

Marianne Myre Pedersen

Seleksjonssituasjonens betydning på personlighetskårer

Master i arbeids- og organisasjonspsykologi

(Trondheim, Høst 2015)

Norges teknisk-naturvitenskapelige universitet (NTNU)

Fakultet for samfunnsvitenskap og teknologiledelse

Psykologisk institutt

Forord

Proessen med å skrive en masteroppgave har vært krevende, men ikke minst lærerik. Jeg vil rette en stor takk min veileder, Eva Langvik, som har fulgt meg gjennom hele prosessen, guidet meg, kommet med konstruktive tilbakemeldinger, og gode innspill både teoretisk og metodisk. En takk rettes til Magnus Odéen og Cut-e Norge As for tilgang til og hjelp med datamaterialet, samt gode innspill. Takk til Mårten Skjøstad og Stine Beate Pederson og Gjensidige forsikring AS som har bidratt til at oppgaven fikk datamateriale fra en reell seleksjonsprosess. Takk til alle som sa ja til å delta i studien. Til slutt vil jeg takke Nicolay, for all støtte under denne prosessen. Uten deg hadde jeg aldri vært der jeg er i dag. Datamaterialet som er benyttet i studien var allerede innsamlet profildata på personlighetstesten Shapes. Det ene datamaterialet U1 ble hentet fra Cut-e Norge AS, mens datamaterialet i U2 ble hentet fra Gjensidige Forsikring AS. Alt datamaterialet ble anonymisert av Cut-e Norge AS. Problemstillingen ble utformet og analysene gjennomført av undertegnede, med råd fra veileder Eva Langvik. Studien var ikke en del av et større forskningsprosjekt.

Sammendrag

Formålet med denne studien var å undersøke om man kan observere en typisk personlighetprofil for et bestemt firma i en seleksjonssituasjon. Profildata på den semi-ipsative personlighetstesten Shapes ble anvendt. Studien benyttet et uavhengig-utvalgsdesign, med data fra en reell seleksjonsetting (N = 416) og kurssetting (N = 34). Studiens hypoteser ble utforsket ved bruk av korrelasjonsanalyse, uavhengig t-test og hierarkisk multiple regresjonsanalyse. Resultatene viste signifikante gjennomsnittsforskjeller. Jobbsøkere skåret i gjennomsnitt signifikant høyere på dimensjonene: resultatbevisst $d = .43$, samvittighetsfull $d = .35$, og analytisk $d = .49$, mens signifikant lavere på dimensjonen uavhengig $d = .64$. Testsituasjon var en signifikant prediktor for de fire personlighetsdimensjonene. Resultatene viste også at jobbsøkere brukte signifikant lengre tid enn kursgruppen på selve gjennomføringen av testen, $d = .88$. Resultatene indikerer at testsituasjon har en betydning for skårene på personlighetstesten. Det er imidlertid viktig å påpeke at man ikke kan trekkes noen slutning om årsak til forskjellen på bakgrunn av denne studien. Det oppfordres derfor til mer forskning på feltet personlighetstester i personellseleksjon, samt av tester som benytter et semi-ipsativt format.

Innhold

Introduksjon	1
Lav kriterievaliditet	1
Faking, et legitimt problem?	2
Shapes management	3
Formålet med studien	4
Teoretisk rammeverk	5
Faking	5
Hva sier litteraturen	5
Er faking en trussel mot kriterievaliditeten til personlighetstester?	8
Tiltak for å håndtere faking	8
Advarsler	9
Strategier for å oppdage faking	9
Konstruksjon av testformat	9
Hypoteser	13
Metode	14
Design og prosedyre	14
Forskningsetikk	15
Utvalg	15
Instrument	16
Shapes Management	16
<i>Administrering og skåring</i>	16
<i>Valideringsskalaer i Shapes</i>	18
Statistisk analyse	18
Ekstreme skårer	18
Korrelasjonsmatrise for personlighetsdimensjonene i Shapes	19

Uavhengig t-test for testsituasjon og personlighetsdimensjoner	19
Hierarkisk multiple regresjonsanalyse	19
Uavhengig t-test for gjennomføringen av selve testen	20
Resultat	20
Interkorrelasjoner for dimensjonene i Shapes	20
Forskjeller i dimensjonsskårer	22
Hierarkisk multiple regresjonsanalyse	24
Resultatbevisst	24
Samvittighetsfull	24
Analytisk	24
Uavhengig	24
Forskjeller i selve gjennomføringen av testen	25
Diskusjon	27
Hovedfunn	27
Testsituasjon og personlighetsskårer	28
Hvordan påvirker testsituasjon gjennomføringen av testen	29
Styrker og svakheter ved studien	30
Forskningsdesign	30
Studiens utvalg	31
Statistiske analyser	31
Hvordan kan det semi-ipsative formatet til Shapes ha påvirket resultatene?	32
<i>Negative korrelasjoner mellom skalaer</i>	33
<i>Variansen i ipsative skalaer</i>	33
<i>Forholdet mellom ipsative skalaer og andre variabler</i>	33
Implikasjoner og videre forskning	34
Konklusjon	35
Referanser	36

Appendiks A: Personlighetsdimensjonene i Shapes	i
--------------------------------------------------------	----------

Appendiks B: Adalloc	ii
-----------------------------	-----------

Figurer

1	Skjerm bilde ved administrering av Shapes	17
A1	Definisjoner av personlighetsdimensjonene i Shapes	i
B1	Bilde av hvordan adalloc fungerer i praksis	iii

Tabeller

1	Deskriptiv statistikk for kjønn og alder (for U1 og U2), N (%)	16
2	Interkorrelasjonsmatrise for dimensjonene i Shapes	21
3	Deskriptiv statistikk, M , SD og M_{diff}	23
4	Regresjonsanalyser med resultatbevisst, samvittighetsfull, analytisk og uavhengig som avhengige variabler. Situasjon, alder og kjønn som uavhengige variabler.	26

Introduksjon

Dagens arbeidsliv er i stadig endring. I den sammenheng anses ansattes kompetanse som en særlig viktig faktor for bedrifter som ønsker å være konkurransedyktige (Sandal, 2011). Å kunne forutsi individers arbeidsprestasjon er av stor økonomisk betydning for organisasjoner og bedrifter (Schmidt & Hunter, 1998). Det har de siste 20 årene vært en økning i bruken av personlighetstester i personellseleksjon (Fan et al., 2012; Rothstein & Goffin, 2006; Sandal, 2011; Dilchert, Ones, Viswesvaran & Deller, 2006; Converse et al., 2008). Personlighetstester brukes for å vurdere om en kandidats personlige egenhet passer til den gitte stillingen (Sandal, 2011).

Etterhvert som anvendelsen av personlighetstester i personellseleksjon har økt, stilles det også større krav fra både akademia og praktikanter når det kommer til testens kvalitet. I den akademiske verden er det høyt fokus på testen konstruktvaliditet (Anderson & Ones, 2003). Fra et praktisk perspektiv er den viktigste egenskapen til personellvurderingsmetoder, i følge Schmidt og Hunter (1998), deres evne til å predikere yrkesprestasjon. Kommersielle testutviklere har derfor hatt et sterkt fokus på høy kriterievaliditet (Anderson & Ones, 2003).

Anvendelsen av personlighetstester i personellseleksjon har ført til bekymring (se Dilchert et al., 2006) og stor debatt i forskningsfeltet (Morgeson et al., 2007; Ones, Dilchert, Viswesvaran & Judge, 2007). Kritikken mot bruk av personlighetstester i seleksjonssammenheng kan oppsummeres i to hovedpunkt.

Lav kriterievaliditet

Det første punktet går på at det ikke finnes tilstrekkelig bevis for den prediktive validiteten til personlighetstester (se Salgado, Anderson & Tauriz, 2014), altså at de ikke er i stand til å predikere jobbprestasjoner (Morgeson et al., 2007). På bakgrunn av dette kan man ikke forsvare avgjørelser innenfor personellseleksjon tatt på bakgrunn av resultatene fra personlighetstester (Morgeson et al., 2007). Til tross for den sterke kritikken mot personlighetstesters lave validitet er personlighet generelt ansett som en nyttig prediktor i en rekke settinger (se R. Hogan, 2005), og personlighetstester blir

hyppig brukt som verktøy i seleksjonssammenheng både i Norge og resten av verden (Anderson & Ones, 2003; Dilchert et al., 2006; Rothstein & Goffin, 2006).

Faking, et legitimt problem?

Den andre og kanskje hyppigste kritikken rettet mot bruken av personlighetstester i dag, er at kandidater kan fordreie svarene slik at resultatet ikke er i overensstemmelse med virkeligheten (R. Hogan, 2005; Griffith & Peterson, 2006; Christiansen, Burns & Montgomery, 2005; White, Young, Hunter & Rumsey, 2008; Rosse, Stecher, Miller & Levin, 1998). Det er i litteraturen brukt en rekke ulike begreper for å beskrive denne type atferd (se MacCann, Ziegler & Roberts, 2011), inkludert; *self-enhancement*, *response distortion*, *impression management* og *faking* (J. Hogan, Barrett & Hogan, 2007; Rosse et al., 1998; Tett & Christiansen, 2007; Morgeson et al., 2007; Ones, Viswesvaran & Reiss, 1996). Denne studien vil benytte begrepet *faking*, da det tyder på å være det mest brukte begrepet i litteraturen (Griffith & Peterson, 2006, 2011; J. Hogan et al., 2007; Kuncel & Borneman, 2007; Morgeson et al., 2007; Tett & Christiansen, 2007). *Faking* er blitt definert som et bevisst forsøk på å presentere misvisende og/eller villedende informasjon om ens personlighet, interesser, erfaringer, tidligere atferd, og holdninger for å påvirke andre (Kuncel & Borneman, 2007). *Faking* refererer i denne studien til et bevisst forsøk på å presentere en selv på en fordelaktig måte, slik at ens responser står i samsvar med oppfattet situasjon (Paulhus, 1984; Fan et al., 2012).

Både forskere og praktikanter uttrykker bekymring (Barrick, Mount & Judge, 2001; Goffin & Christiansen, 2003; Donovan, Dwight & Hurtz, 2003; White et al., 2008; Rosse et al., 1998), særlig når det gjelder bruk av personlighetstester i *høy-risiko* situasjoner (Christiansen et al., 2005; Barrick et al., 2001; Ones et al., 1996). Testing i høy-risiko situasjoner vil si at testens resultat kan føre med seg viktige konsekvenser for testtaker, for eksempel få tilbud om jobb eller bli forfremmet. På den andre siden vil utfallet av testen i lav-risiko situasjoner ikke ha konsekvenser for testtaker (Khorramdel, 2014). Litteraturen tyder på at det er en generell enighet om at *faking* av

personlighetstester forekommer til en viss grad (MacCann et al., 2011; Morgeson et al., 2007). Det er imidlertid uenighet rundt hvorvidt faking utgjør et problem eller ikke ved bruk av personlighetstester i personellseleksjon (J. Hogan et al., 2007; Morgeson et al., 2007; Ones et al., 1996; Viswesvaran & Ones, 1999; Barrick & Mount, 1996). Noen argumenterer for at det ikke er grunn til bekymring rundt bruk av personlighet i personalseleksjon (Barrick & Mount, 1996; Ones et al., 1996), og at den lave kriterievaliditeten til personlighetstester er et viktigere og større problem (se Morgeson et al., 2007). J. Hogan et al. (2007) har foreslått at tilnærmingene som er brukt til å studere faking ikke er tilstrekkelig og at tidligere forskning på feltet ikke har undersøkt problemet grundig nok. Andre har argumentert for at å gi ønsket respons er et tegn på tilpasningsdyktighet hos individet (R. Hogan, 2005). Til tross for at flere argumenterer for at motivert faking ikke representerer en stor trussel mot bruk av personlighetstester i personalseleksjon (Barrick & Mount, 1996; Hough, Eaton, Dunnette, Kamp & McCloy, 1990; Ones & Viswesvaran, 1998; Ones et al., 1996), har stadig nye studier vist det motsatte. En rekke funn indikerer at faking representerer et legitimt problem (Fan et al., 2012; Viswesvaran & Ones, 1999; Birkeland, Manson, Kisamore, Brannick & Smith, 2006; Morgeson et al., 2007; Donovan et al., 2003; Rosse et al., 1998; Hough, 1998)

Shapes management

Shapes management (heretter Shapes) ble lansert i 2002 av Cut-e Norge AS (2015). Shapes er utviklet med spesielt fokus på personlighetstrekk som er viktige for arbeidslivet. Shapes er bygget opp av 18 yrkesrelaterte personlighetsdimensjoner, som er basert på en kompetansemodell (Cut-e Norge AS, 2015). Testen er utviklet til bruk både i seleksjon, utvikling og rådgivning (Cut-e Norge AS, 2015; Det Norske Veritas, 2015). Det finnes ingen studier eller publiserte artikler av Shapes i vitenskapelige tidsskrift, noe som gjøre Shapes til en ikke-vitenskapelig test. Testen har imidlertid gjennomgått en sertifisering hos Det Norske Veritas (DNV), hvor kvaliteten av testen vurderes gjennom å undersøke testens psykometriske egenskaper, inkludert dens validitet og reliabilitet (Det Norske Veritas, 2015). I løpet av 2015 har over 20.000

personer blitt testet med Shapes (personlig kommunikasjon, Cut-e Norge AS, november 2015). Informasjonen om og studiene av Shapes som presenteres i denne oppgaven er hentet fra testmanualen til Shapes.

