1.56 | | |
| | Papier | 0.32 | 1.78 | | 0.40 |
| Opleidingsniveau | HBO | -0.58 | 1.44 | | |
| | WO | -0.07 | 1.66 | | 0.33 |
| Geslacht | Man | -0.11 | 1.78 | | |
| | Vrouw | -0.46 | 1.39 | | -0.21 |
| Leeftijd | >30 | -0.13 | 1.65 | >30 | -0.09 |
| | 30-45 | -0.29 | 1.47 | 30-45 | -0.21 |
| | >45 | -0.61 | 1.58 | >45 | -0.29 |
| | >45 | -0.61 | 1.58 | >30 | -0.29 |

Diagrammen. In Tabel 2 staan de gemiddelden, de standaardafwijkingen en Cohen's d. Voor de subtest diagrammen zijn alle gevonden verschillen significant. Bij de gecontroleerde afname is de score gemiddeld hoger dan in de ongecontroleerde afname ($t=-11.347$; $p<0.005$). De effectgrootte is -0.88 , wat wil zeggen dat er een groot negatief effect is voor de ongecontroleerde afname. Bij deze subtest zijn de scores van de online afnamen gemiddeld hoger dan de scores van de papieren versie ($t=10.113$; $p<0.005$). De effectgrootte is -0.94 . Dit effect is groot te noemen. WO'ers scoren gemiddeld hoger dan HBO'ers ($t=-10.634$; $p<0.005$) waarbij de effectgrootte met 0.46 matig positief is voor het WO-niveau. Vrouwen behalen op deze subtest gemiddeld hogere scores dan mannen, maar dit is met een effectgrootte van 0.07 een zeer klein positief effect te noemen.

Tabel 2 Diagrammen: gemiddelden, standaardafwijkingen en Cohen's d

| Variabele | | M | SD | Leeftijds groep | Cohen's d |
|------------------|-----------------|-------|------|-----------------|-----------|
| Conditie | Gecontroleerd | 0.41 | 1.07 | | |
| | Ongecontroleerd | -0.72 | 1.47 | | -0.88 |
| Afnamevorm | Online | 0.52 | 1.47 | | |
| | Papier | -0.68 | 1.07 | | -0.94 |
| Opleidingsniveau | HBO | -0.94 | 1.39 | | |
| | WO | -0.29 | 1.48 | | 0.46 |
| Geslacht | Man | -0.64 | 1.41 | | |
| | Vrouw | -0.54 | 1.55 | | 0.07 |
| Leeftijd | >30 | -0.40 | 1.39 | >30 | -0.18 |
| | 30-45 | -0.65 | 1.43 | 30-45 | -0.13 |
| | >45 | -0.84 | 1.62 | >45 | -0.29 |
| | | | | >30 >45 | |

Figurenreeksen. In Tabel 3 staan de gemiddelden, de standaardafwijkingen en Cohen's d van de thetawwaarden van de respondenten voor de subtest figurenreeksen. Ook hier zijn alle verschillen significant. Als gekeken wordt naar de afnameconditie, is te zien dat de scores in de ongecontroleerde conditie gemiddeld lager zijn ($t=-10.035$; $p<0.005$) en dat dit een normaal negatief effect is ($d=-0.58$). Ook hier wordt een hoger gemiddelde gevonden voor de papieren afnamen ($t=10.488$; $p<0.005$) net als bij de subtest cijferreeksen. Er bestaat een normaal positief effect ($d=0,66$) voor de papieren afnamevorm. WO-opgeleiden scoren op deze test gemiddeld hoger dan HBO-opgeleiden ($t=-7.470$; $p<0.005$). Er is sprake van een klein positief effect voor WO-opgeleiden. Vrouwen scoren gemiddeld lager dan mannen, hetgeen met een effectgrootte van -0.04 een heel klein negatief effect genoemd kan worden.

