



## **The double dupe**

### **– Den dubbla duperingen**

**Författare: Magnus Axelsson och Marcus Dalgren**  
**Handledare: Leif Rydstedt**

**Enskilt arbete i/Examensarbete Arbets- och organisationspsykologi 10 poäng,  
fördjupningsnivå 1**

**10 p Uppsats**

**Institutionen för Individ och samhälle**

**Oktober 2005**

## Abstract

This experiment has investigated whether three persons using the Thomas PPA were significantly better than chance in judging if the results being fed back were correct. The respondents' ability to judge if the correct test had been fed back was also investigated. 30 respondents were split into three groups and half of them got their tests exchanged. The three test users then met ten respondents each, fed back the test and then judged if it was right test. The respondents also judged how well the results corresponded with their self image and after the experiments' purpose had been revealed they also guessed if it was their test. The results indicate that the test users were not better than chance and that the respondents allowed themselves to be duped.

*Sammanfattning.* I experimentet har undersökts huruvida tre personer som använder Thomas PPA var markant bättre än slumpen på att avgöra om resultat som återkopplades stämde. Det undersöktes också om respondenter kunde avgöra om de fått rätt test återkopplat till sig. 30 respondenter delades in i tre grupper där hälften fick sina test roterade sinsemellan. Tre återkopplare fick träffa tio respondenter var, återkoppla testet och skatta om det var rätt test som hade återkopplats. Respondenterna fick skatta hur väl de ansåg att resultatet stämde in med deras självbild och efter att experimentets egentliga syfte delgetts också skatta om det verkligen var deras test som hade återkopplats. Resultaten pekar åt att återkopplarna inte var skickligare än slumpen och att respondenterna lät sig duperas.

I Personal & Ledarskap (2005) står det att läsa om att allt fler tillverkare av arbetspsykologiska test lämnar in sina produkter för granskning av Stiftelsen För Tillämpad psykologi (STP). Anledningen är att kunderna börjat kräva STP-granskade test. Som testtillverkare betalar man 50 000 kr för att få sitt test utvärderat av den oberoende instans i Sverige som sysslar med sådant, STP. Enligt Personal & Ledarskaps uppskattning omsätter arbetspsykologiska instrument någonstans mellan 200-500 miljoner kronor i Sverige per år och personlighetstest är de som säljer mest. De tre mest sålda testen i Sverige är MBTI, OPQ32 och Thomas PPA (även kallat *Thomas-testet*).

År 2004 skrev vi B-uppsats (tillsammans med ytterligare en klasskamrat) med titeln: *Vad borgar för hög kvalitet i personlighetstest?* (Axelsson, Dalgren & Lansinger, 2004). Uppsatsen var en litteraturstudie som syftade till att ge fingervisningar om vilka krav man som kund/testköpare bör ställa på ett personlighetstest. Andra frågor som togs upp var vilka krav som ställs på en testanvändare och hur den svenska testmarknaden ser ut. Det resultat våra litteraturstudier genererade talade för att testköpare bör ställa höga krav på att testet är konstruerat med grund i vetenskap. Även om personlighetstestning inte är en exakt mätmetod torde vetenskaplig grund kunna utgöra någon form av kvalitetsstämpel jämfört med test som saknar sådan grund i vedertagna teorier. Dessutom ställs tämligen ringa krav för att få använda personlighetstest trots att det rör sig om psykologiska instrument som kan ha stor betydelse i till exempel urvals/-rekryteringssammanhang. Det förekommer exempelvis test på svenska marknaden vars tillverkare endast kräver att användaren går tre dagars utbildning för att bli certifierad på produkten och få använda den i rekryteringssammanhang, utan tidigare utbildning i beteendevetenskap. En uppenbar risk finns att allt för stor vikt läggs vid resultatet av ett personlighetstest, även om det syftar till att bara ge ett bidrag till beslutsunderlag, då det ger en bild av personen ”svart på vitt”.