Formålet med studien

Formålet med denne studien har vært å undersøke om man kan observere en typisk profil for et bestemt firma i en seleksjonssituasjon. Dette er et viktig spørsmål å undersøke da det kan bringe nyttig kunnskap omkring hvilken betydning gjennomføringen av tester i høy-risikosituasjon kan ha for skårene på testen. Det blir stadig uttrykt bekymring både fra forskere og praktikanter for at individer kan, og vil forvrengt sine skårer på personlighetstester som benyttes i personellseleksjon. Det vil derfor være nyttig å undersøke forskjeller mellom ulike grupper og deres personlighetsskårer. Det er verd å merke seg at det ikke alltid er de testene med mest vitenskapelig dokumentasjon som er de mest brukte på arbeidsmarkedet. Da Shapes er en ikke-vitenskapelig test med et format som avviker fra de fleste standarder, har det vært et ønske at denne studien skulle bidra til utvidet kunnskap omkring Shapes og testens semi-ipsative format, samt hvordan dette kan påvirke individers skårer. Data fra en reell rekrutteringsprosess gjennomført hos Gjensidige forsikring AS (høy-risiko situasjon) og fra kurssetting (lav-risiko situasjon) hos Cut-e Norge AS ble benyttet. Den overordnede problemstillingen for denne studien kan derfor formuleres som følger:

“Hvilken betydning har testsituasjon for skårene på personlighetstesten Shapes?”

Oppgaven vil først redegjøre for relevant teori og empiri. Deretter blir hypotesene for studien fremlagt. Det vil følge en gjennomgang av metoden, hvor det vil bli redegjort for studiens prosedyre, etiske hensyn, utvalg og instrument, samt valg av statistiske analyser. Videre vil resultatene bli presentert og studiens viktigste funn vil bli drøftet i lys av tidligere forskning. Til slutt vil metodiske begrensninger vil bli drøftet, før det avsluttes med implikasjoner for videre forskning og en konklusjon.

Teoretisk rammeverk

Faking

Litteraturen innenfor feltet faking og personlighetstester har fokusert på fire hovedproblemstillinger (se Dilchert et al., 2006). Mange studier er gjort for å undersøke om individer kan forvrengne sine responser (Viswesvaran & Ones, 1999; Jackson, Wroblewski & Ashton, 2000), og hvorvidt det er individuelle forskjeller i evnen og motivasjonen til å fake (McFarland & Ryan, 2000). Forskere har forsøkt å besvare spørsmålet om individer faktisk faker sine responser i høy-risiko situasjoner (Birkeland et al., 2006; Robie, Zickar & Schmit, 2001), og hvilken betydning faking kan ha for avgjørelser i personellseleksjon (Donovan et al., 2003). Effekten av faking på kriterievaliditeten og begrepsvaliditeten til personlighetstester har fått oppmerksomhet (Ellingson, Smith & Sackett, 2001; Stark, Chernyshenko, Chan, Lee & Drasgow, 2001), og til slutt har forskere foreslått og testet effekten av ulike tiltak for å håndtere og identifisere faking på personlighetstester (Holden, Wood & Tomashewski, 2001; Hough, 1998).

Hva sier litteraturen. For å besvare spørsmålet om individer kan fake sine responser på personlighetstester har forskere benyttet ulike tilnærminger. Den første tilnærmingen har ofte benyttet laboratoriestudier for å sammenligne responser fra individer i ulike situasjoner, ved bruk av instruksjoner. Deres formål har vært å avdekke det absolutt maksimale nivået av faking på personlighetstester (Viswesvaran & Ones, 1999). Den andre tilnærmingen har vært interessert i å avdekke det operasjonelle nivået av faking i reelle seleksjonssituasjoner. For å undersøke dette benyttes ofte et forskningsdesign hvor responsene fra forskjellige grupper, for eksempel jobbsøkere og studenter, kursdeltakere, og den generelle befolkningen, blir sammenlignet (Viswesvaran & Ones, 1999).

Flere studier har funnet støtte for at individer kan fake personlighetstester, når instruert til det (Viswesvaran & Ones, 1999; Birkeland et al., 2006; Morgeson et al., 2007; Jackson et al., 2000). Jackson et al. (2000) fant i sin laboratoriestudie at individer som ble instruert til å svare som jobbsøkere hadde høyere skårer enn når de svarte etter

vanlige instruksjoner. Resultatene fra studien viste en betydelig forskjell i gjennomsnittsskårer med nesten et helt standardavvik for tester med likert-skala, mens tester som brukte tvunget-valg format viste en økning på mindre enn en tredjedel av et standardavvik (Jackson et al., 2000).

Sterk kritikk er rettet mot studier av faking i laborariesettinger. Flere har hevdet at funn fra slike studier ikke kan generaliseres til reelle seleksjonssituasjoner (J. Hogan et al., 2007; D.B. Smith & Ellingson, 2002; D. Smith & Robie, 2004; Fan et al., 2012). D. Smith og Robie (2004) hevder at det i laboriestudier bare er evnen til å fake som kan bli undersøkt, da settingen ikke tillater individer å variere i motivasjon til å fake, slik som i reelle settinger.

Forskningen på feltet støtter antakelsen at individer har evnen til å fake skårene sine på personlighetstester, når de blir instruert til å gjøre det (Viswesvaran & Ones, 1999; Birkeland et al., 2006; Morgeson et al., 2007; Jackson et al., 2000). Det er derimot større usikkerhet når det kommer til spørsmålet om individer forvrenger sine svar i virkelige settinger (Dilchert et al., 2006; D. Smith & Robie, 2004; Robie et al., 2001; Viswesvaran, Deller & Ones, 2007). I studier av jobbsøkere i reelle seleksjonssituasjoner blir ofte jobbsøkernes skårer på personlighetstesten sammenlignet med valideringsskalaer eller skårene til andre grupper (ikke-jobbsøkere). Høyere gjennomsnittsskåre hos jobbsøkergruppen benyttes som en indikator på faking (Viswesvaran & Ones, 1999; Dilchert et al., 2006; Hough, 1998; Robie et al., 2001). Denne forskningen bygger på antakelsen at faking grunner i et indre ønske hos individet om å kontrollere bildet av seg selv og presentere en selv i henhold til sosial ønskelighet (Dilchert et al., 2006). Jobbsøkere vil forsøke å fremstille seg i et positivt lys eller i lys av stillingen de søker. På bakgrunn av dette er det antatt at jobbsøkere derfor vil skåre høyere på medmenneskelighet, samvittighetsfull og emosjonell stabilitet (D. Smith & Robie, 2004). Det er ikke mangel på forskning omkring dette spørsmålet, men bevisene synes å være varierte (se D. Smith & Robie, 2004).

Resultatene fra en meta-analyse utført av Birkeland et al. (2006) fant bevis for at individer som søker jobb forvrenger sine skårer for å fremstille seg selv i et positivt lys.

Funnene indikerte at jobbsøkere forvrengte sine skårer på de dimensjonene som ble oppfattet som særlig jobbrelevant (Birkeland et al., 2006). Det ble funnet en betydelig forskjell i gjennomsnittsskårer på fire av personlighetstrekkene i fem-faktor modellen, ekstrovert ($d = .11$), emosjonelt stabil ($d = .44$), samvittighetsfull ($d = .45$) og åpenhet ($d = .13$).

En rekke studier har funnet støtte for at individer i reelle seleksjonsprosesser kan forvrengte sine responser, og at de gjør det (Rosse et al., 1998; Hough, 1998; Robie et al., 2001; Donovan et al., 2003). Birkeland et al. (2006) har benyttet uavhengig-utvalg for å undersøker faking i reelle situasjoner, og rapporterer generelt konsistente gjennomsnittsforskjeller mellom jobbsøkere og ikke-jobbsøkere. Studien viser effekter som varierer mellom .48 og .65 (Birkeland et al., 2006). Dette er, som forventet, lavere enn de effektstørrelsene som generelt er funnet i studier som benytter avhengig-utvalg, .47 til .93 (Birkeland et al., 2006).

I følge MacCann et al. (2011) er det en generell enighet blant forskere på feltet at omtrent en av fire faker på personlighetstester brukt i høy-risiko situasjoner. I en studie utført av Rosse et al. (1998) hvor det ble brukt uavhengig-utvalgsdesign, kom det frem at jobbsøkere i gjennomsnitt skåret 0.69 standardavvik høyerer enn individer som tok testen i en lav-risiko situasjon.

Studier har også funnet støtte for antakelsen om individuelle forskjeller i hvilken grad og hvordan individer faker (Barrick & Mount, 1996; J. Hogan et al., 2007; Morgeson et al., 2007; Rosse et al., 1998; D. Smith & Robie, 2004; Viswesvaran & Ones, 1999; McFarland & Ryan, 2000; Mersman & Shultz, 1998). En meta-analyse av studier som benyttet uavhengig-utvalg for å undersøke faking i reelle situasjoner fant større grad av faking på trekkene planmessig og emosjonell stabilitet hos jobbsøkere, samt en signifikant men mye mindre grad av faking også på trekkene ekstrovert og åpenhet (Birkeland et al., 2006). Dette indikerer at individuelle forskjeller i tendenser til faking i personellseleksjon utgjør et legitimt problem både for praktikanter og forskere (Barrick & Mount, 1991; J. Hogan & Holland, 2003; Salgado, 1997).

Er faking en trussel mot kriterievaliditeten til personlighetstester?.

Når det gjelder hvorvidt faking har en effekt på kriterievaliditeten til personlighetstester viser forskning også her varierte resultater. Flere studier har vist støtte for kriterievaliditeten til personlighetstester (Hough, 1998; Ones et al., 1996; Hough et al., 1990), selv brukt i personellseleksjon (Hough et al., 1990; Barrick, 2005; Ones, Viswesvaran & Schmidt, 1993). Meta-analyser har vist at personlighetsvariabler, spesielt planmessighet, har tilstrekkelig kriterievaliditet til å rettferdiggjøre bruken av personlighetstester i personellseleksjon (Barrick & Mount, 1991; Ones & Viswesvaran, 1998; Ones, 2005; Salgado, 1997; Tett, Jackson & Rothstein, 1991). Andre studier indikerer at faking kan ha en negativ effekt på testens kriterievaliditet (Komar, Brown, Komar & Robie, 2008; Converse, Peterson & Griffith, 2009). Til tross for debatten omkring den mulige trusselen faking utgjør mot personlighetstesters prediktive validitet, blir personlighetstester fremdeles ansett som nyttig i personalseleksjon (Ones, 2005). En ny meta-analyse gjennomført av Salgado et al. (2014) indikerer at personlighetstester er mer valide enn tidligere antatt. Bevis for kriterievaliditeten til personlighetstester har ført til optimisme rundt anvendelsen av slike tester i personalseleksjon (Barrick & Mount, 1991; Hough & Oswald, 2000; Ones et al., 1993; Tett et al., 1991). Det hevdes imidlertid at selv om kriterievaliditeten til personlighetstester ikke blir påvirket av faking (Schmitt & Oswald, 2006), vil faking på personlighetstester kunne ha konsekvenser for utfallet av seleksjonsprosessen (Mueller-Hanson, Heggstad & Thornton III, 2003; Rosse et al., 1998). Hvem som får tilbud om jobb kan bli påvirket gjennom at listen med individer som får tilbud om jobb forandres, med større sjanse for å inkludere de som har fakset (Mueller-Hanson et al., 2003; Rosse et al., 1998).

Tiltak for å håndtere faking

En rekke strategier er foreslått for å håndtere problemet knyttet til sosialt ønskelige svar og faking. Disse strategiene kan deles i to hovedkategorier, strategier for å forhindre faking og strategier for å identifisere de individene som trolig har utøvet slik atferd (Kuncel & Borneman, 2007; Dilchert et al., 2006)

Advarsler. For å forebygge faking på personlighetstester er det foreslått å fraråde testtakere mot å gi sosialt ønskelige svar (Kuncel & Borneman, 2007; Vasilopoulos, Cucina & McElreath, 2005). Dette kan gjøres enten ved å advare om at faking kan bli oppdaget eller at det å gi falske svar vil ha negative konsekvenser (Dilchert et al., 2006; Dwight & Donovan, 2003). Dwight og Donovan (2003) fant i sin studie at advarsler mot faking har en relativt liten effekt. Andre har vist støtte for at advarsler mot faking kan være en nyttig tilnærming for å redusere faking på personlighetstester (Fan et al., 2012). Mer forskning trengs på dette området for å avgjøre om advarsler er en god tilnærming til denne problematikken, eller om de kan ha en negativ effekt på begrepsvaliditeten til personlighetstester, skape negative reaksjoner hos testtakere, eller resulterer i egne skjevheter (Dilchert et al., 2006).