Tabel 3 Figurenreeksen: gemiddelden, standaardafwijkingen en Cohen's d

| Variabele | | M | SD | Leeftijds groep | Cohen's d |
|------------------|-----------------|-------|------|-----------------|-----------|
| Conditie | Gecontroleerd | 1.17 | 2.14 | | |
| | Ongecontroleerd | 0.12 | 1.40 | | -0.58 |
| Afnamevorm | Online | 0.14 | 1.40 | | |
| | Papier | 1.44 | 2.41 | | 0.66 |
| Opleidingsniveau | HBO | -0.03 | 1.39 | | |
| | WO | 0.47 | 1.62 | | 0.33 |
| Geslacht | Man | 0.28 | 1.60 | | |
| | Vrouw | 0.21 | 1.48 | | -0.04 |
| Leeftijd | >30 | 0.51 | 1.71 | >30 | -0.17 |
| | | | | 30-45 | |
| | 30-45 | 0.24 | 1.41 | 30-45 | -0.31 |
| | | | | >45 | |
| | >45 | -0.17 | 1.55 | >30 | -0.45 |
| | | | >45 | | |

Ravenitems. In tabel 4 staan de gemiddelden, de standaardafwijkingen en Cohen's d. Ook hier geldt dat alle gevonden verschillen significant zijn. In deze subtest zijn de gemiddelde scores in de ongecontroleerde conditie lager dan die in de gecontroleerde conditie ($t=-9.073$; $p<0.005$). Met een effectgrootte van -0.58 is hier sprake van een normaal negatief effect. Wat betreft de afnamevorm is te zien dat de mensen die de papieren versie hebben gemaakt gemiddeld hoger scoren dan de mensen die de online versie hebben gemaakt ($t=6.623$; $p<0.005$). Er bestaat een normaal positief effect voor de papieren afnamevorm ($d=0.48$). WO-opgeleiden scoren gemiddeld hoger dan HBO-opgeleiden ($t=-6.964$; $p<0.005$). Dit is een klein positief effect ($d=0.31$) voor WO-opgeleiden. Vrouwen scoren gemiddeld lager op deze subtest dan mannen ($t=-3.233$; $p<0.005$). Er is sprake van een klein negatief effect voor vrouwen ($d=-0.14$).

Tabel 4 Ravenitems: gemiddelden, standaardafwijkingen en Cohen's d

| Variabele | | M | SD | Leeftijds groep | Cohen's d |
|------------------|-----------------|-------|------|-----------------|-----------|
| Conditie | Gecontroleerd | 0.50 | 0.50 | | |
| | Ongecontroleerd | -0.17 | 1.02 | | -0.58 |
| Afnamevorm | Online | -0.14 | 1.04 | | 0.48 |
| | Papier | 0.45 | 1.35 | | |
| Opleidingsniveau | HBO | -0.27 | 1.00 | | 0.31 |
| | WO | 0.05 | 1.11 | | |
| Geslacht | Man | -0.01 | 1.31 | | -0.14 |
| | Vrouw | -0.16 | 1.01 | | |
| Leeftijd | >30 | 0.07 | 1.13 | >30 | -0.14 |
| | | | | 30-45 | |
| | 30-45 | -0.07 | 1.10 | 30-45 | -0.26 |
| | | | | >45 | |
| | >45 | -0.36 | 0.93 | >30 | -0.40 |
| | | | >45 | | |

Er zijn in dit onderzoek, zoals al eerder gemeld, drie groepen te onderscheiden, namelijk een groep die de online versie in een gecontroleerde conditie maakte, een groep die de online versie in een ongecontroleerde conditie maakte en een groep die de papieren versie maakte in een gecontroleerde conditie. Door middel van variantieanalyse zijn de gemiddelden van deze drie groepen vergeleken om te kijken of er significante verschillen optraden. Bij elke subtest zijn significante verschillen gevonden tussen de drie groepen (zie tabel 7).

