

Mälardalen högskola
Akademin för hållbar samhälls- och teknikutveckling

En utvärdering av arbetspsykologiska testet Shapes med test-retestmetod

Linn Aasen och Maria Thunberg

C-uppsats i psykologi, HT 2012
Handledare: Jacek Hochwälder
Examinator: Eric Hansen

En utvärdering av arbetspsykologiska testet Shapes med test-retestmetod

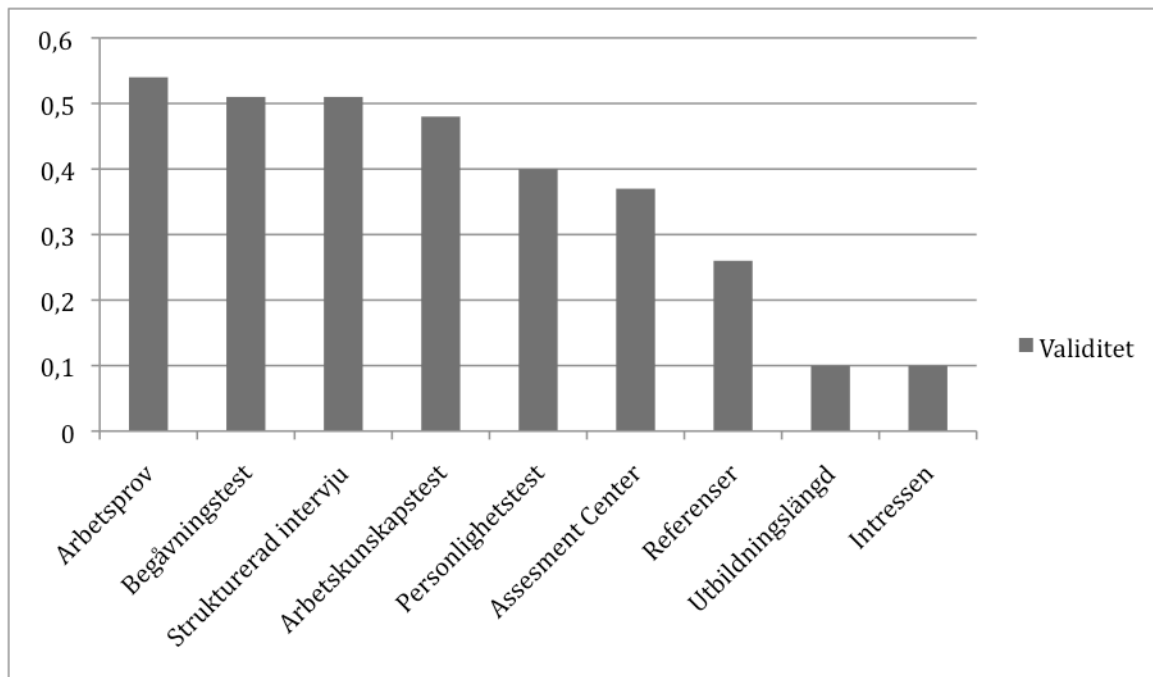
Linn Aasen och Maria Thunberg

Många rekryteringsmetoder som används i dag, som CV och referenstagning, har enligt studier låg validitet, däremot visar många studier att personlighet är stabilt över tid och därför är en mer valid prediktor för att kunna predicera arbetsprestation. Studiens syfte var att undersöka Shapes, ett internetbaserat personlighetstest, och dess arton kompetensdimensioner med test-retest utifrån tre frågeställningar. En datainsamling med 91 deltagare (29 män och 62 kvinnor) gjordes vid två tillfällen med tre veckors mellanrum. Deltagarna delades upp i kön- och åldersgrupper för att se skillnad över tid inom grupperna. Medelvärdesskillnader och korrelationer över tid jämfördes där resultatet visade att det fanns positiva samband mellan båda testtillfällena på samtliga dimensioner. Sju av de arton dimensionerna hade skillnader på medelvärde från testtillfälle ett och testtillfälle två. Detta skulle kunna bero på att Shapes inte är tillräckligt tillförlitligt för att mäta dessa dimensioner eller att dessa dimensioner inte är stabila över tid.

Keywords: test-retest, shapes, personality test, personnel selection

Att välja rätt urvalsmetod i en rekryteringsprocess är oerhört viktigt då det kan vara omfattande med mängder av ansökningar som ska sällas med risk att rätt person försvinner tidigt i processen. För att göra urval används till stor del CV-läsning, ostrukturerade intervjuer och referenstagning för att bedöma en persons framtida arbetsinsats, metoder som statistiskt visat sig ha lägre validitet (Schmidt & Hunter, 1998). Det kostar samtidigt pengar i form av personal som ska läsa och analysera informationen. Referenser som används väljs själv ut av den sökanden och sällan används den nuvarande chefen vilket gör referenserna föråldrade (Mabon, 2002). Självbiografisk information som ett CV är också, som referenser, selekterat av den sökanden, vilket gör att de bästa sidorna visas (Mabon). En annan risk med CV är att många inom rekrytering inte vill se luckor i ett CV vilket leder till att exempelvis tidigare sjukdom eller personliga problem blir en orsak till att en person sällas bort (Mabon).

Vad gäller intervjuer finns det flera fallgropar, exempelvis att olika personer upplever den sökanden och informationen som ges om denne på olika sätt, eller likhetseffekten där intervjuaren föredrar de personer som liknar intervjuaren själv (Skorstad, 2008). Det är skillnad mellan olika typer av intervjuer, en undersökning har exempelvis funnit att det krävs tre till fyra ostrukturerade intervjuer på en person för att ge samma information som en strukturerad intervju (Sackett & Lievens, 2008). Figur 1 visar resultatet i Schmidt och Hunters (1998) metaanalys av validiteten om de olika urvalsmetoderna för att predicera allmän arbetsprestation, där 1.0 är perfekt validitet. De olika mått på arbetsprestation som användes var bland annat gradering av prestation men även försäljningsstatistik och produktionsmängd.