Denna C-uppsats spinner vidare på spåret personlighetstestning och inkluderar också en empirisk studie i form av ett duperingsexperiment. Inom ramen för denna C-uppsats har tillåtelse givits av Skandinaviska Ledningsgruppen (SLG) för att använda oss av Thomas PPA för att utföra experimentet. I studien har 30 personer fått göra Thomas-testet och sedan fått det återkopplat (fått sin profil presenterad för sig) av en på testet certifierad användare. Tre testanvändare har fått återkoppla testet till tio personer var.

Vi kommer nu kort redogöra för tidigare genomförda duperingsexperiment som legat till grund för vårt hypotesskapande.

### *Trankells duperingsexperiment*

År 1953 genomförde Trankell ett duperingsexperiment bland sina studenter (Trankell, 1961). Försökspersonerna var redan väl bevandrade i kritiskt och vetenskapligt tänkande, vilket borde tala för att de skulle ställa sig kritiska i aktuellt sammanhang. Några var dessutom redan yrkesarbetande. Trankells kurs behandlade personlighetsteorier och personlighetsmätningar. Vid tidpunkten för experimentet skulle ämnet "Självkänedom genom handstilstolkning" behandlas. Studenterna erbjöds att delta i en studie där en erkänd grafolog skulle analysera samtliga deltagares handstilar. Trankell lät påskina att han själv inte hade vidare stor tilltro till metoden. Emellertid genomfördes experimentet genom att studenterna fick skriva en viss text (samma för alla) och sedan underteckna med sin namnteckning. De handskrivna texterna gick iväg på analys och en vecka senare kom resultaten i personliga förseglade kuvert. Deltagarna instruerades att inte visa sina resultat för varandra. Kuverten innehöll också en blankett med frågor rörande hur träffsäker de ansåg analysen vara. Blanketten skulle fyllas i efter genomläsning av analysen.

Samtliga studenter ansåg att grafologin inte skulle förkastas som användbar metod och många tyckte att svaren de fått stämde överraskande bra. När Trankell sedermera läste upp sitt resultat uppstod viss förvåning. Det blev då uppenbart att samtliga deltagare fått precis samma utlåtande. Försökspersonerna insåg att de accepterat ett standardutlåtande som något individuellt utformat. Trankells slutsats var att ingen personlighetsinventerande metod kan sägas ha ett värde bara för att undersökt person anser utlåtandet därav vara korrekt. Således menar Trankell (1961) att face validitet (även kallad *ytlig* validitet, se närmare förklaring på sidan 6) inte säger något om metodens personlighetsinventerande förmåga. Trankell menar vidare att människor ofta tror att de är mer speciella än vad de faktiskt är och detta är förklaringen till varför de låter sig duperas på exempelvis ovan beskrivet sätt.

### *Gordons duperingsexperiment*

Gordon genomförde år 2000 en studie på några av sina studenter, liknande den Trankell gjorde. Dock är inte studien publicerad och Gordon själv hänvisar till sin hemsida ([www.gordonconsulting.se](http://www.gordonconsulting.se)).

43 studenter deltog i experimentet inom ramen för en kurs i socialpsykologi. Vid tidpunkten behandlades ämnet "moderna personlighetsmätande instrument". Försökspersonerna fick göra ett personlighetstest som sedan skulle tolkas av en specialist. För att påvisa duperingseffekt oberoende av specifikt test använde Gordon tre olika tester som fördelades slumpmässigt bland deltagarna. En vecka efter att studenterna gjort testet erhöll de sitt resultat på särskilt papper som var undertecknat av en psykolog. Försökspersonerna fick läsa sitt resultat enskilt och sedan svara på i vilken utsträckning utlåtandet överensstämde med deras självbild. 35% ansåg att utlåtandet stämde till punkt och pricka. 49% ansåg att utlåtandet stämde väl med något undantag och 16% ansåg att utlåtandet stämde till vissa delar men inte till andra. Ingen tyckte att utlåtandet var helt fel.