Strategier for å oppdage faking. En strategi for å oppdage faking har vært bruk av skalaer for sosial ønskelighet (løgn-skalaer?) i normative personlighetstester. Sosial ønskelighets-skalaer er ment å identifisere ekstreme skårer og dermed de individene som faker. Selv om korrigerings på bakgrunn av sosial-ønskelighets skalaer har vært den mest brukte strategien (Goffin & Christiansen, 2003) for å identifisere faking på personlighetstester, har studier vist at denne metoden har begrenset nytteverdi (Ellingson, Sackett & Hough, 1999; Sackett & Lievens, 2008; Schmitt & Oswald, 2006; Dilchert et al., 2006).

Konstruksjon av testformat. En strategi som er foreslått for å forhindre faking av personlighetstester bygger på selve konstruksjonen av testen (Kuncel & Borneman, 2007). Testens format kan benyttes for å tilpasse vanskelighetsgrad og transparens av testledd i personlighetstester (Khorramdel, 2014). Ved å gjøre mållkonstruktet mindre tydelig for testtaker, kan konstruksjonen av testen dermed bidra til å redusere faking (Dilchert et al., 2006; Viswesvaran & Ones, 1999; Ziegler, Schmidt-Atzert, Buhner & Krumm, 2007).

Tvunget valg, referert til som forced choice (FC) format ble introdusert som en strategi for å overkomme problemet med faking på personlighetstester (Rust, 1999; Villanova, Bernardin, Johnson & Dahmus, 1994; Bartram, 2007; Christiansen et al.,

2005; Jackson et al., 2000; Khorramdel, 2014; Joubert, Inceoglu, Bartram, Dowdeswell & Lin, 2015). Tvunget-valg-format er i flere studier funnet å signifikant redusere effekten av svarskevheter (response bias), faking og sosialt ønskelige svar (Cheung & Chan, 2002; Jackson et al., 2000; Christiansen et al., 2005; Saville & Willson, 1991). Ved å bruke tvunget-valg-format blir individene tvunget til å velge mellom to eller flere alternativer som er tilsvarende like attraktive, men som måler hvert sitt underliggende konstrukt (Christiansen et al., 2005). Ved å sette testledd som er omtrent like attraktive sammen kan effekten av sosial ønskelighet reduseres, noe som gjør tvunget-valg-format nyttig i seleksjonssettinger (Christiansen et al., 2005; Jackson et al., 2000; Converse et al., 2008). Formatet gjør det umulig for individet å skåre veldig høyt eller veldig lavt på alle dimensjoner (Clemans, 1966; Heggstad, Morrison, Reeve & McCloy, 2006; Baron, 1996), og resulterer dermed i høyere differensiering mellom et individs skalaskårer (Baron, 1996; Brown & Bartram, 1999). Det er også antatt at tvunget-valg-format har mindre transparente testledd sammenlignet med single-stimulus format (Jackson et al., 2000; Martin, Bowen & Hunt, 2002), og er derfor mer robust mot faking (Heggstad et al., 2006; Khorramdel, 2014). På den andre siden har studier vist at individer også kan fake sine responser på personlighetstester som benytter tvunget-valg-format (Heggstad et al., 2006).

Bruken av tvunget-valg format er imidlertid sterkt kritisert, da formatet resulterer (dersom skåret etter klassisk testteori) i helt eller delvis ipsative skårer (se Hicks, 1970), som er vist seg å ha psykometriske utfordringer (Johnson, Wood & Blinkhorn, 1988; Meade, 2004; Tenopyr, 1988; Brown & Maydeu-Olivares, 2011). Matematisk sett er data ipsative når summen av alle innhentede skårer for ethvert individ på den gitte testen er konstant, altså når et gitt antall responser alltid leder til den samme totalsummen (Clemans, 1966). Hver skåre for et individ er avhengig av individets skårer på andre variabler eller testledd, men uavhengig av, og dermed ikke sammenlignbare med resultatet til andre individer (Hicks, 1970). Som et resultat er det bare mulig med intraindividuelle, ikke interindividuelle sammenligninger ved bruk av ipsative skårer (Cattell, 1944; Hicks, 1970; McLean & Chissom, 1986; Baron, 1996; Closs, 1996). I

personellseleksjon brukes testskårene til å gjøre sammenligninger mellom individene. Individens skårer blir dermed tolket fra et normativt perspektiv, slik at høyere skårer indikerer høyere nivå av trekket som måles, sammenlignet med andre individer (Heggestad et al., 2006). Ipsative skårer skiller seg fra normative i at de måler relative heller enn absolutte verdier (Brown & Bartram, 2009). Dette har skapt bekymring for bruk av personlighetstester med tvunget-valg-format i personellseleksjon (Dilchert et al., 2006).

Fra et psykometrisk perspektiv er det flere karakteristikk ved ipsative skårer som er problematiske (se Clemans, 1966). Flere har derfor identifisert ulike problemer assosiert med ipsative skårer (Cattell, 1944; Clemans, 1966; Cornwell & Dunlap, 1994; Johnson et al., 1988; Loo, 1999; Hicks, 1970; Baron, 1996). I tester som benytter tvunget-valg format blir testledd rangert etter grad av preferanse til andre testledd, noe som skaper gjensidig avhengigheten mellom skårer (og dimensjonen som blir målt av disse), som igjen fører til at antakelsen om uavhengige feilskårer (error) ikke blir møtt (Baron, 1996). Flere har derfor stilt seg kritiske til om reliabilitetsanalyse er hensiktsmessig (Meade, 2004; Baron, 1996). Det er generell enighet blant forskere om at format som bruker tvunget valg vanskeliggjør tolkningen av reliabilitetsanalyse, da reliabilitetsmålet blir forvrengt. I hvor stor grad eller hvilken retning er det derimot uenighet om (se Baron, 1996; Johnson et al., 1988; Tenopyr, 1988). Hovedproblemet ved bruk av tradisjonell reliabilitets-statistikk på ipsative skårer er at de bryter med de fleste forutsetningene som ligger til grunn for denne typen analyse.

På bakgrunn av den konstante sluttskåren har ipsative data en rekke uønskede statistiske egenskaper som man ikke finner hos typiske normative data (Clemans, 1966; Hicks, 1970). Flere har konkludert med at ipsative data ikke egner seg til vanlige statistiske analysemetoder på bakgrunn av at summen av variablene er konstant for alle individer (Chan, 2003; Closs, 1996; Cornwell & Dunlap, 1994; Johnson et al., 1988). Et resultat av at summen av alle skalaskårer i ipsative mål er konstant, og dermed har null varians, er at summen av alle kollonner eller rader i interkorrelasjonsmatrisen også vil være lik null (Clemans, 1966). For en typisk normativ skala vil man generelt forvente

positive interkorrelasjoner mellom de samme skårene. Dette er ikke tilfellet for ipsative skalaer som viser et stort antall negative interkorrelasjoner (Baron, 1996; Clemans, 1966), som resulterer i at korrelasjonene mellom skalaene også vil ha negative verdier (Clemans, 1966). Når variansen er lik vil interkorrelasjonene vanligvis bli $-1/(m - 1)$, hvor m er antall skalaer (Clemans, 1966). De negative fordreide verdiene er problematiske og gjør at korrelasjonene mellom ipsative skalaer ikke kan tolkes. Dermed blir resultatene fra både faktoranalyse og reliabilitetsanalyser meningsløse (Closs, 1996; Tenopyr, 1988; Meade, 2004). Det er tydelig at fordreid begrepsvaliditet er ett problem for ipsative tester (Brown & Maydeu-Olivares, 2012b). Kriterievaliditeten til personlighetstester blir også påvirket av de psykometriske egenskapene til ipsative mål (Brown & Maydeu-Olivares, 2012b). Som et resultat av den konstante skalaskåren i ipsative tester vil skalaens korrelasjon til ethvert eksternt kriteriemål også være lik null (Clemans, 1966; Hicks, 1970). På bakgrunn av kritikken over har flere uttrykt sterk kritikk mot bruk av parametriske statistiske analyser på data som er ipsative (Guilford, 1952; Clemans, 1966; Hicks, 1970; Closs, 1996; Johnson et al., 1988). En stor mengde studier har også støttet opp om denne kritikken (Clemans, 1966; Hicks, 1970; Johnson et al., 1988; Tenopyr, 1988; Closs, 1996; Meade, 2004; Baron, 1996). Noen har gått så langt som å fraråde bruk av ipsative personlighetstester i personellseleksjon (Heggestad et al., 2006; Hicks, 1970). Til tross for dette blir ipsative personlighetstester stadig brukt innenfor samfunnsvitenskapelig forskning og i praksis (Meglino & Ravlin, 1998), også i seleksjonssammenheng (Saville & Willson, 1991).

I følge Hicks (1970) kan graden av ipsativitet ha alvorlige konsekvenser for tolkningen av skårer og beslutninger tatt i personellseleksjon, fordi individer med svært ulik absoluttskåre kan ha samme relative disposisjon av egenskaper, og dermed de samme ipsative skårene. Ofte blir begrepet ipsative brukt som et synonym for gjensidig avhengig og refererer til en type avhengighet mellom de variablene som blir målt. I litteraturen er ofte de ulike typene ipsative data blitt behandlet som om de var identiske, noe som kan være problematisk da de innehar ulike psykometriske egenskaper (Meade, 2004).

Flere personlighetstester er designet eller skåret slik at de produserer delvis ipsative skårer (Hicks, 1970; McLean & Chissom, 1986). FC-format kan bli karakterisert som delvis-ipsativt dersom ikke alle kriteriene for å være helt ipsative blir møtt (Hicks, 1970; Clemans, 1966). For eksempel vil tester hvor individer bare delvis rangerer testledd eller hvor alternativene blir vektet ulikt karakteriseres som delvis ipsative (Hicks, 1970; Meade, 2004).

I følge Hicks (1970) kan delvis ipsative tester være mer valide mål i situasjoner som for eksempel personellseleksjon, hvor faking kan være en trussel. Delvis ipsative tester har psykometriske egenskaper både fra ipsative og normative skårer. Ved bruk av delvis ipsative mål kan man dermed utnytte at det er vanskeligere å fordreie svarene på testen, samtidig som man slipper de statistiske begrensningene ved helt ipsative data. Det argumenteres for at ipsative data som oftest har høyere operasjonell validitet enn normative (Bartram, 2007). Det er imidlertid mangel på forskning som har undersøkt denne problemstillingen, og de få resultatene som foreligger er sprikende (Converse et al., 2008; Heggestad et al., 2006; Jackson et al., 2000).

Hypoteser

På bakgrunn av det teoretiske rammeverket som er blitt presentert, vil følgende hypoteser bli undersøkt:

H1: På grunn av det semi-ipsative formatet vil det være negative korrelasjoner mellom flere av personlighetsdimensjonene.

H2: Noen av dimensjonsskårene i høy-risiko og lav-risiko situasjon vil være ulike.

H3: Deltakerne vil bruke lengre tid på testen i seleksjonssettingen.

H4: Svarene vil være mer konsistente i seleksjonssetting.

Metode

Design og prosedyre

Datamaterialet som ble benyttet i denne studien var innsamlet profildata på personlighetstesten Shapes. Datamaterialet ble hentet fra to ulike utvalg, uavhengig-utvalgs-design.

Alle individene i studien har gjennomført personlighetstesten via internettadministrering, da Shapes er et elektronisk web-basert spørreskjema. Alle kandidater mottok på forhånd av testtakingen en mail med informasjon om testen og innlogging. Testen ble administrert via www.mapTQ.com, hvor kandidatene registrerte seg og logget inn med egne brukernavn og passord.

Det første utvalget (U1) var bestod av individers profildata fra personlighetstesten Shapes, gjennomført i forbindelse med sertifiseringskurs hos Cut-e. Det kan derfor si at U1 har gjennomført testen i det man kan kalle en lav-risiko situasjon, det vil si en situasjon hvor testresultatene ikke vil ha viktige konsekvenser for individet. I sammenheng med kursdeltakelse ble alle deltakerne skriftlig informert om, og samtykket til at datamaterialet fra kurset kan brukes til forskning og utvikling (personlig kommunikasjon, Cut-e Norge AS, september 2014).