Figur 1 Validitet vid olika urvalsmetoder (Schmidt & Hunter 1998).

Det finns olika typer av arbetspsykologiska test som kan användas för att mäta hur en person kommer att prestera i framtiden: begåvnings- personlighets- intresse- och motivationstest. Personlighetstest tillhör den typ av test som inte ska visa någon form av maxprestation utan ska visa ett medel över hur en person är och förklara hur denne vanligtvis agerar (Mabon 2002).

Personlighetstest kan delas upp i normativa-, ipsativa- och adaptiva test. I ett normativt personlighetstest får testpersonen välja på en skala hur väl ett påstående eller ett adjektiv passar in med hans eller hennes uppfattning om sig själv (Skorstad, 2008). Ytterligare ett kännetecken för normativa test är att de jämförs med en referensgrupp. Då denna typ av test handlar om gradering blir det möjligt för jämförelse mellan testpersoner (Mabon, 2002).

I ett ipsativt personlighetstest får testpersonen begränsade val (forced choice) och testpersonen kan inte gradera sina svar. Ett exempel är att testpersonen måste välja på vilket av två olika påstående som bäst beskriver hur hon eller han är på arbetet. Dessa två påståenden utgör ett så kallat block. En annan variant kan vara att testpersonen enbart får välja på "stämmer" och "stämmer inte" när de ska ta ställning till ett påstående (Skorstad, 2008). En nackdel med ipsativa test är att det inte blir möjligt med jämförelse emellan olika testpersoner då de inte har möjlighet att gradera sina svarsalternativ (Mabon, 2002).

Det senaste på marknaden är adaptiva test vilket är en kombination av normativ och ipsativ testmetodik. Ett exempel på adaptivt test kan vara att testpersonen får ta ställning till ett block innehållande tre olika påstående som exempelvis "Jag planerar långt fram i tiden", "Jag tycker om att ta ledningen" och "Jag stödjer nya idéer och initiativ" (ipsativt - forced choice) men testpersonen ska också disponera ut sex poäng som fördelas mellan de olika påståendena (normativt). Därefter registreras responsen baserat på poängens fördelning och utifrån detta delas de in i nya block som senare viktas olika beroende på tidigare poäng (adaptivt) (Skorstad, 2008). Anledningen till att det kallas adaptivt är för att testet och dess innehåll förändras beroende på vilken respons testpersonen ger.

Fördelar med att använda test istället för de traditionella metoderna vid rekrytering kan vara att det skapar en bättre struktur, det är vetenskapligt, effektivt, objektivt, skapar en möjlighet till jämförelse mellan sökande samt ger samma chans för alla sökande (Skorstad, 2008). Kritik som förs mot personlighetstest är att deltagaren fuskar och förskönar sitt resultat, så kallad faking. Mabon (2002) menade att om en person ger falsk information för att

få ett jobb kommer denne själv att vantrivas då den hamnar i en roll som inte matchar dennes personlighet och blir därför själv lidande. De flesta testleverantörer har dessutom kontrollfrågor och fler studier visar att faking förekommer sällan och har liten betydelse (Mabon, 2002). Robie, Brown och Beaty (2007) undersökte om faking förekommer när personer ska fylla i formulär om sin personlighet i samband med att de söker anställningar, och fann att de *flesta* är ärliga, och att det är svårare att fuska vid ipsativa test. Det går också att förebygga faking genom att det sätts en låg snitt-gräns (gräns på resultat som godkänns) i samband med en sällningsprocess så de som är ärliga inte skjuts undan av de resultat som kommer från de personer som utför faking (Sackett & Lievens, 2008)

Inom testbranschen brukar personlighet beskrivas genom två teorier, antingen typteori där människor grupperas utifrån likheter, eller via traitteorier där de individuella skillnaderna mellan människor används för att hitta särdrag (trait). En mängd personbeskrivande ord samlades ihop för att finna dessa särdrag inom traitteorin och bildade olika kluster baserat på vilka dimensioner som behövdes. Genom utveckling bildades bland annat "Catells 16PF" och Costa och McCrae's "The Big Five" (femfaktormodellen). Många test har idag en viss grund i femfaktormodellen även om de inte direkt mäter de fem faktorerna. Personlighetsfaktorerna är (a) *extroversion* som beskriver hur utåtriktad en person är, (b) *öppenhet* för nya upplevelser, (c) *samvetsgrannhet*, (d) *vänlighet* och (e) *neurotisism* som visar om en person har benägenhet för kraftiga känslomässiga reaktioner (Skorstad, 2008).

I en studie som undersökte personlighet i åldrarna 21-96 år fann att efter åldern 30 skedde inga större förändringar i personlighet trots att en människa normalt upplever omvälvande händelser genom livet (Costa & McCrae, 1988). Denna forskning har kritiserats då andra forskare, exempelvis Robert och DelVecchio (2000) menade att personlighet stabiliseras först vid 50 år, samtidigt som de bekräftar att personligheten är relativt stabilt över tid. I en senare artikel (Costa & McCrae, 2006) bemötte de kritik med att personligheten kan förändras i vuxen ålder men då över lång tid och mestadels innan trettioårsåldern samtidigt som de menade att personligheten är till viss del genetisk och därmed stabil. En annan studie över tid visade på att de faktorer som tillhör Big Five är mer stabila över tid än andra känslomässiga drag, både i kortare och längre mätningar, från några månader till tiotal år (Vaidya, Gray, Haig & Watson, 2002). Dessa studier är baserade på mellan-individ, men även en studie som har mätt traits på inom-individ (Terracciano, McCrae & Costa, 2010) fann att efter en människa har passerat 30 år personlighetsdragen stabila.

De flesta studier som görs över tid fokuserar på ålder men det finns även studier där hänsyn har tagits till kön. Soto, John, Gosling & Potter (2010) har gjort en studie på människor i åldern 10-65 år med över en miljon deltagare som fann att det fanns en skillnad i hur personligheten (utifrån Big Five) förändras olika mellan män och kvinnor under uppväxten gällande neuroticism och öppenhet, då exempelvis kvinnors medelvärde på neurotisism ökar från barndomen upp till tonåren för att sedan minska i vuxen ålder, medan mäns medelvärde på samma faktor minskar successivt från barndomen genom hela livet.