Tabel 5 Anova

| Variabele | | F | a |
|----------------|----------------|----------|------|
| Diagrammen | Tussen groepen | 51.70535 | 0.00 |
| Cijferreeksen | Tussen groepen | 16.35688 | 0.00 |
| Figurenreeksen | Tussen groepen | 57.99638 | 0.00 |
| Ravenitems | Tussen groepen | 36.18653 | 0.00 |

Naar aanleiding van deze uitkomsten kan nu uitspraak worden gedaan over de opgestelde hypothesen. H_1 kan worden verworpen. In alle subtests is er een duidelijk aantoonbaar verschil tussen de gecontroleerde en ongecontroleerde conditie. In de gecontroleerde conditie is de gemiddelde score hoger. Op basis van de uitgevoerde onafhankelijke t-toetsen kan H_2 ook worden verworpen. De afnamevorm heeft wel invloed op de uitkomst. Afgezien van de subtest diagrammen, zijn de gemiddelde scores op de overige subtesten op papier hoger. Het feit dat alleen voor de subtest diagrammen het omgekeerde geldt is opmerkelijk. Respondenten met een WO-opleiding scoren op alle subtests gemiddeld

hoger dan respondenten met een HBO-opleiding. Hierdoor is het aannemelijk dat H_3 klopt. De scores van vrouwen zijn gemiddeld iets lager dan die van mannen, behalve op de subtest diagrammen. Op deze subtest scoren vrouwen gemiddeld hoger. De verschillen zijn wel significant waardoor H_4 verworpen wordt. De effectgrootte is echter zo klein dat het hier gaat om erg kleine verschillen in de gemiddelden waardoor de invloed van geslacht klein is.

Naast de verschillen in de gemiddelde scores is voor zowel de gecontroleerde als voor de ongecontroleerde online afnamen ook gekeken naar het aantal mensen dat gestoord is tijdens de afname en of mensen de test geconcentreerd gemaakt hebben. Deze vragen stonden in de evaluatievragenlijst aan het eind van de test. In de gecontroleerde conditie geeft 27,3% aan gestoord te zijn tijdens het maken van de test. In de ongecontroleerde conditie is dit 35,8% (zie tabel 6). Verder geeft 94% van de respondenten in de gecontroleerde conditie aan dat zij de test geconcentreerd gemaakt hebben. In de ongecontroleerde conditie geeft 91,2% aan de test geconcentreerd te hebben gemaakt (zie tabel 7).

Tabel 6 Percentages: Geconcentreerd ja/nee

| Conditie | Geconcentreerd | % |
|-----------------|----------------|------|
| Gecontroleerd | Ja | 94 |
| | Nee | 6 |
| Ongecontroleerd | Ja | 91,2 |
| | Nee | 8,8 |

Tabel 7 Percentages: Gestoord ja/nee

| Conditie | Gestoord | % |
|-----------------|----------|------|
| Gecontroleerd | Ja | 27,3 |
| | Nee | 72,7 |
| Ongecontroleerd | Ja | 64,2 |
| | Nee | 35,8 |

De meeste mensen geven aan de test geconcentreerd te hebben ingevuld. In hoeverre dat verschil maakt voor de score op de test valt te zien in tabel 8. Uit deze tabel blijkt dat de scores van de mensen die de test geconcentreerd hebben gemaakt significant hoger zijn dan van de mensen die aangeven niet geconcentreerd te zijn geweest. Concentratie is dus van invloed op de testscore. Dit komt overeen met hetgeen in H_5 verondersteld werd.

Tabel 8 Concentratie: verschillen in score

| | Geconcentreerd | M | t | a | Cohen's d |
|----------------|----------------|-------|-------|--------|-----------|
| Cijferreeksen | Ja | -0.26 | -4.18 | <0.005 | -0,39 |
| | Nee | -0.82 | | | |
| Diagrammen | Ja | -0.59 | -4.77 | <0.005 | -0,39 |
| | Nee | -1.13 | | | |
| Figurenreeksen | Ja | 0.23 | -5.53 | <0.005 | -0,52 |
| | Nee | -0.41 | | | |
| Ravenitems | Ja | -0.06 | -7.86 | <0.005 | -0,63 |
| | Nee | -0.67 | | | |