Datamaterialet (profildata på Shapes) for det andre utvalget (U2) ble hentet fra en reell rekrutterings- og seleksjonsprosess hos et av Norges største forsikringsselskap, Gjensidige forsikring AS. Situasjonen hvor individene i U2 gjennomførte Shapes, kan derfor karakteriseres som en høy-risiko situasjon, hvor testresultatene kan ha konsekvenser for hvorvidt man fikk jobben eller ikke. Det ble sendt ut mail for å innhente samtykke, fra HR-partner hos Gjensidige forsikring AS, på vegne av undertegnede. Mailen inneholdt et informasjonsskriv om studien, hensikten med studien, hvilke data som skulle benyttes, hva datamaterialet skulle brukes til hva det ville bety å delta, at kandidatene stod fritt til å takke ja eller nei til å delta i studien, samt at datamaterialet ville være anonymisert og ikke få noen konsekvens for kandidatens nåværende eller et eventuelt fremtidig forhold til Gjensidige forsikring AS. Det ble videre informert at kandidatene ved å svare positivt på mailen ville samtykke i

deltakelse. Informasjonsskrivet ble sendt ut til totalt 159 personer som hadde gjennomført en rekrutteringsprosess hos Gjensidige forsikring AS i løpet av 2014. Åtte av disse nådde ikke frem grunnet feil e-post adresse. Det ble sendt ut en påminnelse en uke etter første e-post ble sendt ut. Svarprosent var 21 %.

Forskningsetikk

Før studien ble utført ble Norsk Samfunnsvitenskapelig Datatjeneste AS (NSD) kontaktet for å avklare om studien måtte meldes. På bakgrunn av datamaterialet som var ønsket å bruke i studien, allerede innsamlet data, samt at alt datamaterialet skulle anonymiseres før det ble overlevert til undertegnede, ble studien ansett som ikke meldepliktig.

Alt datamaterialet som ble brukt i studien (både for U1 og U2) ble anonymisert av Cut-e Norge AS, før det anonymiserte materialet ble overlevert til analyser. For at datamaterialet for U1 skulle være fullstendig anonymisert ble det nødvendig å kategorisere enkelte demografiske variabler, inkludert alder, antall år utdanning, år med ledererfaring og år utdanning. Dette var også nødvendig for datamaterialet i U2, i tillegg ble det lagt stor vekt på å holde datamaterialet og kandidatenes identitet anonymt, all kontakt gikk derfor via HR-partner hos Gjensidige forsikring AS, som var eier av datamaterialet for U2.

Utvalg

U1 besto av totalt 451 individer, hvorav 35 ble slettet på bakgrunn at ekstreme skårer (se nærmere forklart under uteliggere). Etter denne prosedyren besto U1 av totalt 416 individer, hvorav 289 (71.6 %) var kvinner og 118 (28.4 %) menn. Alder var kategorisert i grupper, med færrest respondenter i den yngste gruppen (<35) for U1. U2 besto av totalt 34 individer, hvorav 16 (47.1 %) kvinner og 18 (52.9 %) menn. Frekvensen av kjønn og alder er presentert i Tabell 1.

Tabell 1

Deskriptiv statistikk for kjønn og alder (for U1 og U2), N (%)

	N	Kjønn		Alder		
		Menn	Kvinner	<35	36-45	>45
U1	416	118 (28.4)	298 (71.6)	99 (23.8)	159 (38.2)	158 (38.0)
U2	34	16 (47.1)	18 (52.9)	18 (52.9)	10 (29.4)	6 (17.6)

Instrument

Shapes Management. Shapes er et adaptivt og computerbasert spørreskjema. Testen bygger på en kompetansemodell med 18 yrkesrelaterte personlighetsdimensjoner: styrende, overbevisende, sosialt sikker, utadvendt, omgjengelig, observerende, veloverveid, resultatbevisst, planleggende, samvittighetsfull, analytisk, teoretisk, oppfinnsom, endringsvillig, uavhengig, ambisiøs, konkurranseorientert, og energisk. Se appendiks A for definisjoner av personlighetsdimensjonen i Shapes. Med høyt fokus på den praktiske nytten, har testen som mål å kunne predikere 18 arbeidsrelaterte kompetanser (Cut-e Norge AS, 2015). Kompetanser blir i testdokumentasjonen til Shapes definert som yrkesrelatert atferd som har betydning for i hvilken grad en person når sin yrkesrelaterte mål og gjennomfører sine oppgaver på en god måte (Cut-e Norge AS, 2014).

Administrering og skåring. Kandidaten blir presentert for ulike utsagn og tar stilling til hvor godt de ulike påstandene stemmer overens med bildet kandidaten har av seg selv. Shapes er basert på Cut-es egne måleteknologi adaptive allocation of consent (adalloc) (Cut-e Norge AS, 2015). Adalloc gir shapes et semi-ipsativt svarformat, en kombinasjon av ipsative og normative elementer (Cut-e Norge AS, 2015). I følge Cut-e Norge AS (2014) gir adalloc-metoden raskere og mer presise målinger av psykologiske konstrukt enn andre kjente psykometriske målemetoder. Bakgrunnen for dette er i følge Cut-e Norge AS (2014) at personlighetsdimensjonene som blir målt blir

vurdert avhengig av hverandre, samt at instrumentet gjennom hele testadministrasjonen tilpasser seg den enkelte kandidatens responser. Denne typen måleteknologi gir dermed profiler som i høy grad skiller mellom kandidater, samtidig som den gir kort administrasjonstid (Cut-e Norge AS, 2015). En av styrkene ved adaloc er i følge Cut-e Norge AS (2014) metodens motstandskraft mot sosial ønskverdighet og faking. I en studie hvor individer blir instruert til å svare i tråd med kravene til en stilling, viste resultatene at gjennomsnittet for inventorer som benyttet adaloc kun økte i ubetydelig grad (3 %). I følge Cut-e Norge AS (2015) er økningen i gjennomssnittsskårer et resultat av at individene i en seleksjonssetting i større grad vil benytte alle poengene for å kunne presentere seg selv på en mest mulig fordelaktig måte (Se appendiiks B for utdypende informasjon om adaloc).

Administreringen av Shapes skjer ved at kandidaten blir presentert for 3 testledd i en blokk, hvor kandidaten skal distribuere 6 poeng etter hvor godt utsagnene eller testleddene passer kandidatens atferd (Cut-e Norge AS, 2015). Alle testledd som blir presentert i en blokk måler forskjellige personlighetsdimensjoner.

The screenshot displays the 'Kartlegging av kompetanser' (Mapping of competencies) interface. At the top left is the 'cut-e' logo. The main heading is 'Kartlegging av kompetanser'. Below this, a question asks: 'Hvor nøyaktig beskriver disse utsagnene din atferd?' (How accurately do these statements describe your behavior?). A note says: 'Del ut poeng ved å benytte pluss- (+) og minus- (-) knappene' (Distribute points by using plus (+) and minus (-) buttons). There are three statements, each with a rating scale of five circles and plus/minus buttons:

- Statement 1: 'Jeg trives med å selge inn en idé' (I enjoy selling an idea). Rating: 2 orange circles filled.
- Statement 2: 'Jeg utvikler oppfinnsomme løsninger' (I develop creative solutions). Rating: 2 orange circles filled.
- Statement 3: 'Jeg liker å arbeide mot utfordrende mål' (I like to work towards challenging goals). Rating: 1 orange circle filled.

Below the statements, there is a section for distributing points: 'Maks 6 poeng å fordele:' (Max 6 points to distribute:). One orange circle is filled in this section. At the bottom right, there are three buttons: 'Avslutt' (End), 'Hjelp' (Help), and 'Neste' (Next).

Figur 1. Skjerm bilde ved administrering av Shapes

Som vist i Figur 1, har kandidaten delt ut to poeng til testleddet “Jeg trives med å selge inn en idé”, to poeng til utsagnet “Jeg utvikler oppfinnsomme løsninger”, og ett poeng til testleddet “Jeg liker å arbeide mot utfordrende mål”. Ett poeng er ikke utdelt. Kandidaten står helt fritt til selv å bestemme hvordan poengene skal fordeles. Kandidaten kan velge å fordele alle poengene til ett testledd eller hvilken fordeling som måtte være ønskelig (1-4-1, 2-3-1, 3-3-0, 0-1-5 osv.). Det er mulig å fordele samme antall poeng til alle testledd, og kandidaten kan dersom ønskelig velge ikke å fordele poeng dersom ingen av testleddene som presenteres i blokken oppleves å beskrive egen atferd (Cut-e Norge AS, 2015).

Valideringsskalaer i Shapes. Det er lagt inn tre kontrollpunkter som kan brukes for å identifisere ekstreme skårer eller profiler med stor variasjon. Den første av disse er en konsistensskala (målt fra 1-9). Konsistens vurderes ut fra testtakers resultat på de 18 personlighetsdimensjonene. Det er anbefalt at konsistensskårer på 1 og 2 bør undersøkes nærmere eller ikke brukes (Cut-e Norge AS, 2015). Det andre kontrollpunktet er poengfordeling. I gjennomsnitt fordeler testtakere 97 % ($SD = 5\%$) av poengene (Cut-e Norge AS, 2013). Intervallet mellom 82 % og 100 % dekker 99,9 % av alle besvarelser. Det anbefales derfor at poengfordeling under 82 % brukes med varsomhet, da profilen vil være vanskelig å tolke (Cut-e Norge AS, 2013). Det siste kontrollpunktet er tidsbruk for å fullføre Shapes. Det tar i gjennomsnitt 17 minutter å fullføre Shapes. Det finnes ingen tidsbegrensning, tidsbruk mellom 10-30 minutter anses som normalt (Cut-e Norge AS, 2013). Det er nærmest umulig å gjennomføre testen på under 7 minutt. Testutviklerene anbefaler at kandidater med testtid under 9 minutter bør undersøkes nærmere (Cut-e Norge AS, 2013).

Statistisk analyse

Datamaterialet ble analysert ved bruk av SPSS versjon 22. Forberedende analyser ble utført for å undersøke at alle antakelser for bruk av analysene ble møtt (Field, 2009).

Ekstreme skårer. Uteliggere representerer tilfeller hvor skårene er vesentlig forskjellig fra alle andre i et bestemt datasett (Field, 2009). For å identifisere uteliggere

ble det gjort en inspeksjon av histogram og boxplots. Boxplots gav en indikasjon av variansen i skårene innen hver gruppe og en visuell inspeksjon av forskjellene mellom gruppene. Å slette uteliggere på bakgrunn av denne inspeksjonen ville vært altfor strengt i dette tilfellet, z-skåre på 3,29 ble derfor benyttet som skjæringspunkt. Dette ekskluderer bare ekstreme skårene som virkelig er uteliggere, ved at 99,9 % av skårene vil falle innenfor 3.29 standardavvik. Det ble gjort et valg å fjerne uteliggere fra kun det ene utvalget (U1). Bakgrunnen for denne avgjørelsen var en antakelse om at kursgruppen kunne være påvirket av at situasjonen ikke hadde betydning og at individene derfor ikke tok gjennomføringen av testen seriøst, mens det i gruppen for jobbsøkere (U2) ble antatt at alle tok testen seriøst. Ekstreme skårer kan oppstå nettopp fordi man forsøker å presentere seg selv mer fordelaktig, og det er nettopp dette som har vært ønskelig å undersøke med denne studien.

Korrelasjonsmatrise for personlighetsdimensjonene i Shapes.

Sammenhengen mellom personlighetsdimensjonene i Shapes ble undersøkt ved bruk av Pearson korrelasjon. Korrelasjonsmatrisen beskriver forholdet mellom dimensjonene, både i styrke og retning (Field, 2009).

Uavhengig t-test for testsituasjon og personlighetsdimensjoner.

Uavhengig t-test ble gjennomført for å besvare H1, om gjennomsnittsskårene til individer i høy-risiko og lav-risiko gruppen var ulike, samt hvilke personlighetsdimensjoner som viste seg å være signifikant forskjellig i de to gruppene. Cohen's d ble utregnet for å undersøke effektstyrker (Cohen, 1992).

Hierarkisk multiple regresjonsanalyse. Hierarkisk multiple regresjonsanalyse ble benyttet for å undersøke hvor mye av variansen i skårer på personlighetsdimensjonene i Shapes som kan forklares på bakgrunn av situasjon, når alder og kjønn er kontrollert for. Hierarkisk multiple regresjonsanalyse ble valgt da denne metoden egner seg godt for å undersøke den prediktive evnen til en uavhengig variabel på en avhengig variabel, samtidig som man tar høyde for andre kontrollvariabler.

Det ble utført fire regresjonsanalyser, en for hver av de fire

personlighetsdimensjonene. Avhengig variabel var personlighetsdimensjonen, mens både kjønn, alder og testsituasjon var uavhengige variabler. Kjønn og alder ble lagt inn i steg 1, situasjon i steg 2. Kursdeltakere ble satt som referansegruppe. Kjønn er kodet som dummyvariabel (mann = 0, kvinne = 1).