Det finns flera metaanalyser där personlighet utifrån Big Five och arbetsprestation har korrelationer mellan .13 och .28 (Tett, Jackson & Rothstein, 1991) eller .09 och .21 (Schmidt et al., 2008). Skorstad (2008) menade att trots att korrelationerna är relativt låga har de ändå högre prediktionsvaliditet än många av de urvalsverktyg som används idag.

Barrick och Mount (1991) har undersökt validiteten hos personlighetstest men med fokus på olika yrken då de ansåg att undersöka prediktionsvaliditeten *generellt* är onödigt då olika yrken kräver olika personligheter. De undersökte tre yrkeskategorier och fann att endast samvetsgrannhet korrelerade med arbetsprestation på samtliga, annars var det olika personlighetsdrag som gav goda arbetsresultat beroende på vilken yrkeskategori det handlade om.

Traditionellt har personlighetstest utförts med papper och penna, men idag börjar internetbaserade test ta form på marknaden. Sakett och Lievens (2008) beskrev flera fördelar med internetbaserade test: det sparar tid för respondenten och pengar för företaget, samt skapar en viss image av organisationen. Deltagaren får mindre chans till att läsa igenom i förväg och fundera ut hur denne ska svara och då anses risken för faking mindre (Sakett & Lievens). En nackdel skulle kunna vara att testet inte utförs av den personen som avses göra det, men kan undvikas med att testdeltagaren får information om att stickprov kan komma att ske och att deltagaren då kan få göra om testet under kontroll (Skorstad, 2008). En annan nackdel med internetbaserade test menade Skorstad är att alla människor inte har samma förståelse för hur datorer fungerar och testadministratören därför måste se till att kunskap finns innan testet skickas ut.

Då personlighetstest i rekryteringssammanhang syftar till att predicera framtida arbetsprestation är det viktigt att forskare jämför sina tester mot kriterier som definierar olika typer av arbetsprestation, kriterierelaterad validitet (Mabon, 2002). För att höja den kriterierelaterade validiteten visar studier att personlighetstestets items bör kontextualiseras när de används i rekryteringssammanhang (Bing, Whanger, Davison & VanHook 2004, Hunthausen, Truxillo, Bauer & Hammer, 2003). Genom att sätta items i ett specifikt sammanhang ”jag är noggrann med detaljer *på jobbet*” istället för ”jag är noggrann med detaljer” minskar tolkningsproblem (Sackett & Lievens, 2008). Att definiera arbetsprestation är inte enkelt. Det kan betyda olika saker för olika människor i olika situationer och det är ett begrepp som forskare inte har enats om. Forskare använder sig främst av två typer av kriterier för definiera och mäta arbetsprestation. En av de mest förekommande är subjektiva kriterier och innebär exempelvis att en chef bedömer en medarbetares arbetsprestation genom skattning, chefsevaluering (Schmidt & Hunter, 1998). Det finns också objektiva kriterier som exempelvis försäljningsstatistik, lön eller befordringar (Sturman, Cheramie & Cashen, 2005). Detta kan vara ett hinder när jämförelser mellan olika studier görs då olika definitioner och mätning av arbetsprestation har använts.

Enligt Bartram (2005) har forskare tidigare genom olika metoder (begåvningsstest, motivations och personlighetsformulär) försökt att predicera samma sak, nämligen arbetsprestation, men har dock inte relaterat de olika metoderna med ett relevant kriterium som faktiskt omfattar arbetsprestation (Bartram, Robertson & Callinan, 2002). Enligt Bartram är det begreppet *kompetens* som bäst kan förklara arbetsprestation. Bartram definierar kompetens som “sets of behavior that are instrumental in the delivery of desired results or outcomes” (Bartram et al., 2002, s. 7). Genom faktoranalys av olika kompetens- och motivations- och begåvningsstest samt graderingsinstrument har Bartram skapat en generell taxonomi av 112 kompetenser som kan fördelas i åtta kompetensdimensioner (Bartram et al.). I sin meta-analys matchade Bartram kompetensdimensionerna med åtta personlighetsfaktorer: neurotism, extraversion, öppenhet, sympatisk och samvetsgrannhet (Big Five) samt med faktorerna behov av framgång, behov av makt och kontroll samt begåvning. Detta för att få bättre kunskap om vilka personlighetsfaktorer som kan predicera vilka kompetenser. Dessa åtta personlighetsfaktorer gav namnet åt Great Eight. En central aspekt med denna typ av modell är att det blir möjligt för differentiering, att det går att se hur personlighet påverkar vissa aspekter av arbetsprestation. Något som är mycket viktigt i testsammanhang då målet ofta är att jämföra vilken sökande som är bäst lämpad för jobbet (Bartram).

Reliabilitet handlar om ett mätinstruments tillförlitlighet och precision. Vid konstruktionen av ett personlighetstest vill man att testet ska ge samma resultat varje gång man undersöker samma egenskap (Skorstad, 2008). Reliabiliteten kan mätas på olika sätt men de två vanligaste metoderna är att undersöka den interna konsistensen samt test-retest metoden. Intern konsistens har att göra med i vilken utsträckning items mäter samma egenskap (Skorstad). När olika items, som mäter samma egenskap, korrelerar är testet reliabelt.

Cronbachs alpha är det vanligaste måttet för uppskattning av den interna konsistensen och bör vara över 0.70 (Skorstad).