In tabel 7 valt te zien dat het aantal mensen dat tijdens de test gestoord werd beduidend groter is in de ongecontroleerde conditie. Opmerkelijk is dat dit geen aantoonbaar nadelig effect heeft gehad op de testscore. In tabel 9 valt op dat de mensen die aangeven gestoord te zijn tijdens de test toch gemiddeld hoger scoren dan degenen die niet gestoord zijn. Deze gegevens zijn echter niet significant. Alleen voor de subtest diagrammen werd een significant verschil gevonden ($t=1.48$; $p<0.05$). Bij deze test scoorden de mensen die gestoord waren lager dan de mensen die niet gestoord waren. Gezien deze resultaten kan H_6 verworpen worden. Er is geen aanwijzing gevonden dat het feit dat iemand al dan niet gestoord werd, invloed heeft op de testscore.

Tabel 9 Gestoord tijdens afname: verschillen in score

| | Gestoord | M | t | a | Cohen's d |
|----------------|----------|-------|-------|------|-----------|
| Cijferreeksen | Ja | -0.20 | -2.06 | 0.14 | -0,10 |
| | Nee | -0.37 | | | |
| Diagrammen | Ja | -0.72 | 1.48 | 0.04 | 0,08 |
| | Nee | -0.60 | | | |
| Figurenreeksen | Ja | 0.19 | -0.66 | 0.51 | -0,03 |
| | Nee | 0.15 | | | |
| Ravenitems | Ja | -0.11 | -0.09 | 0.93 | 0,00 |
| | Nee | -0.11 | | | |

Conclusie en discussie

Doel van dit onderzoek was het onderzoeken in welke mate de afnameconditie en de afnamevorm van een test de testuitslag beïnvloeden. Gezien het toegenomen gebruik van ongecontroleerde online tests binnen de A&O psychologie zijn dit enkele belangrijke kwesties. In dit onderzoek zijn een aantal bevindingen gedaan die relevant zijn wanneer het gaat om de inzet van tests. Wellicht dat bedrijven deze bevindingen mee kunnen nemen in hun beslissing over welke tests in te zetten.

Misschien wel de verrassendste bevinding is het feit dat wanneer mensen zich bevinden in een gecontroleerde setting tijdens het maken van een online test, ze beter scoren dan wanneer ze zich in een ongecontroleerde setting bevinden. Dit in tegenstelling tot wat men zou verwachten op basis van de literatuur. Door Segall (Tippins et. al., 2006) wordt beschreven dat mensen in een ongecontroleerde setting de uitslag zullen proberen te manipuleren om beter uit de verf te komen, hetgeen logischerwijs hogere scores in de ongecontroleerde setting op zou leveren. Het feit dat juist de gecontroleerde groep gemiddeld hoger scoort is dus verrassend. Dit kan wellicht verklaard worden door het feit dat de mensen die de test in een ongecontroleerde setting hebben gemaakt, minder zorg hebben gedragen voor een rustige omgeving. Hierdoor kunnen zij afgeleid zijn door invloeden uit hun omgeving, hetgeen de concentratie niet ten goede komt. Dit zou de lagere score kunnen verklaren.

Echter lijkt het weinig invloed te hebben of mensen al dan niet gestoord zijn tijdens de test. Er zijn geen resultaten gevonden die erop wijzen dat dit een negatieve invloed heeft. Het kan zijn dat mensen wel minder geconcentreerd waren en afgeleid werden, maar hier in de evaluatievragenlijst geen melding van hebben gemaakt of misschien dat ze zich niet bewust waren van het feit dat ze minder geconcentreerd waren. Het zou ook kunnen dat mensen in een gecontroleerde situatie meer druk voelden waardoor ze in staat waren zich beter te concentreren waardoor hun scores gemiddeld hoger uitvielen. Er moet wel worden opgemerkt dat er in dit onderzoek niets van de uitslag van de test afhing. Het zou kunnen dat mensen hierdoor toch minder serieus en geconcentreerd waren tijdens het maken van de test dan wanneer er consequenties aan de uitslag verbonden waren. Wanneer de test in een selectiesituatie had plaatsgevonden, zouden de mensen in de ongecontroleerde conditie wellicht meer zorg hebben gedragen voor een rustige omgeving waarin ze de test hadden kunnen maken.