Uavhengig t-test for gjennomføringen av selve testen. Det var interessant å undersøke hvorvidt testsituasjon hadde betydning for selve gjennomføringen av Shapes. Uavhengig t-test ble derfor benyttet med testsituasjon som avhengig variabel og de tre kontrollpunktene i Shapes; konsistens, poengfordeling og testtid som uavhengige variabler, for å undersøke H3 og H4.

Resultat

Interkorrelasjoner for dimensjonene i Shapes

Analysen av sammenhengen mellom dimensjonene i Shapes viser flere høye korrelasjoner, både positive og negative. Den høyeste positive korrelasjonen ble funnet mellom dimensjonene teoretisk og veloverveid ($r = .39$) og for dimensjonene endringsvillig og oppfinnsom ($r = .39$). Den høyeste negative korrelasjonen ble funnet mellom personlighetsdimensjonene samvittighetsfull og styrende ($r = -.30$).

I testdokumentasjonen (Cut-e Norge AS, 2015) beskrives også interkorrelasjonene for personlighetsdimensjonene i Shapes, ut fra norske testdata ($N = 3464$).

Interkorrelasjonene strekker seg til $r = .33$ (Cut-e Norge AS, 2015).

Interkorrelasjonsmatrisen for denne studien viser både flere og høyere interkorrelasjoner mellom personlighetsdimensjonene i Shapes, enn det som er beskrevet i testdokumentasjonen.

Tabell 2

Interkorrelasjonsmatrise for dimensjonene i Shapes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Styrende	1																	
2. Overbevisende	.26**	1																
3. Sosialt sikker	.13**	.14**	1															
4. Utadvendt	-.10*	.02	.36**	1														
5. Omgjengelig	-.15**	-.23**	.07	.17**	1													
6. Observerende	-.10*	-.01	-.05	-.02	.06	1												
7. Veloverveid	-.21**	-.11*	-.16**	-.24**	-.10*	.16**	1											
8. Resultatbevisst	.07	.04	-.19**	-.15**	-.20**	-.16**	-.06	1										
9. Planleggende	-.10*	-.28**	-.22**	-.25**	.07	-.20**	.14**	.11*	1									
10. Samvittighetfull	-.30**	-.26**	-.12*	-.06	.18**	-.12*	.03	-.04	.37**	1								
11. Analytisk	-.28**	-.20**	-.19**	-.25**	-.07	-.02	.36**	.03	.16**	.06	1							
12. Teoretisk	-.09	-.07	-.18**	-.18**	-.25**	.17**	.39**	-.03	-.07	-.19**	.36**	1						
13. Oppfinnsom	-.03	.02	-.10*	-.03	-.24**	.06	.05	-.08	-.25**	-.28**	.01	.21**	1					
14. Endringsvillig	-.03	-.11*	-.11*	-.12*	-.12*	-.07	.03	-.04	-.12**	-.21**	.01	.07	.39**	1				
15. Uavhengig	.01	.03	-.10*	-.15**	-.06	.01	.10*	-.03	-.10*	-.11*	-.01	.02	.12**	.14**	1			
16. Ambisiøs	.12*	.02	-.10*	-.18**	-.24**	-.17**	-.16**	.20**	-.001	-.09	-.03	.05	.05	.06*	-.06	1		
17. Konkurransorientert	.12**	.27**	-.01	.08	-.27**	-.15**	-.21**	.08	-.15**	-.08	-.16**	-.20**	-.03	-.08	-.04	.19**	1	
18. Energisk	.03	-.09*	-.16**	-.03	-.01	-.18**	-.20**	.10*	.04	.07	-.14**	-.16**	.09	.12*	.01	.17**	-.10*	1

* $p < .05$; ** $p < .001$. $N = 450$

Forskjeller i dimensjonsskårer

De numeriske forskjellene i gjennomsnitt på de 18 personlighetsdimensjonene var størst for dimensjonene: resultatbevisst, samvittighetsfull, analytisk, uavhengig og konkurranseorientert. Jobbsøkergruppen hadde i gjennomsnitt høyere skåre på resultatbevisst, samvittighetfull, analytisk og konkurranseorientert, sammenlignet med kursgruppen. Gjennomsnittsskåren på dimensjonen uavhengig var derimot lavere for jobbsøkergruppen. Tabell 3 viser gjennomsnitt, standardavvik og differansen i gjennomsnitt for alle personlighetsdimensjonene i Shapes.

Det var en signifikant forskjell i gjennomsnittsskåre på personlighetsdimensjonen resultatbevisst for høy-risiko ($M = 24.29$, $SD = 3.40$) og lav-risiko gruppen ($M = 22.70$, $SD = 3.77$; $t(448) = -2.38$, $p = .018$). Forskjellen i gjennomsnitt ($M_{diff} = -1.59 \pm 1.31$ (95 % konfidensintervall) var liten til medium (cohen's $d = .43$) (Cohen, 1992).

Høy-risiko ($M = 24.63$, $SD = 4.61$) og lav-risiko gruppen ($M = 22.95$, $SD = 4.78$) viste en signifikant forskjell i gjennomsnittsskåre på personlighetsdimensjonen samvittighetsfull; $t(448) = -1.98$, $p = .049$. Forskjellen i gjennomsnitt ($M_{diff} = -1.68 \pm 1.67$ (95 % konfidensintervall) var liten til medium ($d = .35$) (Cohen, 1992).

Det ble observert en signifikant høyere skåre på personlighetsdimensjonen analytisk i jobbsøker gruppen ($M = 19.50$, $SD = 5.07$), sammenlignet med kursgruppen ($M = 17.39$, $SD = 4.29$; $t(448) = -2.72$, $p = .007$). Forskjellen i gjennomsnitt ($M_{diff} = -2.11 \pm 1.53$ (95 % konfidensintervall) var liten til medium ($d = .49$) (Cohen, 1992).

Jobbsøkergruppen ($M = 17.53$, $SD = 3.22$) viste en signifikant lavere skåre på dimensjonen uavhengig, sammenlignet med kursgruppen ($M = 19.84$, $SD = 3.62$; $t(448) = 3.61$, $p = .001$). Forskjellen i gjennomsnitt ($M_{diff} = 2.32 \pm 1.26$ (95 % konfidensintervall) var betydelig, $d = .64$ (Cohen, 1992).

Sammenlignet med kursgruppen, var skårene fra seleksjonssituasjonen signifikant høyere på resultatbevissthet, samvittighetsfullhet og analytisk, samt signifikant lavere på uavhengig. Effektstørrelser varierte fra .35 til .64.

Tabell 3

Deskriptiv statistikk, M, SD og M_{diff}

Var.	U1		U2		M_{diff}
	(N = 416)		(N = 34)		
	M	SD	M	SD	
Styrende	22.44	6.04	21.43	6.99	1.03
Overbevisende	18.68	5.29	17.70	5.59	0.99
Sosialt sikker	22.16	4.58	21.61	6.14	0.55
Utadvendt	19.75	5.18	18.92	5.98	0.83
Omgjengelig	24.40	4.36	23.76	3.57	0.64
Observerende	21.57	4.83	20.20	5.16	1.37
Veloverveid	20.10	3.76	20.34	3.92	-0.24
Resultatbevisst	22.70	3.77	24.29	3.40	-1.59
Planleggende	21.06	4.38	22.23	3.37	-1.17
Samvittighetsfull	22.95	4.78	24.63	4.61	-1.68
Analytisk	17.39	4.29	19.50	5.07	-2.11
Teoretisk	19.66	4.40	19.47	4.17	0.19
Oppfinnsom	16.54	5.59	16.99	5.87	-0.45
Endringsvillig	19.69	4.23	20.20	3.74	-0.51
Uavhengig	19.84	3.62	17.53	3.22	2.32
Ambisiøs	22.33	4.42	22.40	3.23	-0.07
Konkurransorientert	14.82	6.68	16.77	5.80	-1.95
Energisk	24.51	4.57	25.74	4.55	-1.24

M_{diff} = Differansen i gjennomsnittsskårer mellom U1 og U2

Hierarkisk multiple regresjonsanalyse

Hierarkisk multiple regresjonsanalyse ble gjennomført for å undersøkt hvor mye av variansen i de fire personlighetsdimensjonene, resultatbevisst, samvittighetsfull, analytisk og uavhengig, som kan forklares av ulike testsituasjoner, når alder og kjønn er kontrollert for, se tabell 4. Forberedende analyser ble gjennomført for å forsikre at ingen antakelser var brutt (normalitet, lineærhet, multikolaritet og homogen varians). Kjønn og alder ble lagt inn i steg 1, mens situasjon ble lagt inn i steg 2.

Resultatbevisst. Kjønn og alder forklarte 0.5 % av variansen i personlighetsdimensjonen resultatbevisst. Etter at situasjon ble lagt til, ble totalt 1.7 % av variansen forklart av modellen, $F(3, 446) = 2.57, p < .054$. Situasjonsvariabelen forklarer 1.2 % av variansen i den avhengige variabelen, når kjønn og alder er kontrollert for, (R squared change = .012, F-change (1, 446) = 5.41, $p < .02$). I den siste modellen var bare situasjon statistisk signifikant ($\beta = .11, p < .02$).

Samvittighetsfull. Kjønn og alder forklarte 5.9 % av variansen i personlighetsdimensjonen samvittighetsfull. Etter at situasjon ble lagt til, ble totalt 6.9 % av variansen forklart av modellen, $F(3, 446) = 11.01, p < .001$. Situasjonsvariabelen forklarer 1 % av variansen i den avhengige variabelen, når kjønn og alder er kontrollert for, (R squared change = .010, F-change (1, 446) = 4.92, $p < .027$). I den siste modellen var alle variablene statistisk signifikant, hvor kjønn hadde høyest beta verdi ($\beta = .22, p < .001$), deretter alder ($\beta = .12, p < .013$) og situasjon ($\beta = .10, p < .027$).

Analytisk. Kjønn og alder forklarte 0.8 % av variansen i personlighetsdimensjonen analytisk. Etter at situasjon ble lagt inn totalt 2 % av variansen forklart av modellen, $F(3, 446) = 3.06, p < .028$. Situasjonsvariabelen forklarer 1.2 % av variansen i den avhengige variabelen, når kjønn og alder er kontrollert for, (R squared change = .012, F-change (1, 446) = 5.63, $p < .018$). I den siste modellen ble situasjon funnet som eneste statistisk signifikante variabel ($\beta = .11, p < .018$).

Uavhengig. Kjønn og alder forklarte 2.5 % av variansen i personlighetsdimensjonen uavhengig. Etter at situasjon ble lagt inn ble totalt 5 % av variansen forklart av modellen, $F(3, 446) = 7.77, p < .001$. Situasjonsvariabelen forklarer

2,5 % av variansen i den avhengige variabelen, når kjønn og alder er kontrollert for, (R squared change = .025, F-change (1, 446) = 11.59, $p < .001$). I den siste modellen var både alder og situasjon statistisk signifikant, hvor situasjon hadde høyest beta verdi ($\beta = -.16, p < .001$), deretter alder ($\beta = -.12, p < .013$).

Forskjeller i selve gjennomføringen av testen

Det ble også gjennomført uavhengig t-test, med situasjon som avhengig variabel og test-tid, konsistens og poengfordeling som uavhengige variabler, for å undersøke H4. Analysene viser statistisk signifikant forskjell i testtid, mens de to andre variablene ble vist ikke signifikant. Antakelsen om homogen varians ble testet ved Levene's test, og viste at variansen i de to utvalgene var ulik, $F(448) = 9.08, p = .003$, og derfor rapporteres justert verdi for t . Det var en signifikant forskjell i testtid for høy-risiko ($M = 21.76, SD = 9.47$) som brukte betydelig lengre tid, sammenlignet med lav-risiko gruppen ($M = 16.03, SD = 6.26; t(35) = -3.47, p = .001$). Forskjellene i gjennomsnitt ($M_{diff} = -5.74 \pm 3.35$ (95 % konfidensintervall) var betydelig, $d = 0.88$ (Cohen, 1992).

Tabell 4

Regresjonsanalyser med resultatbevisst, samvittighetsfull, analytisk og uavhengig som avhengige variabler. Situasjon, alder og kjønn som uavhengige variabler.

Variabler	Totale Utvalg															
	Resultatbevisst				Samvittighetsfull				Analytisk				Uavhengig			
	β	t	R^2	ΔR^2	β	t	R^2	ΔR^2	β	t	R^2	ΔR^2	β	t	R^2	ΔR^2
<i>Step 1</i>			.005				.059				.008				.025	
Kjønn	-.05	-1,06			.22**	4.64			-.02	-.49				-.09	-1.85	
Alder	-.04	-.01			.12*	2.51			.06	.127				-.12*	-2.49	
<i>Step 2</i>			.017	.012			.063	.010			.020	.012			.050	.025
Situasjon	11*	2.33			.10*	2.22			.11*	2.37				-.16**	-3.40	

* $p < .05$; ** $p < .001$.