Test-retest syftar till att undersöka stabilitet över tid genom att testa samma person med samma test vid två olika tidpunkter (Mabon, 2002). Om exempelvis resultatet från ett test visar att personen är extrovert ska samma test visa att personen är extrovert vid en senare mätning. Reliabilitetskoefficienten är det som anger korrelationen mellan testresultaten och dess värde är alltid mellan 0 och 1. Perfekt reliabelt testresultat innebär 1 och 0 betyder att all varians beror på felvarians (det som påverkar poängtalerna utan att reflektera den egenskap som mäts) (Skorstad, 2008). Som nämnts ovan har personlighet visat sig vara stabilt över tid vilket gör undersökning av test-retest aktuellt för personlighetstestning. Få test har idag reliabilitetstestats med test-retest-metoden, då Cronbachs alpha har varit den reliabilitet som undersöks mest (Skorstad). En risk med test-retest undersökningar är minneseffekter då dessa kan komma att påverka resultatet (Cohen & Swerdlik, 2005).

Testet Shapes är ett adaptivt, internetbaserat personlighetstest som används som urvalsinstrument vid bland annat rekrytering för att kartlägga en persons kompetensprofil. Shapes är skapat av företaget Cut-e som genomför arbetspsykologisk testning på internet sedan 2002. Cut-e levererar olika typer av test och tjänster till kunder som exempelvis FN, Statoil och Manpower. Shapes bygger på Bartrams (2005) Great Eight personlighetsfaktorer *vänlighet, extraversion, öppenhet, emotionell stabilitet, samvetsgrannhet, behov av makt, behov av framgång och begåvning*. I Shapes har varje sådan faktor ett flertal underfaktorer och i Shapes kallas dessa underfaktorer för *kompetensdimensioner* och totalt finns det 18 stycken. Istället för begreppet neurotism använder Shapes begreppet emotionell stabilitet. Nedan följer en beskrivning av Shapes 18 kompetensdimensioner och deras kopplingar till Great Eight. Vissa kompetensdimensioner har förklaring likt "laddar negativt på emotionell stabilitet" som betyder att om personen hög poäng på den aktuella dimensionen har den låg poäng på emotionell stabilitet (Cut-e, shapes technical manual, 2010).

1. Ledande: tar gärna ledningen; leder och instruerar andra; ser till att teamet fokuserar på målet

Prediceras med personlighetsfaktor: *behov av makt*

2. Övertygande: tycker om att sälja och förhandla; har lätt att övertyga andra; kan påverka andra och vinna deras stöd

Prediceras med personlighetsfaktor: *behov av makt*

3. Socialt självsäker: uppträder trygg och säker i sin roll; känner sig obesvärad vid presentationer och formella tillställningar

Prediceras med personlighetsfaktor: *behov av makt, laddar negativt mot emotionell stabilitet*

4. Sällskaplig: är talför, livfull och har lätt för att umgås; trivs i grupper; kan underhålla andra och ha ett stort kontaktnät

Prediceras med personlighetsfaktor: *extraversion*

5. Hänsynsfull: är teamorienterad och hänsynsfull; arbetar gärna ihop med andra; stödjer andra och är trevlig att umgås med

Prediceras med personlighetsfaktor: *vänlighet, laddar negativt mot emotionell stabilitet*

6. Betraktande: intresserar sig för andras beteenden; observerar och analyserar andras motiv och reaktioner

Prediceras med personlighetsfaktor: *laddar negativt mot behov av framgång*

7. Genomtänkt: har ett brett perspektiv; beaktar de vidare konsekvenserna av planer och aktiviteter

Prediceras med personlighetsfaktor: *samvetsgrannhet*

8. Resultatorienterad: är pragmatisk och målinriktad i sitt arbetssätt; fokuserar på att uppnå resultat; löser problem snabbt och effektivt

Prediceras med personlighetsfaktor: *behov av framgång*

9. Systematisk: respekterar regler och föreskrifter; avslutar påbörjade arbeten och är noga med kvaliteten

Prediceras med personlighetsfaktor: *samvetsgrannhet*

10. Samvetsgrann: är noga med att följa procedurer; respekterar regler och föreskrifter; avslutar påbörjade arbeten och är noga med kvaliteten

Prediceras med personlighetsfaktor: *samvetsgrannhet*

11. Analytisk: förstår snabbt komplexa sakförhållanden och problem; ser ett problems kärna; bedömer information noga

Prediceras med personlighetsfaktor: *begåvning*

12. Konceptuell: arbetar gärna med koncept, modeller och strategier; tycker om intellektuella utmaningar

Prediceras med personlighetsfaktor: *begåvning*

13. Idérik: kan tänka utanför ramarna; utvecklar uppfinningsrika och nyskapande idéer och lösningar

Prediceras med personlighetsfaktor: *öppenhet*

14. Öppen för förändringar: stödjer nya idéer; tycker om förändringar och nyheter; letar ständigt efter möjligheter till förbättringar

Prediceras med personlighetsfaktor: *öppenhet*

15. Självständig: handlar på eget initiativ; följer den egna övertygelsen; uttalar fritt sina uppfattningar och åsikter

Prediceras med personlighetsfaktor: *emotionell stabilitet*

16. Ambitiös: ställer upp krävande mål för egen del; uppskattar svåra utmaningar; vill komma framåt och lyckas

Prediceras med personlighetsfaktor: *behov av framgång*

17. Tävlingsinriktad: tycker om att tävla mot andra; har vinnarinstinkt; talar gärna om sina egna erfarenheter

Prediceras med personlighetsfaktor: *behov av makt, laddar negativt mot vänlighet*

18. Energisk: visar drivkrafter och beslutsamhet; har en energisk inställning till arbetet, förblir fokuserad och behåller kontroller även under press

Prediceras med personlighetsfaktor: *behov av framgång, mot emotionell stabilitet*

Syfte och frågeställning

Med bakgrund mot tidigare forskning som visar att traditionella urvalsmetoder inom rekrytering, som exempelvis ostrukturerade intervjuer och referenstagning, har lägre validitet än personlighetstest är det viktigt att fortsätta forska inom personlighetstestning. Då personlighetstestet Shapes inte har utvärderats med en test-retestmetod tidigare var syftet med studien att undersöka huruvida Shapes visar på stabilitet över tid med en test-retestmetod, samt om det finns en skillnad inom kön och åldersklasser mellan testtillfälle ett och två. Studiens frågeställningar var:

- (1) Finns det signifikanta skillnader mellan testtillfälle ett och testtillfälle två på samtliga dimensioner i personlighetstestet?
- (2) Finns det någon signifikant skillnad inom åldersgrupper vid testtillfälle ett och testtillfälle två?
- (3) Finns det någon signifikant skillnad inom kön vid testtillfälle ett och testtillfälle två?