Gordon ([www.gordonconsulting.se](http://www.gordonconsulting.se)) påpekar att hans resultat av studien överensstämmer med Trankells. Det utlåtande som Gordons studenter fått på sina papper byggde till viss del på texter ur manualer för på marknaden förekommande personlighetstest. I enlighet med Trankells tankar påpekar Gordon att det finns uppenbar risk att människor låter sig duperas av "färdiga" resultatbeskrivningar som köpts dyrt. Gordon hävdar vidare att vi lever i en tidsålder där individen är i fokus, där vi överöses med betygelse om att vi är unika. Personlig utveckling och självständighet är honnörsord och människor skall betraktas som individer och inte som delar av ett kollektiv. Således anser Gordon att duperingseffekten får större chans att framträda i och med det rådande synsättet.

Dagens människa är alltså enligt Gordon benägen att tro att ett utlåtande stämmer när det är grundat på ett personlighetsinventorium och överensstämmer med självbilden eller bilden av den man önskar vara. Benägenhet att vara kritisk och inse att utlåtandet kan stämma in på vem som helst är således inte lika stor. Gordon hävdar bestämt att människan har svårt att se sig själv men däremot lätt för att fantisera om sin intrapsykiska beskaffenhet. Att bli imponerad av ett utlåtande som stämmer med självbilden ligger således nära till hands men som Trankell också hävdar så utgör inte den ytliga validiteten något bevis på metodens värde. Gordon höjer också ett varningens finger för de krav som ställs på användarna av personlighetstest. Gordon skriver att det blir allt vanligare att även sådana som inte har någon beteendevetenskaplig utbildning kan gå en kort kurs i tillverkarens regi, certifieras och sedan vara behöriga att använda ett test. Att kunna bedöma människor anser Gordon vara ett komplext område och kräver betydligt mer materialinsamlande än vad som görs genom nyttjandet av ett personlighetstest.

### *Barnumeffekten*

Den duperingseffekt som både Trankell och Gordon sett har döpts till barnumeffekten efter P.T. Barnum som var en välkänd cirkusdirektör som ansåg att grunden till hans framgång låg i två grundläggande principer: "Det föds en idiot varje minut" och "ha något för alla" (Boyce & Geller, 2002). Barnumeffekten refererar till människors tendens att acceptera allmängiltiga och tvetydiga påståenden om deras personlighet som riktiga och dessutom direkt riktade till dem.

I en studie av McDonald och Standing (2002) studerades barnumeffekten i relation till det som på engelska kallas "self-serving bias". Self-serving bias innebär att människor har ett behov av att se och uppleva sig själva som bra människor, oberoende av objektiva fakta. I detta sammanhang pekar det då på att människor bör ha en större tendens att acceptera positiva beskrivningar av sig själva. Self-serving bias visade sig vara starkare än Barnumeffekten då deltagarna i studien skattade att de positiva beskrivningarna stämde bättre överens med deras personlighet än de som var neutrala eller negativa.

I en annan studie utförd av Standing och Keays (1986) fick försökspersonerna i experimentet varsin lista med 66 adjektiv och ombads välja ut de som stämde överens med deras personlighet och de som inte gjorde det. Tre veckor senare fick halva gruppen tillbaka en lista med adjektiv som beskrev dem och den andra hälften fick en beskrivande profil som genererats med ett datorprogram. I båda dessa grupper byttes resultaten ut för hälften av gruppen för att studera om detta ger någon effekt i deras skattning av hur väl resultatet stämmer överens med deras personlighet.

En skillnad mellan denna studie jämfört med många av de tidigare var att de felaktiga resultaten som återgavs var olika för varje person och hade genererats utifrån en av de andra personernas resultat i experimentet. Trots detta var det inga signifikanta skillnader i skattningar mellan grupperna som fick tillbaka sitt riktiga material och de som fick tillbaka någon annans material. I alla fallen tenderade skattningarna att vara mycket positiva och ansågs ge en korrekt beskrivning av respondentens personlighet (Standing & Keays, 1986).

### *Psykometri*

Definitionen av begreppet psykometri är vetenskapen om mätning av psykologiska egenskaper och beteenden. Psykometrin ligger därmed till grund för konstruktionen och utvecklingen av psykologiska test (Mabon, 2004).