Concentratie lijkt wel een rol te spelen bij het maken van de test. De mensen die aangaven de test geconcentreerd te hebben gemaakt, scoorden gemiddeld hoger dan de mensen die aangaven niet geconcentreerd te zijn geweest tijdens het maken van de test. Het is dus van

belang dat als het om bijvoorbeeld een selectiesituatie gaat, benadrukt wordt dat mensen er zorg voor moeten dragen dat ze geconcentreerd aan de test kunnen beginnen. Dit geldt voor zowel gecontroleerde als ongecontroleerde afname. In de gecontroleerde conditie zorgt een testleider meestal voor een rustige plek waar men zich goed kan concentreren. Wanneer iemand een test elders, bijvoorbeeld thuis, maakt is het extra van belang te benadrukken dat men zorgt voor zo min mogelijk afleiding teneinde zich zo goed mogelijk te kunnen concentreren. Dit blijft uiteindelijk de verantwoordelijkheid van de kandidaat zelf, maar het is aannemelijk dat iemand die graag door een selectie komt, zelf zorg zal dragen voor een optimale testomgeving

Gezien de resultaten lijkt er van het door Tippins en Segall (Tippins et. al. 2006) genoemde vals spelen in de ongecontroleerde conditie weinig sprake te zijn. Wanneer mensen vals spelen zouden de scores hoger uit moeten vallen in die conditie waarin vals gespeeld is, in dit geval dus in de ongecontroleerde conditie. De scores in deze conditie zijn gemiddeld echter lager dan die in de gecontroleerde groep wat het onwaarschijnlijk maakt dat er vals is gespeeld. Mocht er wel vals gespeeld zijn, dan heeft dit geen positieve invloed gehad op de uitslag van de test. Het feit dat er voor de respondenten niets afhing van de test, kan er aan bij hebben gedragen dat mensen geen behoefte voelden om vals te spelen. Segall (Tippins et. al., 2006) zegt dat hoe meer er van een test afhangt hoe groter de kans dat mensen vals spelen. Dit kan verklaren waarom er in dit onderzoek geen aanwijzingen gevonden zijn dat er is vals gespeeld.

Een andere bevinding is dat de afnamevorm van invloed is op de uitslag van de test. In tegenstelling tot de bevindingen van Chuah (2006) wat betreft persoonlijkheidsvragenlijsten, zijn er in dit onderzoek wel verschillen gevonden tussen de papieren afnamevorm en de online afnamevorm. De papieren afnamevorm levert bij drie subtests gemiddeld hogere scores op dan de online afnamevorm. Op de subtest diagrammen is dit echter omgekeerd en zijn de gemiddelde scores voor de online afnamevorm hoger. Het zou dus kunnen zijn dat mensen de voorkeur geven aan een papieren test en dat dit ook hun score beïnvloedt.

Het verschil tussen de scores van mannen en vrouwen is erg klein. Van de vier subtesten is er geen enkele test waarbij opvallende afwijkende gemiddelde scores voor mannen of vrouwen zichtbaar zijn. Geslacht lijkt dus weinig invloed te hebben op de uitslag van zowel de online test als de papieren test.

Leeftijd lijkt wel een invloed te hebben op de uitslag van de test. De gemiddelde scores worden lager naarmate de leeftijd hoger wordt. Dit is niet echt verrassend aangezien in eerder onderzoek genoemd wordt dat intelligentie af lijkt te nemen naarmate men ouder wordt

(Bugg, 2006). Leeftijd kan echter ook invloed hebben op hoe goed men met een computer om kan gaan. De vraag is in hoeverre dit mee heeft gespeeld in de uitslag van de test. Het is dus de vraag of de lagere gemiddelde score inderdaad het gevolg is van een afname van intelligentie bij oudere mensen of dat de lagere gemiddelde score het gevolg is van het feit dat oudere mensen minder ervaring hebben om met computers te werken en zaken van een computerscherm te lezen. Dit is iets wat verder onderzocht zou kunnen worden.