Metod

Urval och deltagare

Urval gjordes med ett tillgänglighetsurval via Facebook och det fanns inga speciella krav förutom att deltagarna skulle vara över 18 år. Det var 200 personer som tillfrågades varav 143 personer deltog vid första testtillfället, vilket gav en svarsfrekvens på 72 % och vid andra testtillfället deltog 91 personer, vilket gav en svarsfrekvens på 64 %. Detta gav ett bortfall på 58 personer vid test ett och 52 personer vid test två. Utav de 91 deltagarna var 29 män och 62 kvinnor. Ålder för deltagarna var mellan 18 och 69 år ($M = 35.43$ år $SD = 11.96$). Deltagarna delades in i två åldersgrupper utifrån medianvärdet där yngre var deltagarna mellan 18 – 33 år och äldre var deltagarna mellan 34 – 69 år.

Material

För att mäta deltagarnas 18 kompetensdimensioner (nedan kallat för dimensioner) användes det internetbaserade personlighetstestet *Shapes* (Cut-e, 2010). Testet innehöll relevanta frågor om bakgrundsinformation och dessa var kön samt ålder. Vidare fanns det frågor om exempelvis utbildningsnivå eller antal år i arbetslivet men då denna information inte är relevant för studiens frågeställningar har de uteslutits ur materialet. Testet bestod av 48 block. Ett block är tre påståenden som deltagaren ska välja mellan och i testet Shapes innehåller varje block 3 påstående (items). Totalt innehåller Shapes 144 items. Exempel på items är: ”Jag tycker om att sälja”, ”Jag utvecklar innovativa lösningar”, ”Jag arbetar gärna med tydliga prioriteringar”. För varje block ska deltagaren fördela sex poäng mellan de tre items som bäst beskriver honom eller henne, i Figur 2 visas ett exempel på hur ett block kan se ut. Deltagaren har även valet att fördela noll poäng. Höga poäng innebär att items stämmer väl in på deltagaren och låga poäng att det inte gör det. Det är Cut-e’s egna system som genererar

poäng på testet. Poängen jämförs sedan mot en generell svensk normgrupp. Normgruppen är insamlad mellan 2008-2010 och baserad på 7803 personer. Baserat på poängen skapas sedan ett resultat i systemet för varje dimension på en niogradig skala (standard nine) som visar hur deltagarna hamnar i en normalfördelning. Denna skalpoäng är det som sedan används i analyserna.

Shapes är adaptivt vilket innebär att testet anpassar items efter hur deltagarna fördelar sina poäng. Denna teknologi kallas för Adaloc (adaptive allocation of consent). Adaloc strukturerar upp testet in i sektorer. En sektor är antalet block som krävs för att mäta de 18 dimensionerna. Då varje dimension mäts med ett item och varje block innehåller tre items krävs det sex blocks för att mäta de 18 dimensionerna och därmed skapa en sektor. I den första sektorn skapas deltagarens grundpoäng och dimensionerna sorteras sedan efter denna poäng. Från sektor två och framåt bildas sedan nya block utifrån första sektorns poäng vilket resulterar i att items med liknande poäng grupperas. Per sektor viktas varje block, där ett block av items som har höga värden viktas högre än block av items med låga värden. Poängen från fördelade items i sektorerna förändrar stegvis deltagarens värden på de 18 skalorna. Effekten blir att sex poäng som fördelas i ett block med hög viktning har en större påverkan än block med låg viktning. Shapes har tidigare utvärderats med Cronbach's alfa och dimensionerna låg mellan .72 och .85 förutom systematisk som låg på .62 (Cut-e, shapes technical manual, 2010).

Hur väl stämmer dessa beskrivningar överens med ditt beteende?

Du kan dela ut poäng genom att använda + och - knapparna.

Jag tycker det är roligt med konkurrens.	 + -
Jag arbetar gärna mot utmanande mål.	 + -
Jag tycker om att sälja in en idé.	 + -
Du kan fortfarande dela ut 2 poäng:	

Figur 2. Ett exempel på ett block i testet Shapes

Procedur

En av författarna fick förfrågan från företaget Cut-e om att göra ett reliabilitetstest över tid på testet Shapes genom test-retestmetod. Då testet är onlinebaserat ansågs det som bästa alternativ att få kontakt med potentiella deltagare när de redan satt framför datorn. Kontakt togs med 200 personer via Facebook genom författarnas egna personliga kontakter på respektive Facebook-konto. I informationen beskrevs syftet, vad deltagarna förväntades utföra, att deras uppgifter skulle hanteras med konfidentialitet samt att deras deltagande var frivilligt. Informationen beskrev också att deltagarna fick möjlighet att läsa om sin egen kompetensprofil när båda testen var utförda, en rapport som genereras automatiskt i systemet. Det framgick tydligt i informationen att all data skulle aidentifieras efter att testen var avslutade, samt att deltagarna enbart själva hade tillgång till sina kompetensprofiler. Om de

ville delta skulle de ta kontakt och inloggningsuppgifter till testet skickades ut. De som samtyckte och valde att genomföra testet fick i samband med inloggningen testinstruktioner via testsidan samt kontaktuppgifter till en av författarna om frågor skulle uppstå. Efter tre veckor skickades e-post ut till samtliga som deltagit som innehöll nya inloggningsuppgifter för del två och information om hur de kunde läsa rapporten som testet genererat.

Resultat

För att besvara frågeställning ett användes Spearmans rangkorrelation för att se samband mellan de arton dimensionerna vid testtillfälle ett och testtillfälle två. Utfallet av denna analys presenteras i Tabell 1. Samtliga dimensioner har ett positivt samband med sig själva mellan testtillfälle ett och testtillfälle två. Dock varierar dessa samband inom grupperna mellan de olika tillfällena. Till exempel har män .56 på dimensionen självständig och .83 på dimensionen övertygande. Vidare skiljer det sig även mellan grupperna, exempelvis dimensionen analytisk har yngre .59 och äldre .88.