Diskusjon

Formålet med denne studien var å undersøke hvilken betydning testsituasjon hadde for skårer på personlighetstesten Shapes. Datamateriale fra en høy-risiko og en lav-risiko situasjon gjorde det mulig å undersøke forskjeller i gjennomsnittsskårer hos individer som gjennomførte testen i en reell seleksjonssituasjon, sammenlignet med kurssetting. Studien hadde også et ønske om å bidra til kunnskap om faking og personlighetstester. Samt å undersøke betydning av det semi-ipsative formatet til Shapes.

Jeg vil nå presentere en oppsummering av hovedfunnene fra denne studien, før studiens hypoteser vil bli drøftet i lys av tidligere forskning. Metodiske begrensninger vil bli vurdert før oppgaven avsluttes med implikasjoner for videre forskning og en konklusjon.

Hovedfunn

Resultatene viser at sammenhengen mellom personlighetsvariablene i Shapes i stor grad er i overensstemmelse med det som er forventet gitt det semi-ipsative formatet; flere svake negative korrelasjoner ble observert mellom dimensjonene i Shapes. Dette støtter H1.

Gruppen som har gjennomført Shapes i en seleksjonssetting har en signifikant høyere skåre på personlighetsdimensjonen resultatbevisst, samvittighetsfull og analytisk, samt en signifikant lavere skåre på dimensjonen uavhengig, sammenlignet med individer som har tatt testen i en kurssituasjon. Alle forskjellene kan sies å være betydelige basert på effektstørrelsene som varierte fra .35 til .64. Funnene støtter dermed H2, at noen av dimensjonsskårene i høy-risiko og lav-risikosituasjon ville være ulike.

De fire personlighetsdimensjonene som viste en signifikant forskjell ble undersøkt ytterligere for å se hvor mye av variansen i de ulike personlighetsdimensjonene som kunne bli forklart av testsituasjonen. Resultatene viste at testsituasjon var en signifikant prediktor for alle fire variabler. Resultatene fra denne analysen støtter også H2. Den største forklarte variansen ble funnet for variabelen uavhengig, hvor

testsituasjon alene forklarte 1.2 % av variansen til personlighetsdimensjonen.

Det ble også funnet en signifikant forskjell i hvor lang tid testtakerne brukte på å gjennomføre testen, som indikerte at jobbsøkere tar seg lengre tid på gjennomføringen. Jobbsøkere brukte i gjennomsnitt nesten 6 minutter lengre tid på gjennomføringen, sammenlignet med kursdeltakere. Resultatene av den uavhengige t-testen viser støtte for H3, mens avviser hypotesen om at svarene ville være mer konsistente i seleksjonssettingen (H4).

Testsituasjon og personlighetsskårer

En av hypotesene for studien var at jobbsøkere ville skåre høyere på noen av personlighetsdimensjonene i Shapes, sammenlignet med kursdeltakere. I en høy-risiko situasjon kan utfallet av testen ha viktige konsekvenser for individet. Når personlighetstester blir brukt med det formål å ta en beslutning som vil påvirke hvorvidt testtaker får tilbud om jobb eller ikke, kan dette gi testtaker motivasjon til å produsere skårer som vil øke individets sjanse til å få den ønskede jobben (Dilchert et al., 2006). Hypotesen bygger på antakelsen om at individer i en høy-risiko situasjon vil ha motivasjon til å fordreie skårene sine på personlighetstesten, og dermed presentere seg selv på ønsket måte. Hva individene antar at er forventet eller ønsket å skåre kan derfor ha noe å si for hvordan de velger å skåre. Det kan derfor tenkes at individene vil vise en slags profil som er fordelaktig i den gitte bransjen, eller for stillingen de søker. Det ble ikke gjort noen antakelse om hvilke dimensjoner jobbsøkerne ville skåre høyere på da alle dimensjonene i Shapes er arbeidsrelaterte og kan derfor ses som viktige for en jobb.

Resultatene fra studien står i samsvar med funn fra tilsvarende studier, som viser at jobbsøkere har høyere gjennomsnittsskårer på flere personlighetsdimensjoner, sammenlignet med andre grupper (Birkeland et al., 2006; Rosse et al., 1998; Robie et al., 2001; Donovan et al., 2003; D. Smith & Robie, 2004). Rosse et al. (1998) fant i sin studie at jobbsøkere i snitt skåret 0.69 standardavvik høyere enn individer som tok testen i en lav-risikosituasjon. Dette nivået tilsvarer hva som er funnet i laboratoriestudier (D. Smith & Robie, 2004; Robie et al., 2001). Funnene i denne

studien viser at jobbsøkere skårer mellom 0.35 og 0.49 standardavvik høyere enn kursdeltakerne på dimensjonene; resultatbevisst, samvittighetsfull og analytisk, mens de på dimensjonen uavhengig skåret 0.64 standardavvik lavere enn kursdeltakere. Disse funnene er som forventet lavere enn hva som er funnet i laboratoriestudier, men samsvarer med funn fra studien til Robie et al. (2001) hvor det ble rapportert rundt et halvt standardavvik forskjell i gjennomsnitt.

Birkeland et al. (2006) fant i sin meta-analyse av faking i reelle situasjoner at jobbsøkere skårer høyere på trekk som kan anses som særlig jobbrelatert. Jobbsøkere skåret signifikant høyere enn ikke-jobbsøkere på trekkene ekstrovert, emosjonelt stabil, samvittighetsfull og åpenhet. Den største forskjellen ble funnet for samvittighetsfull og emosjonelt stabil som var nesten et halvt standardavvik høyere for jobbsøkergruppene (Birkeland et al., 2006). Resultatene fra denne studien viste at jobbsøkere i snitt skårer høyere på dimensjonene resultatbevisst, samvittighetsfull og analytisk. Det er interessant at det ikke ble funnet en forskjell i dimensjonen sosialt sikker, da denne dimensjonen kan antas å være assosiert med big-5 trekket ekstrovert, som ofte blir sett på som svært attraktivt for flere typer jobber. Dette kan indikere at individene i høy-risikogruppen utgjør en spesiell gruppe. Det kan tenkes at på bakgrunn av at de søker jobb hos Gjensidige forsikring AS, at profilene derfor kan knyttes til en spesiell type jobb, for eksempel økonomer.

Hvordan påvirker testsituasjon gjennomføringen av testen

Det ble også funnet en signifikant forskjell i testtid mellom individer som tok testen i seleksjonssetting og i kurssammenheng. Høy-risikogruppen brukte betydelig lengre tid på å gjennomføre testen. Dette kan tyde på at de tar situasjonen seriøst og at de ønsker å prestere best mulig. Dette kan tolkes som et tegn på at testtakerne føler seg motivert, og at de kanskje derfor vil velge å fordreie sine skårer for å øke sjansen for å få jobben.

Styrker og svakheter ved studien

Studien bidrar til ny kunnskap på feltet personlighetstester i personelleleksjon, da studien undersøker testsituasjonens betydning for skårene på personlighetstesten Shapes, ved bruk av norske data. Gjennom å undersøke Shapes kan studien også bidra til bedre forståelse av personlighetstester som benytter et semi-ipsativt format.

Forskningsdesign. En stor styrke med denne studien er at den benytter data fra en reell seleksjonssituasjon. Dette gir muligheten til å undersøke både evnen og motivasjonen til å fordreie skårene på testen.

Studier som har undersøkt i hvilken grad individer kan fake sine responser på personlighetstester har benyttet et forskningsdesign med enten avhengig- eller uavhengig utvalg (Furnham, 1986). Det er i denne studien benyttet sistnevnte, hvor individer fra to ulike og uavhengige grupper er benyttet. Det er både styrker og begrensninger ved bruk av denne typen forskningsdesign. En styrke er at individene bare gjennomfører testen en gang, og at individet derfor ikke blir påvirket av å ha gjennomført samme test flere ganger. En svakhet ved dette forskningsdesignet er at det ikke tillater å evaluere forandring i skårer over ulike situasjoner da det ikke er benyttet re-test.

Fordi individene i studien ikke ble instruert kan man ikke konkludere med hva som er årsakt til forskjellene i gjennomsnitt. Det kan for eksempel tenkes at den observerte forskjellen skyldes faktiske forskjeller i personlighet mellom individene i de to gruppene. Derfor kan det ikke utelukkes at kursdeltakerne kan ha hatt et ønske å fremme seg selv i et godt lys, selv om ikke resultatene får viktige konsekvenser. Det motsatte er også en mulighet, at kursdeltakere ikke er like motivert som jobbsøkere, og derfor ikke er så nøye med hva de svarer. Dette kan resultere i målefeil, og i verste fall gjøre at skårene fra kursgruppen ikke er reliable. Funnene fra denne studien må derfor tolkes med varsomhet.

Det må også påpekes at denne studien, som mange andre, har undersøkt faking blant jobbsøkere på gruppenivå, ved å undersøke forskjeller mellom en gruppe jobbsøkere og en gruppe ikke-jobbsøkere. Det kan være noe misvisende å anta at faking er konstant for alle individene i en seleksjonssituasjon, spesielt med tanke på bevisene

som støtter individuelle forskjeller både i grad av og hvordan individer faker (Barrick & Mount, 1996; J. Hogan et al., 2007; Morgeson et al., 2007; Rosse et al., 1998; D. Smith & Robie, 2004; Viswesvaran & Ones, 1999). Det argumenteres for at faking er et problem som ville vært bedre besvart på individnivå (Birkeland et al., 2006), da det antas at flere faktorer påvirker et individs valg om å forvrengne sine responser (McFarland & Ryan, 2000). Det kan derfor tenkes at funnene fra denne studien representerer en tendens hos bare noen få individer. Det er også en mulighet for at forskjellene ikke skyldes faking i det hele tatt. Det kan tenkes at personlighetsdimensjonene som jobbsøkergruppen skårer forskjellig på sammenlignet med kursdeltakerne er typisk for bransjen. Forskjellene kan dermed være et resultat av at personer med en spesiell personlighetsprofil søker nettopp denne typen jobb, forskjellene skyldes dermed faktiske personlighetsforskjeller i de to gruppene.

Studiens utvalg. Til tross for at utvalget var stort nok til de analysene som ble benyttet, kan det tenkes at et større utvalg og lik størrelse i de to gruppen kunne ført til flere signifikante forskjeller. Det er anbefalt å ha minst 15 % av utvalget i hver variabel, når man kjører regresjonsanalyser med dummyvariabler. Gruppen med jobbsøkere utgjorde bare 7 % av det totale utvalget og kan dermed ha påvirket de statistiske analysene. En annen begrensning i studien er at utvalget i høy-risikogruppen er relativt lite. Små utvalg kan være en trussel mot den statistiske styrken av funnene (Cohen, 1992). Utvalget fra høy-risikogruppen hadde også relativt lav responsrate (21 %) (Baruch, 1999). Lav responsrate kan være problematisk da de som ikke svarer kan være forskjellig fra de som svarer i forhold til en rekke faktorer (Baruch, 1999). Resultatene fra studien må derfor tolkes med varsomhet. Valget om å fjerne uteliggere fra en gruppe (U1), men ikke den andre (U2) kan også være en feilkilde. Men på bakgrunn av studiens overordnede problemstilling var det hensiktsmessig å gjøre det på denne måten.

Statistiske analyser. Det er flere svakheter som kan trekkes frem ved de statistiske analysene som er benyttet i studien. Å benytte flere t-tester på samme datamateriale øker sjansen for å begå en type-1 feil, at vi feilaktig forkaster H_0 , og konkluderer at det er en genuin forskjell når det ikke er det (Field, 2009). Det finnes

flere måter å korrigere for dette, den vanligste er Bonferroni-korrigerings, hvor p-verdien .05 deles på antall tester (Field, 2009). Dette ville i denne studien blitt altfor strengt. Det ble derfor valgt å legge større vekt på den faktiske forskjellen i snitt mellom de to gruppene. Signifikans ble brukt som et kriterie for å ta dimensjonene med videre til regresjonsanalysen, samtidig som det ble valgt å regne ut Cohen's d, for å si noe om størrelsen på forskjellene.

I tolkningen av resultatene er det blitt lagt større vekt på de faktiske forskjellene i gjennomsnitt, da det er flere svakheter ved studien og ulike egenskaper ved formatet som kan ha påvirket resultatene. Det må også påpekes at siden individene ikke er testet i ulike settinger, kan man ikke fastslå at det er situasjonene som fører til disse forskjellene. Vi kan dermed ikke fastslå et kausalt forhold, kun assosiasjoner, som kan ha mange forklaringer.