Voor de praktijk kan dit onderzoek aanleiding zijn om bepaalde keuzes in het selectieproces te verantwoorden. Doordat in dit onderzoek naar voren is gekomen dat de afnameconditie van invloed is op de uitslag van de test, kan men, afhankelijk van het doel van een onderzoek, kiezen voor een gecontroleerde of ongecontroleerde afnameconditie. Wanneer een test onderdeel is van een selectieprocedure is het wenselijk dat deze test in een gecontroleerde situatie plaatsvindt. Dit levert waarschijnlijk de meest optimale score op van een test. Daarnaast kan controle worden uitgeoefend over de plek waar de testafname plaats vindt en kan men de kans dat iemand vals speelt minimaliseren. Voor situaties waar er niets op het spel staat voor de kandidaat, kan wellicht wel gebruik gemaakt worden van een ongecontroleerde afname van een test. Men moet zich er dan wel bewust van zijn dat de scores op deze test waarschijnlijk lager uitvallen dan wanneer deze persoon de test in een gecontroleerde situatie had gemaakt. Om in een selectieprocedure toch gebruik te maken van een ongecontroleerde afnameconditie lijkt het testen in twee stappen een goede manier om te voorkomen dat vals spelen de selectieprocedure beïnvloedt. Wanneer iemand dan op de ongecontroleerde test een hoge score heeft behaald, maar dit in een zelfde test in gecontroleerde setting niet kan waarmaken, is het aannemelijk dat iemand vals gespeeld heeft en kan iemand uit de selectieprocedure worden verwijderd.

Daarnaast lijkt de afnamevorm ook van invloed te zijn. Hoewel veel bedrijven steeds meer gaan werken met online tests, onder andere vanwege het gebruiksgemak, lijkt het erop dat papieren tests gemiddeld hogere scores opleveren. Als er op basis van deze tests belangrijke beslissingen worden genomen, is dit verschil wel iets om bij stil te staan. Afhankelijk van het gebruikte medium en de gehanteerde norm, zou iemand dus onder of boven de norm uit kunnen komen. Stel dat iemand de online test maakt en onder de norm uitkomt, terwijl als hij de papieren test had gemaakt net hoog had gescoord om boven de norm uit te komen, dan kan de afnamevorm bepalend zijn en verstrekkende gevolgen hebben voor de geteste persoon. Gezien dit belang is het raadzaam om in de toekomst meer onderzoek te doen naar de invloed van de afnamevorm.

Er moet een aantal kanttekeningen gemaakt worden bij dit onderzoek. Allereerst vond het onderzoek, zoals al eerder werd genoemd, niet plaats in een selectiesituatie terwijl een intelligentietest vaak wordt ingezet in het kader van selectie. De gevonden resultaten zijn van toepassing in een situatie waarbij er niets afhangt van de gemaakte test. Een aanbeveling is daarom om dit onderzoek te repliceren in een selectiecontext, zodat onderzocht kan worden of de in dit onderzoek gevonden effecten generaliseerbaar zijn naar selectiesituaties. Dit zou voor de praktijk zeer relevant zijn omdat veel bedrijven belang hebben bij deze informatie. Mochten bevindingen uit dit onderzoek ook naar voren komen uit een onderzoek in een selectiesituatie dan biedt dat bedrijven de mogelijkheid om in een ongecontroleerde setting te testen, hetgeen positieve effecten heeft op selectieprocedures. Hierbij kan onder andere gedacht worden aan het snellere verloop van procedures en kostenbesparing omdat bedrijven niet meer over de faciliteiten hoeven te beschikken om tests af te nemen en geen personeel in dienst hoeven te nemen voor de supervisie/begeleiding.

Daarnaast is er in dit onderzoek sprake van een erg scheve verdeling, waardoor de gevonden effecten een vertekend beeld kunnen geven. Het zou beter zijn geweest als de respondenten gelijkmatiger over de verschillende groepen waren verdeeld. In dit onderzoek was er in verhouding maar een kleine groep toegewezen aan de gecontroleerde conditie. Tevens bestond de groep in de online-gecontroleerde afname overwegend uit WO-studenten, hetgeen wellicht de gemiddeld hogere scores in de gecontroleerde conditie verklaart.