Tabell 1

Samband mellan testtillfälle ett och testtillfälle två inom de olika dimensionerna uppdelat för kön och ålder

Dimension	Kvinnor ($n = 62$) ($t_1 - t_2$) ^a	Män ($n = 29$) ($t_1 - t_2$)	Yngre ($n = 50$) ($t_1 - t_2$)	Äldre ($n = 41$) ($t_1 - t_2$)
Ledande	.84**	.64**	.79**	.79**
Övertygande	.59**	.83**	.58**	.76**
Socialt självsäker	.90**	.74**	.86**	.84**
Sällskaplig	.88**	.82**	.89**	.82**
Hänsynsfull	.63**	.75**	.64**	.73**
Betraktande	.67**	.66**	.68**	.63**
Genomtänkt	.59**	.60**	.64**	.53**
Resultatorienterad	.48**	.55**	.41**	.56**
Systematisk	.72**	.59**	.61**	.84**
Samvetsgrann	.73**	.73**	.78**	.72**
Analytisk	.60**	.85**	.59**	.88**
Konceptuell	.74**	.51**	.52**	.81**
Idérik	.79**	.86**	.76**	.88**
Öppen för förändring	.72**	.57**	.62**	.70**
Självständig	.60**	.56**	.40**	.74**
Ambitiös	.62**	.63**	.62**	.66**
Tävlingsinriktad	.81**	.80**	.91**	.75**
Energisk	.47**	.67**	.43**	.65**

^a Samband mellan dimensionens värden vid testtillfälle ett (t_1) och testtillfälle två (t_2).

** $p < .01$

Vidare för att besvara frågeställning ett utfördes 2 (testtillfälle; t_1 eller t_2) x (kön; man eller kvinna) mixed ANOVA för att se om det fanns skillnader med avseende på dimensionerna. Resultaten visade att det fanns skillnader mellan testtillfälle ett och två på totalt fyra dimensioner och dessa presenteras i Tabell 2. Gränsen för signifikans sattes till $p < .10$, på grund av få deltagare i varje grupp. Även om mellangrupps effekter inte är relevanta för frågeställningarna visar resultatet att det fanns signifikanta mellangrupps effekter med

avseende på kön. På dimensionerna socialt självsäker, analytisk och tävlingsinriktad hade män högre medelvärden än kvinnor vid båda testtillfällena och på dimensionerna samvetsgrann och systematisk hade kvinnor högre medelvärden än män. Ingen signifikant interaktionseffekt uppvisades mellan tid och kön.

Tabell 2

Medelvärde, standardavvikelse och F-värde för huvudeffekt av tid på de dimensioner som har statistisk signifikant skillnad mellan testtillfälle ett och testtillfälle två vid ANOVA tidpunkt x kön

	Testtillfälle ett		Testtillfälle två		<i>F</i> (1, 89)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Ledande	4.47	2.07	4.26	1.97	3.66*
Övertygande	4.08	1.86	4.51	1.96	9.62***
Hänsynsfull	6.03	2.03	5.62	1.90	6.09**
Betraktande	5.23	1.98	5.68	2.08	4.31**

$p < .10^*$, $p < .05^{**}$, $p < .01^{***}$

För att besvara frågeställning två utfördes t-test för beroende mätningar, detta för att se om det fanns någon skillnad inom grupperna kvinnor och män. I Tabell 3 redovisas de signifikanta skillnader som visades i t-testen inom gruppen kvinnor då det inom gruppen män inte fanns några statistiska signifikanta skillnader.

Tabell 3

Medelvärde, standardavvikelse och t-värde på de dimensioner som har statistisk signifikant skillnad mellan testtillfälle ett och testtillfälle två vid t-test inom gruppen kvinnor

	Testtillfälle ett		Testtillfälle två		<i>t</i> (61)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Hänsynsfull	6.00	2.11	5.63	1.95	1.72*
Betraktande	5.35	2.10	5.96	2.08	2.78***
Genomtänkt	4.24	2.08	4.87	2.11	2.59**

$p < .10^*$, $p < .05^{**}$, $p < .01^{***}$

Vidare för att besvara frågeställning ett utfördes 2 (testtillfälle; t1 eller t2) x (åldersgrupp; yngre eller äldre) mixed ANOVA. Totalt var det fem dimensioner som visade på signifikant skillnad mellan testtillfälle ett och testtillfälle två. Fyra av dessa dimensioner (ledande, övertygande, hänsynsfull och betraktande) var samma som förändrades i gruppen kön samt dimensionen genomtänkt har också ändrats mellan de olika testtillfällena. Dimensionerna med signifikanta skillnader presenteras i Tabell 4. Resultatet visar också att två dimensioner har interaktionseffekt. Tidpunkt interagerade med åldersgrupp på dimensionerna ambitiös, $F(1, 89) = 7.12$, $p < .10$ och övertygande $F(1, 89) = 7.78$, $p < .10$. I likhet med kön visar även åldersgrupper signifikanta mellangruppseffekter. Vid båda testtillfällena hade yngre högre medelvärden på dimensionen ambitiös och lägre medelvärde på dimensionen självständig än äldre.

Tabell 4

Medelvärde, standardavvikelse och F -värde huvudeffekt av tid på de dimensioner som har statistisk signifikant skillnad mellan testtillfälle ett och testtillfälle två vid ANOVA tidpunkt x åldersgrupp

	Testtillfälle ett		Testtillfälle två		$F (1, 89)$
	M	SD	M	SD	
Ledande	4.47	2.07	4.26	1.97	3.22*
Övertygande	4.08	1.86	4.51	1.96	7.78***
Hänsynsfull	6.03	2.03	5.62	1.90	5.94**
Betraktande	5.23	1.98	5.68	2.08	7.20***
Genomtänkt	4.46	2.06	4.89	1.96	5.88**

$p < .10^*$, $p < .05^{**}$, $p < .01^{***}$

T-test för beroende mätningar utfördes för att besvara frågeställning tre, för att se om det fanns någon skillnad inom grupperna yngre och äldre. I Tabell 5 och 6 redovisas de signifikanta skillnader på angivna dimensioner.