Hvordan kan det semi-ipsative formatet til Shapes ha påvirket resultatene?. Data er ipsative når summen av alle innhentede skårer for ethvert individ på den gitte testen er konstant (Clemans, 1966). I følge (Hicks, 1970) er en del tester blitt konstruert slik at de lager delvis ipsative skårer. Shapes kan klassifiseres som en slik test, da adalloc åpner for variasjon i totalskåren, samtidig som den gjør at individet ikke kan skåre høyt på alle dimensjoner, høyere skåre på en dimensjon vil nødvendigvis føre til lavere skåre på en annen dimensjon. Dette kan ses som et resultat av at skalaene i Shapes er vektet forskjellig. Dette står i overensstemmelse med Hicks (1970); Clemans (1966) som har beskrevet flere egenskaper et instrument som vil gjøre det til delvis ipsativt, inkludert at skårede alternativer er forskjellig vektet. Adalloc skaper større differensiering mellom skårene på personlighetsdimensjonene for et individ ved at testledd som skåres høyt blir parett sammen med andre testledd med høy skåre. Adalloc vektet disse med en høyere skåre enn dem som blir satt sammen i en blokk med testledd som har fått lav skåre. Videre vil de testleddene i blokken med høy skåre som igjen blir skåret høyt av individet, bli tildelt en høyere vektning enn dem som blir skåret lavt i samme blokk.

Shapes er som tidligere nevnt et semi-ipsativt mål, og har dermed både normative

og ipsative egenskaper. En årsak til at Shapes ikke klassifiseres til fullstendig ipsativt mål er at den har noen normative karakteristikk. Adaloc skaper ulik totalskåre for individer, og individer er gitt friheten til å velge hvordan de ønsker å distribuere poengene.

Da Shapes også innehar ipsative egenskaper, kan det tenkes at disse har påvirket resultatene fra studien.

Negative korrelasjoner mellom skalaer. På bakgrunn av den konstante sluttskåren har ipsative data en rekke uønskede statistiske egenskaper som man ikke finner hos typiske normative data (Clemans, 1966; Hicks, 1970). Flere har konkludert med at ipsative data ikke egner seg til vanlige statistiske analysemetoder på bakgrunn av at summen av variablene er konstant for alle individer (Chan, 2003; Closs, 1996; Cornwell & Dunlap, 1994; Johnson et al., 1988). Den konstante sluttsummen og at skalaskårene i ipsative mål har null varians resulterer i at summen av alle kolumnene eller radene i interkorrelasjonsmatrisen også er null (Clemans, 1966).

Interkorrelasjonsmatrisen for dimensjonene i Shapes viser i denne studien et stort antall lave og negative korrelasjoner. Dette indikerer at det semi-ipsative formatet til Shapes kan ha innvirkning på analysene som igjen vil påvirke resultatet til studien. Flere har argumentert at de negativt fordreide verdien gjør tolkningen av korrelasjoner mellom skalaene problematiske og fraråder derfor bruk av standard statistiske analyser på ipsative data (Meade, 2004; Closs, 1996; Tenopyr, 1988).

Variansen i ipsative skalaer. Et annet problem ved ipsative skalaer er at når summen av alle skalaskåre er konstant vil variansen også være lik null (Clemans, 1966). At ipsative mål har mindre variasjon er noe som kan ha påvirket resultatene av analysen. Liten spredning gjør korrelasjonene mindre og kan dermed ha påvirket effektstyrkene i positiv retning, men regresjonsanalysen i negativ retning.

Forholdet mellom ipsative skalaer og andre variabler. Det er også hevdet at kriterievaliditeten til personlighetstester er påvirket av egenskaper ved ipsative mål (Brown & Maydeu-Olivares, 2012a). Den konstante sluttskåren på ipsative skalaer fører til at skalaenes kovarians med ethvert eksternt kriteriemål også vil være lik

null (Hicks, 1970; Clemans, 1966). Enhver positiv korvarians med eksterne variabler blir kompensert for ved noen negative kovarianser og omvendt (Johnson et al., 1988).

Dersom en av testens dimensjoner er antatt å samvarierte positivt med et kriteriemål, vil de ipsative egenskapene skape falske kompenserende kovarianser og dermed forvrengte forholdet. Det er derfor tenkelig at de ipsative egenskapene ved Shapes kan ha virket inn på forholdet mellom personlighetsdimensjonen og andre variabler. På bakgrunn av dette må man ta høyde for at resultatene fra regresjonsanalysen kan skyldes denne feilkilden, og at man ved bruk av en normativ versjon av Shapes kanskje ville funnet andre resultater.

Hicks (1970) har argumentert for at graden av ipsativitet kan ha alvorlige konsekvenser for tolkningen av skårer og beslutninger tatt i personellseleksjon. Da det er vanskelig å fastslå akkurat i hvor stor grad av ipsativitet det er i Shapes, må man derfor ta forbehold om at de overnevnte egenskapene også kan ha påvirket resultatene i denne studien og at de derfor må ses som indikasjoner, heller enn konkrete funn.

Implikasjoner og videre forskning

Funnene i denne studien viser at jobbsøkere har en større gjennomsnittsskåre enn ikke-jobbsøkere, på flere av personlighetsdimensjonene i Shapes. På bakgrunn av resultatene i denne studien foreslås det at mer forskning trengs for å kunne si noe om hvorvidt forskjellene mellom jobbsøkere og ikke-jobbsøkere kan skyldes faking. I denne studien er det benyttet uavhengig-utvalg. Fremtidig forskning vil kunne dra nytte av å kombinere uavhengig utvalg og avhengig utvalg, for å bedre kunne utelukke faktiske forskjeller mellom ulike grupper.

Det ville også være interessant å undersøke forskjellene i skårer for en normativ versjon av Shapes og den originale semi-ipsative versjonen av Shapes, for å si noe om hvilken betydning testens format har for skårene og gjennomføringen av testen.

Forskning som har undersøkt effekten av faking på kriterievaliditeten til personlighetstester har kommet til en generell enighet om at faking ikke påvirker kriterievaliditeten, men mer forskning trengs her. På bakgrunn av funnene fra denne

studien kan man ikke utelukke at individer kan forvrengte sine svar for å fremstille seg selv i lys av stillingen. Hvorvidt dette har konsekvenser for seleksjonsbeslutningene som skal tas har forskning enda ikke blitt enig om, men det tyder på at faking kan påvirke listen over hvilke individer som får tilbud om jobb eller ikke. Undersøkelser av faking i reell personellseleksjon er derfor høyst relevant for å skape bedre forståelse av i hvilken grad slik atferd kan påvirke utfallet av seleksjonsprosessen (Hough, 1998; Rosse et al., 1998). Funnene fra studien kan tyde på at personlighetstesten Shapes kan kartlegge egenskaper som er ansett som spesielt relevante innenfor en gitt bransje. Det argumenteres derfor for at personlighetstester kan bidra til nyttig informasjon i en seleksjonssetting og at bruken av personlighetstester som supplement i personellseleksjon er anbefalt.

Konklusjon

Formålet med studien var å undersøke hvilken betydning testsituasjon har for skårene på personlighetstesten Shapes. Denne studien bidrar til den eksisterende litteraturen ved å undersøke forskjellene i gjennomsnittsskårer, mellom en høy-risiko og en lav-risikogruppe, på en av de mest brukte personlighetstestene på det norske arbeidsmarkedet. Resultatene viste en betydelig signifikant forskjell for fire av 18 personlighetsdimensjoner. Jobbsøkergruppen skåret signifikant høyere på dimensjonene, resultatbevisst, samvittighetsfull og analytisk, sammenlignet med kursgruppen. Skårene på dimensjonen uavhengig var derimot signifikant lavere for jobbsøkergruppen sammenlignet med kursgruppen. Resultatene indikerer at testsituasjon har en betydning for skårene på personlighetstesten. Denne konklusjonen er basert på den betydelige effektstørrelsen, som varierte fra .35 til .64. Det er imidlertid viktig å påpeke at man ikke kan trekkes noen slutning om årsak til forskjellen i personlighetsskårene. Mer forskning er nødvendig for å kunne svare på dette spørsmålet. Studien har imidlertid bidratt med nyttig kunnskap på feltet personellseleksjon og personlighetstester.

Referanser

- Anderson, N. & Ones, D.S. (2003). The construct validity of three entry level personality inventories used in the uk: cautionary findings from a multiple-inventory investigation. *European Journal of Personality*, 17(S1), S39–S66.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69(1), 49–56.
- Barrick, M.R. (2005). Yes, personality matters: Moving on to more important matters. *Human performance*, 18(4), 359–372.
- Barrick, M.R. & Mount, M.K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1), 1–26.
- Barrick, M.R. & Mount, M.K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of applied psychology*, 81(3), 261.
- Barrick, M.R., Mount, M.K. & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and assessment*, 9(1-2), 9–30.
- Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, 15(3), 263–272.
- Baruch, Y. (1999). Response rate in academic studies—a comparative analysis. *Human relations*, 52(4), 421–438.
- Birkeland, S.A., Manson, T.M., Kisamore, J.L., Brannick, M.T. & Smith, M.A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4), 317–335.
- Brown, A. & Bartram, D. (1999). Development and psychometric properties of OPQ32r: Supplement to the OPQ32 technical manual. *SHL Group Ltd., Thames Ditton, UK*.
- Brown, A. & Bartram, D. (2009). Doing less but getting more: Improving forced-choice measures with IRT.

- Brown, A. & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, *71*(3), 460–502.
- Brown, A. & Maydeu-Olivares, A. (2012a). Fitting a thurstonian IRT model to forced-choice data using mplus. *Behavior research methods*, *44*(4), 1135–1147.
- Brown, A. & Maydeu-Olivares, A. (2012b). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological methods*, *18*(1), 36. doi: 10.1037/a0030641
- Cattell, R.B. (1944). Psychological measurement: normative, ipsative, interactive. *Psychological Review*, *51*(5), 292.
- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, *30*(1), 99–121.
- Cheung, M.W. & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling*, *9*(1), 55–77.
- Christiansen, N.D., Burns, G.N. & Montgomery, G.E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*(3), 267–307.
- Clemans, W.V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monograph*(14).
- Closs, S. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, *69*(1), 41–47.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.
- Converse, P.D., Oswald, F.L., Imus, A., Hedricks, C., Roy, R. & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, *16*(2), 155–169.
- Converse, P.D., Peterson, M.H. & Griffith, R.L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment*, *17*(1), 47–60.

- Cornwell, J.M. & Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology*, 67(2), 89–100.
- Cut-e GmbH. (2008). *adalloc - adaptive allocation of consent* (V2.1 utg.). Ikke utgitt, Cut-e GmbH.
- Cut-e Norge AS. (2013). *autorisasjonskurs -modul2: personlighet- og verditest*. Ikke utgitt, Cut-e Norge AS.
- Cut-e Norge AS. (2014). *Brukermanual for shapes-batteriet* (V1.0 utg.). Ikke utgitt, Cut-e Norge AS.
- Cut-e Norge AS. (2015). *Testdokumentasjon - shapes (management)* (V1.3 utg.). Ikke utgitt, Cut-e Norge AS.
- Det Norske Veritas. (2015). *Sertifisering av arbeidspsykologiske tester*. Hentet 12/06/2015 fra <http://tinyurl.com/nkzywbs>
- Dilchert, S., Ones, D.S., Viswesvaran, C. & Deller, J. (2006). Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, 48(3), 209.
- Donovan, J.J., Dwight, S.A. & Hurtz, G.M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, 16(1), 81–106.
- Dwight, S.A. & Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16(1), 1–23.
- Ellingson, J.E., Sackett, P.R. & Hough, L.M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84(2), 155.
- Ellingson, J.E., Smith, D.B. & Sackett, P.R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, 86(1), 122.
- Fan, J., Gao, D., Carroll, S.A., Lopez, F.J., Tian, T.S. & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within

- selection contexts. *Journal of Applied Psychology*, 97(4), 866.
- Field, A. (2009). *Discovering statistics using spss* (3. utg.). Sage publications.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and individual differences*, 7(3), 385–400.
- Goffin, R.D. & Christiansen, N.D. (2003). Information exchange article correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and assessment*, 11(1/2).
- Griffith, R.L. & Peterson, M.H. (2006). *A closer examination of applicant faking behavior*. IAP.
- Griffith, R.L. & Peterson, M.H. (2011). One piece at a time: the puzzle of applicant faking and a call for theory. *Human Performance*, 24(4), 291–301.
- Guilford, J.P. (1952). When not to factor analyze. *Psychological Bulletin*, 49(1), 26.
- Heggestad, E.D., Morrison, M., Reeve, C.L. & McCloy, R.A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, 91(1), 9.
- Hicks, L.E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological bulletin*, 74(3), 167.
- Hogan, J., Barrett, P. & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270.
- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: a socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100.
- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, 18(4), 331–341.
- Holden, R.R., Wood, L.L. & Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology*, 81(1), 160.
- Hough, L.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11(2-3), 209–244.

- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of applied psychology*, 75(5), 581.
- Hough, L.M. & Oswald, F.L. (2000). Personnel selection: Looking toward the future—remembering the past. *Annual review of psychology*, 51(1), 631–664.
- Jackson, D.N., Wroblewski, V.R. & Ashton, M.C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13(4), 371–388.
- Johnson, C.E., Wood, R. & Blinkhorn, S. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, 61(2), 153–162.
- Joubert, T., Inceoglu, I., Bartram, D., Dowdeswell, K. & Lin, Y. (2015). A comparison of the psychometric properties of the forced choice and likert scale versions of a personality instrument. *International Journal of Selection and Assessment*, 23(1), 92–97.
- Khorramdel, L. (2014). The influence of different rating scales on impression management in high stakes assessment. *Psychological Test and Assessment Modeling*, 56(2), 154.
- Komar, S., Brown, D.J., Komar, J.A. & Robie, C. (2008). Faking and the validity of conscientiousness: A monte carlo investigation. *Journal of Applied Psychology*, 93(1), 140.
- Kuncel, N.R. & Borneman, M.J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, 15(2), 220–231.
- Loo, R. (1999). Issues in factor-analyzing ipsative measures: The learning style inventory (LSI-1985) example. *Journal of Business and Psychology*, 14(1), 149–154.
- MacCann, C., Ziegler, M. & Roberts, R. (2011). Faking in personality assessments: reflections and recommendations. I M. Ziegler, C. MacCann & R. Roberts (red.), *New perspectives on faking in personality assessment* (s. 309-329). Oxford

University Press, USA.

- Martin, B.A., Bowen, C.-C. & Hunt, S.T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, *32*(2), 247–256.
- McFarland, L.A. & Ryan, A.M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*(5), 812.
- McLean, J.E. & Chissom, B.S. (1986). Multivariate analysis of ipsative data: Problems and solutions.
- Meade, A.W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, *77*(4), 531–551.
- Meglino, B.M. & Ravlin, E.C. (1998). Individual values in organizations: Concepts, controversies, and research. *Journal of management*, *24*(3), 351–389.
- Mersman, J.L. & Shultz, K.S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences*, *24*(2), 217–227.
- Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K. & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*(3), 683–729.
- Mueller-Hanson, R., Heggstad, E.D. & Thornton III, G.C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, *88*(2), 348.
- Ones, D.S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, *18*(4), 389–404.
- Ones, D.S., Dilchert, S., Viswesvaran, C. & Judge, T.A. (2007). In support of personality assessment in organizational settings. *Personnel psychology*, *60*(4), 995–1027.
- Ones, D.S. & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human performance*, *11*(2-3), 245–269.

- Ones, D.S., Viswesvaran, C. & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660.
- Ones, D.S., Viswesvaran, C. & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of applied psychology, 78*(4), 679.
- Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology, 46*(3), 598.
- Robie, C., Zickar, M.J. & Schmit, M.J. (2001). Measurement equivalence between applicant and incumbent groups: An irt analysis of personality scales. *Human Performance, 14*(2), 187–207.
- Rosse, J.G., Stecher, M.D., Miller, J.L. & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*(4), 634.
- Rothstein, M.G. & Goffin, R.D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16*(2), 155–180.
- Rust, J. (1999). The validity of the giotto integrity test. *Personality and Individual Differences, 27*(4), 755–768.
- Sackett, P.R. & Lievens, F. (2008). Personnel selection. *Annu. Rev. Psychol., 59*, 419–450.
- Salgado, J.F. (1997). The five factor model of personality and job performance in the european community. *Journal of Applied psychology, 82*(1), 30.
- Salgado, J.F., Anderson, N. & Tauriz, G. (2014). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology, 52*(2), 359–392.
- Sandal, G.M. (2011). Personlighetstester og utvelgelse av ledere. I S. Einarsen & A. Skogstad (red.), *Det gode arbeidsmiljø : krav og utfordringer* (2. utg.,

- s. 391-407). Bergen : Fagbokforl.
- Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64(3), 219–238.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2), 262.
- Schmitt, N. & Oswald, F.L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, 91(3), 613.
- Smith, D. & Robie, C. (2004). The implications of impression management for personality research in organizations. *Personality and organizations*, 111–138.
- Smith, D.B. & Ellingson, J.E. (2002). Substance versus style: a new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87(2), 211.
- Stark, S., Chernyshenko, O.S., Chan, K.-Y., Lee, W.C. & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86(5), 943.
- Tenopyr, M.L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology*, 73(4), 749.
- Tett, R.P. & Christiansen, N.D. (2007). Personality tests at the crossroads: A response to morgeson, campion, dipboye, hollenbeck, murphy, and schmitt (2007). *Personnel Psychology*, 60(4), 967–993.
- Tett, R.P., Jackson, D.N. & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel psychology*, 44(4), 703–742.
- Vasilopoulos, N.L., Cucina, J.M. & McElreath, J.M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90(2), 306.
- Villanova, P., Bernardin, H., Johnson, D.L. & Dahmus, S.A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of

- motion picture theater personnel. *Personnel Psychology*, 47(1), 73–90.
- Viswesvaran, C., Deller, J. & Ones, D.S. (2007). Personality measures in personnel selection: Some new contributions. *International Journal of Selection and Assessment*, 15(3), 354–358.
- Viswesvaran, C. & Ones, D.S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and psychological measurement*, 59(2), 197–210.
- White, L.A., Young, M.C., Hunter, A.E. & Rumsey, M.G. (2008). Lessons learned in transitioning personality measures from research to operational settings. *Industrial and Organizational Psychology*, 1(03), 291–295.
- Ziegler, M., Schmidt-Atzert, L., Buhner, M. & Krumm, S. (2007). Fakability of different measurement methods for achievement motivation: Questionnaire, semi-projective, and objective. *Psychology Science*, 49(4), 291.

Appendiks A

Personlighetsdimensjonene i Shapes

Personlighetsdimensjon	Definisjon
styrende	trives med å ha ledelsen; leder og dirigerer andre; holder teamet fokusert på målene
overbevisende	trives med salg og forhandlinger; opplever det som lett å overtale andre; påvirker andre og får deres støtte
sosialt sikker	Fremstår som trygg og selvsikker; er komfortabel under presentasjoner og i formelle situasjoner
utadvendt	er pratsom, livlig og utadvendt; trives i grupper; underholder andre og har et bredt kontaktnett
omgjengelig	er team-orientert og hensynsfull; liker å arbeide sammen med andre; er støttende og behagelig å være sammen med
observerende	er interessert i andres atferd; observerer og analyserer andres motiver og reaksjonsmønstre
veloverveid	inntar et bredt perspektiv; tar de langsiktige konsekvensene av planer og aktiviteter med i betraktningen
resultatbevisst	er pragmatisk og målorientert i sin tilnærming; fokuserer på å nå resultater; løser problemer raskt og effektivt
planleggende	er systematisk og velorganisert; foretar klare prioriteringer; planlegger langsiktig og lager realistiske tidsplaner
samvittighetsfull	er nøye med å følge prosedyrer; respekterer regler og retningslinjer; sørger for at arbeidsoppgaver blir fullført og er opptatt av kvalitet
analytisk	setter seg raskt inn i komplekse saker og problemstillinger; identifiserer kjernen i problemer; evaluerer informasjon grundig
teoretisk	liker å arbeide med konsepter, modeller og strategier; trives med intellektuelle utfordringer
oppfinnsom	er nyskapende i sin tenkning; utvikler fantasifulle og originale idéer og løsninger
endringsvillig	støtter nye ideer; trives med endringer og nye tilnæringsmåter; er konstant på utkikk etter forbedringsmuligheter
uavhengig	handler på eget initiativ; følger sin egen retning; uttrykker fritt egne meninger og synspunkter
ambisiøs	setter seg krevende mål; trives med vanskelige utfordringer; ønsker å nå langt og å lykkes
konkurransorientert	liker å konkurrere; er fast bestemt på å lykkes; trives med å snakke om egne prestasjoner
energisk	viser drivkraft og utholdenhet; er energisk i sin tilnærming til arbeidsoppgaver; beholder fokus og selvkontroll under press

Figur A1. Definisjoner av personlighetsdimensjonene i Shapes

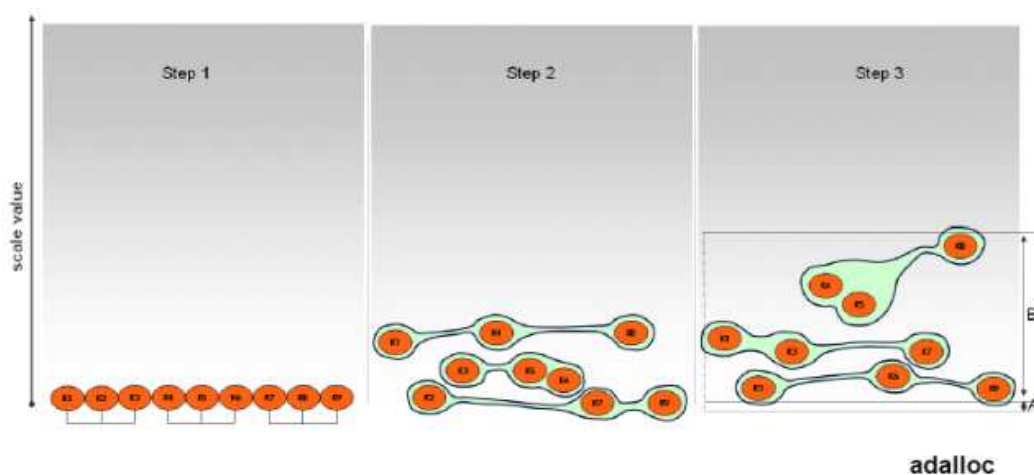
Appendiks B

Adalloc

Adalloc organiserer Shapes i sektorer. Hver sektor består av 6 blokker, da hver av de 18 personlighetsdimensjonene (representert av et testledd) skal presenteres nøyaktig en gang per sektor ($6 \cdot 3 = 18$) (Cut-e Norge AS, 2015). Shapes består av totalt 8 sektorer. Sektor 1 består av 1 testledd fra hver av de 18 personlighetsdimensjonene. I sektor 2 presenteres testledd nummer 2 fra hver dimensjon, i sektor 3 presenteres testledd 3 fra hver dimensjon, og så videre, helt til alle testleddene er presentert. De første seks blokkene av testledd (sektor 1) settes sammen tilfeldig. Når sektor 1 er fullført har hver dimensjon fått tildelt en verdi fra 0 til 6 (Cut-e Norge AS, 2015). For hver sektor blir blokkene vektet ut fra dimensjonsverdiene til korresponderende blokker. En blokk med testledd tilhørende dimensjoner som har en relativt høy verdi blir gitt en høyere blokkverdi enn en blokk med testledd som tilhører dimensjoner som har relativt lave verdier (Cut-e GmbH, 2008).

I sektor 2 blir testleddene fordelt i de ulike blokkene på bakgrunn av deres nåværende verdi (basert på skåringene i sektor 1). Testledd i sektor 2 som representerer hver sine personlighetsdimensjoner blir organisert i blokker på 3, sammen med testledd med tilsvarende verdi (høy, middels, og lav skåre) (Cut-e Norge AS, 2015). Poengene som fordeles i sektor 2 ganges deretter med en faktor som gjenspeiler hvor stor andel av den totale poengsummen som er fordelt på den aktuelle blokkens dimensjoner (Cut-e Norge AS, 2015). Adalloc oppdaterer den enkelte personlighetsdimensjonens verdi etter hver fullførte blokk, dette slik at den totale poengfordelingen fra foregående sektorer blir avspeilt i grupperingen av blokkene i neste sektor. Faktoren som de fordelte poengene ganges med oppdateres tilsvarende (Cut-e Norge AS, 2015). Summen av skalaskårene (råskårene) vil alltid være konstant dersom alle mulige poeng er fordelt (Cut-e GmbH, 2008). Figur B1 gir et bilde av hvordan adalloc fungerer i praksis.

Som beskrevet over ser vi i step 1 at testledd fra hver av de 9 dimensjonene blir presentert i 3 blokker med 3 testledd i hver blokk. I step 2 ser vi at testledd grupperes sammen, basert på verdien de fikk i seksjon 1. Step 2 utgjøres også av tre blokker. I step



Figur B1. Bilde av hvordan adalloc fungerer i praksis

3 blir leddene i hver blokk multiplisert med en faktor som gjenspeiler hvor stor del av den totale poengsummen de representerte dimensjonene i blokken har fått. Slik det fremgår av figuren, gir adalloc høy differensiering mellom de 9 dimensjonene (B i steg 3), samtidig som gjennomsnittet for hver enkelt dimensjon holdes lavt (A i steg 3) (Cut-e Norge AS, 2015).