Gezien de bevindingen wat betreft afnamevorm in dit onderzoek tegengesteld zijn aan eerder bevindingen door andere auteurs, is het aan te bevelen dat ook hier meer onderzoek naar wordt gedaan. In dit onderzoek is de bevinding dat er voor de papieren test hogere scores werden gevonden dan voor de online test. Het aantal respondenten dat de test op papier maakte was relatief klein, met als gevolg de eerder genoemde scheve verdeling. Nieuw onderzoek met gelijkmatiger verdeelde groepen is dan ook zeker op zijn plaats om de invloed van afnamevorm verder te onderzoeken.

Al met al heeft dit onderzoek inzicht gegeven in de mogelijke invloed van afnameconditie en afname vorm op de uitslag van een intelligentie test. De effecten die gevonden zijn in dit onderzoek zijn niet erg groot waardoor er geen expliciete uitspraken gedaan kunnen worden. Er is meer onderzoek nodig om de omvang van de gevonden effecten te onderzoeken, evenals de mogelijk praktische implicaties die deze effecten kunnen hebben.

Referenties

Bugg, J.M., Zook, N.A., DeLosh, E.L., Davalos, D.B., & Davis, H.P. (2006). Age differences in fluid intelligence: Contributions of general slowing and frontal decline. *Brain and Cognition*, 62, 9-16.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, 33, 2, 148-154.

Centraal Bureau voor de Statistiek. *Onderzoek ICT gebruik bij personen (z.a)*.

Verkregen op 7 september 2007 van

<http://statline.cbs.nl/StatWeb/table.asp?STB=T&LA=nl&DM=SLNL&PA=71098ned&D1=0-32&D3=a&LYR=G1:0&HDR=G2>

Chuah, S.C., Drasgow, F., & Roberts, W.B. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, 40, 359-376.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments and Computers*, 31, 572–577.

Halpern, D.F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, 52, 10, 1091-1102.

International Test Commission. (2001). International guidelines for test use. *International Journal of Testing*, 1, 93–114.

Joinson, A. (1999). Social desirability, anonymity, and internet questionnaires. *Behavior Research Methods, Instruments and Computers*, 31, 3, 433-438.

Lievens, F., & Harris, M.M. (2003). Research on internet recruiting and testing: Current status and future directions. In C.L. Cooper & I.T. Robertson (Eds.) *International Review of Industrial and Organizational Psychology*, vol. 16 (pp. 131-165). Chichester: John Wiley & Sons Ltd.

Mooney, J. (2002). Pre-employment testing on the internet: Put candidates a click away and hire at modem speed. *Public Personnel Management*, 31, 1, 41-52.

Naglieri, J.A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist*, 59, 150-162

Naye, C.D., Do, B., Drasgow, F. & Fine, S. (2007). Running Head: Two-step Testing in Selection. *Paper presented at SIOP conference 2007.*

Ployhart, R.E., Weekley, J.A., Holtz, B.C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology*, 56, 733-752.

Potosky, D., & Bobko, P. (2004). Selection testing via the internet: Practical considerations and exploratory empirical findings. *Personnel Psychology*, 57, 1003-1034.

Preckel, F., & Thiemann, H. (2003). Online versus paper-pencil version of a high potential intelligence test. *Swiss Journal of Psychology* 62, 2, 131-138.

Rosenfeld, P., Giacalone, R., Knouse, S., Doherty, L., Vicino, M., Kantor, J., & Greaves, J. (1991). *Impression management, candor, and microcomputer-based organizational surveys: An individual differences approach.* *Computers in Human Behavior*, 7, 23-32.

Tippins et. al., N.T., Beaty J., Drasgow, F., Gibson, W.M., Pearlman, K., Segall, D.O., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 189-225.

Trimbos Instituut. *Persbericht*. Verkregen op 7 september 2007 van
<http://www.trimbos.nl/default20473.html?back=1&date=1>.

University of Illinois IRT laboratory. Verkregen op 18 november 2007 van
http://io.psych.uiuc.edu/irt/intro_main.asp.