Tabell 5

Medelvärde, standardavvikelse och t -värde på de dimensioner som har statistisk signifikant skillnad mellan testtillfälle ett och testtillfälle två vid t -test inom gruppen yngre

	Testtillfälle ett		Testtillfälle två		$t (49)$
	M	SD	M	SD	
Hänsynsfull	6.04	1.95	5.56	2.00	0.13**
Betraktande	5.12	1.88	5.50	2.09	1.72*
Analytisk	4.22	1.95	4.72	2.10	2.59**
Ambitiös	4.96	1.73	4.36	1.74	2.73***

$p < .10^*$, $p < .05^{**}$, $p < .01^{***}$

Tabell 6

Medelvärde, standardavvikelse och t -värde på de dimensioner som har statistisk signifikant skillnad mellan testtillfälle ett och testtillfälle två vid t -test inom gruppen äldre

	Testtillfälle ett		Testtillfälle två		$t (40)$
	M	SD	M	SD	
Ledande	4.39	2.17	3.95	1.87	2.15**
Övertygande	3.59	1.80	4.34	1.91	4.18***
Betraktande	5.37	2.12	5.90	2.06	2.03**
Genomtänkt	4.51	2.36	5.27	2.13	2.31**

$p < .10^*$, $p < .05^{**}$, $p < .01^{***}$

Diskussion

Syftet med studien var att undersöka personlighetstestet Shapes med test-retest-metod för att se om testet är stabilt över tid. Vidare skulle det undersökas om det fanns en skillnad inom grupperna kön och ålder mellan testtillfälle ett och två. Detta syfte skulle uppnås genom att besvara tre frågeställningar. Beträffande frågeställning ett visar resultatet att det fanns positiva samband för samtliga dimensioner och att fem av de arton dimensionerna visar på medelvärdesskillnad mellan de två testtillfällena. Resultat med avseende på frågeställning två visade på att det var lika många dimensioner hos de äldre som hos de yngre som hade någon signifikant förändring mellan testtillfälle ett och två. Gällande frågeställning tre som handlade om skillnader inom gruppen kön över tid visar resultaten att det inte finns några skillnader hos män men tre dimensioner hos kvinnor fick signifikant medelvärdesskillnad mellan testtillfälle ett och två.

Den första frågeställningen handlade om det finns signifikanta skillnader mellan testtillfälle ett och testtillfälle två på samtliga dimensioner i personlighetstestet. Resultatet visar att samtliga dimensioner har positiva samband med varandra, mätt med Spearmans rangkorrelation mellan de olika tidpunkterna. Resultatet visar att sambanden varierar inom och mellan grupperna. Exempelvis inom gruppen yngre har dimensionen självständig .40 och dimensionen sällskaplig .89. Fem av alla dimensioner (ledande, övertygande, hänsynsfull, betraktande och genomtänkt) visar på medelvärdesskillnad mellan de två testtillfällena i de ANOVA som utfördes. Detta skulle kunna innebära att antingen är testet Shapes inte tillräckligt reliabelt för att mäta just dessa dimensioner, alternativt att dessa dimensioner inte är lika stabila över tid som de andra dimensionerna.

Frågeställning två gällde om det fanns någon skillnad inom åldersgrupperna vid testtillfälle ett och testtillfälle två. Enligt Costa och McCrae (1998, 2006) och Terracciano et al. (2010) borde de yngre personerna ha större skillnad än vad de äldre har eftersom uppdelningen av åldersgrupper sker vid trettioårsåldern, den tid då personligheten bör stabiliseras enligt deras teorier. Vid t-testet visade det sig att det var lika många dimensioner som det var skillnad på hos den äldre gruppen (ledande, övertygande, betraktande och genomtänkt) som hos den yngre gruppen (hänsynsfull, betraktande, analytisk och ambitiös). Detta resultat stödjer därför inte Costa och McCrae samt Terracciano et al. forskning. Robert och DelVecchios (2000) teori säger att personlighet stabiliseras först vid 50 år. Detta stämmer mer överens med denna studies resultat då åldersgrupperna är indelade i 18 – 33 och 34 – 69 år, vilket innebär att även i den äldre åldersgruppen finns det ett åldersspann på 16 år där personlighet kan förändras. En förklaring till detta skulle kunna vara att de flesta i åldersgruppen äldre är under 50 år.

Beträffande frågeställning tre som handlade om skillnad inom kön vid testtillfälle ett och testtillfälle två så uppvisade män ingen skillnad inom gruppen medan kvinnor hade tre dimensioner (hänsynsfull, betraktande och genomtänkt) som förändrades över tid. Detta skulle kunna bero på att gruppen med män var betydligt mindre än gruppen med kvinnor. I studien av Soto et al. (2010) fann de att det fanns en skillnad i hur personligheten förändras mellan män och kvinnor över tid i neurotism och öppenhet. I denna studie fanns det också en signifikant skillnad mellan män och kvinnor, men inte på någon underdimension till neurotism eller öppenhet.

Totalt är det sju dimensioner som har förändrats på något sätt mellan testtillfällena i denna studie. Då dimensionerna ledande, övertygande, hänsynsfull, betraktande, genomtänkt, analytisk och ambitiös inte var stabila över tid är detta något som användaren av testet Shapes bör tänka på. Till exempel om ett företag söker en person med högt på dimensionen analytisk kanske denna egenskap bör mätas med kompletterande metoder, som exempelvis

begåvningsstest. En förklaring till att dessa dimensioner inte var stabila kan bero på att människor förändras på just dessa dimensioner under kort tid och därför är inte Shapes eller något annat test ett bra mått på ovan nämnda dimensioner. En intressant aspekt i sammanhanget är tidsfaktorn. Skulle dessa dimensioner förändras även om tiden var längre mellan tillfällena eller skulle det innebära att det blev fler dimensioner som förändrades över tid?

Antalet dimensioner med skillnader skulle kunna bli färre vid ett större urval, då deltagarantalet gjorde att en högre signifikansnivå sattes. En annan aspekt på undersökningen är tiden som gick mellan testen. Det hävdas att personlighet är stabilt över tid men trots de få veckor som gick mellan testen uppkom skillnader.

Studiens största brist var att författarna själva inte ansvarade över själva mätinstrumentet och att resultatet genereras av Cut-e's egna system. Detta gjorde att författarna på grund av sekretessbestämmelser inte hade insyn över vilka poäng som varje item genererar eller vilka frågor som mäter vilken dimension. Detta resulterade i att varken faktoranalys eller Cronbach's alfa kunde utföras. Dock har Cronbach's alfa beräknats av Cut-e 2010 som påvisade höga värden (Cut-e, shapes technical manual, 2010).

Ett hot mot validiteten kan vara att deltagarna utförde test två veckan innan jul, detta kan ha medfört att vissa deltagare kände sig mer stressade än vid testtillfälle ett. Dock finns det alltid en risk vid all form av testning att yttre faktorer påverkar olika deltagare olika mycket som i sin tur kan påverka resultatet. Tiden mellan de båda testen kunde ha varit längre men på grund av tidsramarna för studien kunde inte perioden mellan testtillfällena förlängas.

Till skillnad mot många andra studier som baseras på en homogen grupp som exempelvis studenter har denna ett urval hämtat från Facebook vilket bör innebära att gruppen är mer heterogen. I denna studie var det viktigt att urvalet inte var en homogen grupp då det handlar om personlighet. Däremot hade deltagarantalet kunnat vara högre.

En kritik mot personlighetstest är att deltagarna fuskar och förskönar sitt resultat i hopp om att utge en bra bild av sig själv (Mabon, 2002). Detta är något som kan ha påverkat resultaten i denna studie då författarna kände till deltagarna via sina Facebookkonton och detta kan ha gjort att deltagarna förskönade sitt resultat. Författarna försökte dock minimera denna risk genom att informera att ingen av författarna eller någon annan kunde ta del av deltagarnas kompetensprofiler. Ett annat problem skulle kunna vara minneseffekt, som enligt Cohen och Swerdlik (2005) kan vara en nackdel vid test-retest. Då Shapes är adaptivt är det mindre risk för minneseffekt då frågornas ordningsföljd och upplägg hela tiden ändras beroende på deltagarens svar. Då varje påstående består av block är det också svårare att komma ihåg exakt poäng för varje item i varje block.

Test-retest-metoden är en metod som inte brukar användas och då det inte heller har gjort tidigare på Shapes har studien bidragit med ny kunskap. Utöver detta ligger internetbaserade test i tiden och kommer troligtvis att användas mer i framtiden.

Genom studien har idéer på vidare forskning uppkommit. Till exempel behövs fler studier med en jämnare könsfördelning för att bekräfta om det finns skillnad mellan män och kvinnor på personlighet över tid. Trots många studier kring vilken ålder personlighet stabiliseras tycks forskare inte vara eniga och även i denna studie är det osäkert när personligheten stabiliseras åldersmässigt. Mer studier behövs i ämnet. Ytterligare förslag på vidare forskning är att komplettera denna studie med en kvalitativ undersökning då det är viktigt att få en djupare förståelse kring deltagarnas upplevelse av testet i sig.

Avslutningsvis visade det sig att det fanns skillnader på några av dimensionerna men att de flesta dimensionerna i testet Shapes var stabila över tid. Även om personlighetstest har sina brister och kan antas inte mäta allt perfekt har ändå personlighetstest högre prediktionsvaliditet än många av de traditionella urvalsmetoderna som används idag. Detta innebär att i rekryteringssammanhang bör därför personlighetstest vara ett av verktygen för att

rätt person ska hamna på rätt plats.

Referenser

- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Bartram, D. (2005). The great eight competencies. A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185-1203. doi: 10.1037/0021-9010.90.6.1185
- Bartram, D., Robertson, I.T., & Callinan, M. (2002). Introduction: A framework for examining organizational effectiveness. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1-10). Chichester, UK; Wiley.
- Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150-157.
- Brent, W. R., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3-25. doi: 10.1037//0033-2909.126.1.
- Cohen, R. J., & Swerdlik, M. S. (2005). *Psychological testing and assessment. An introduction to tests and measurement (6th ed.)*. New York, NY: The McGraw-Hill Companies.
- Costa, P. T. Jr., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO personality inventory. *Journal of Personality and Social Psychology, 54*, 853-863.
- Costa, P. T. Jr., & McCrae, R. R. (2006). Age changes in personality and their origins: Comment on Roberts, Walton, and Viechtbauer. *Psychological Bulletin, 132*, 26-28.
- Cut-e. (2010). shapes technical manual. Hamburg: cut-e.
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology, 88*, 545-51
- Mabon, H. (2002) *Arbetspsykologisk testning: om urvalsmetoder i arbetslivet*. Stockholm: Psykologiförlaget.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*, 3-25.
- Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology, 21*, 489-509. doi: 10.1007/s10869-007-9038-9
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419-450. doi:10.1146/annurev.psych.59.103006.093716
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.
- Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*, 827-868. doi:10.1111/j.1744-6570.2008.00132.x
- Skorstad, E. (2008). *Rätt person på rätt plats, Psykologiska metoder för rekrytering och ledarskapsutveckling*. Lund: Studentlitteratur.

- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2010). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*, 330-348.
- Sturman, M. C., Chermie, R. A., & Cashen, L. H. (2005). The impact of job complexity and performance measurement on the temporal consistency, stability and test-retest reliability of employee job performance ratings. *Journal of Applied Psychology, 90*, 269-283
- Terracciano, A., McCrae, R. R., & Costa, P. T. Jr. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality, 44*, 31-37.
- Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.
- Vaidya, J., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology, 83*, 1469-1484.