I denna del kommer de två psykometriska begreppen reliabilitet och validitet beskrivas kortfattat tillsammans med en mer ingående förklaring av vad ipsativitet innebär då det instrument som används i studien är ett ipsativt test. Michells kritik av psykometrin kommer även att tas upp och relateras till ipsativiteten.

## *Reliabilitet*

Reliabiliteten är ett mått på testverktygets precision och mäts framförallt på två olika sätt. Dessa två metoder kallas för test-retest och intern konsistens. Test-retest reliabilitet innebär att testet fylls i av samma person vid två olika tillfällen och de två olika resultaten jämförs sedan. Eftersom personlighetsdrag antas vara någorlunda stabila över tid bör resultaten vara snarlika. Det är önskvärt att korrelationskoefficienten för test-retest reliabiliteten ligger så nära +1,0 som möjligt och standarden för personlighetstest ligger enligt Friedman och Schustack (1999) på +0,8.

Stor varians mellan de två testtillfällena behöver dock inte indikera ett mätfel. Det kan även vara så att det skett en verklig förändring under den tid som gått och då skall resultaten variera. För att undvika att testrespondenten kommer ihåg sina svar från det första tillfället är det enligt Kline (2000) brukligt att låta minst tre månader gå mellan testtillfällena. Intern konsistens ger i sin tur ett mått på hur likartade testfrågorna är. Det är alltså önskvärt att frågorna inom en dimension är likartade och har hög intern konsistens medan ett lågt värde skulle peka på att frågorna mäter helt olika saker. Det finns dock en fin balansgång att gå här och Raymond Cattell (1977, i Kline, 2000) anser att de dimensioner som mäts i ett personlighetstest är ”breda” och att för hög intern konsistens därför ej är önskvärt. Om ett tests interna konsistens är väldigt hög så är frågeområdet också väldigt smalt och det kan då i värsta fall röra sig om olika varianter av samma fråga. Testet kommer då ha hög intern konsistens men låg validitet (Kline 2000).

Intern konsistens kan mätas genom split-halfmetoden vilket innebär att testet delas i två lika stora delar som sedan korreleras med varandra. Denna delning är dock något godtycklig då testet kan delas på en mängd olika sätt och de olika delningarna ger olika ofta olika resultat. Reliabiliteten är även kopplad till testets längd, ju fler frågor desto högre reliabilitet. Detta innebär att split-half reliabiliteten är en underskattning av den verkliga reliabiliteten (Mabon, 2004).

## *Validitet*

Det finns en mängd olika validitetsbegrepp inom psykometri och de är alla mätmetoder för att utröna om ett test verkligen mäter de variabler som det avser att mäta. Om validiteten är hög innebär detta att testet med största sannolikhet mäter det som det avser att mäta. Metoderna för att mäta validiteten är dock inte lika enkla eller självklara som de metoder som utvecklats för reliabiliteten. De olika metoder som används återspeglas i olika validitetsbegrepp och de som kommer att tas upp är face validitet, samstämmig validitet, kriterierelaterad validitet och begreppsvaliditet.

Face validitet är det ytligaste validitetsmättet och mäts genom att fråga testutförarna om de upplever att testet stämmer väl överens med det som skall testas (Nevo, 1985). De tidigare nämnda studierna av Trankell (1961) och Gordon ([www.gordonconsulting.se](http://www.gordonconsulting.se)) har dock visat att detta mått inte är helt oproblematiskt då studierna indikerar att vi tenderar att anse att resultat stämmer väl överens med vår självbild även om det inte är vårt egna testresultat vilket indikerar att vi är benägna att ge ett positivt utlåtande delvis oberoende av resultatet som presenteras. Face validitet är även ett mått på testets genomskinlighet och upplevda meningsfullhet. Om frågorna i testet är konstiga eller upplevs som irrelevanta av de svarande kan detta leda till brister i motivation med konstiga svar som följd (Mabon, 2004).

Även om det är viktigt att de som testas fyller i testen efter bästa förmåga så har genomskinligheten även sina nackdelar. Hög genomskinlighet innebär ofta även att det är lätt för testutföraren att räkna ut vad testet mäter och detta kan i sin tur leda till att hon försöker manipulera resultaten i sin egen fördel.

Detta är ett stort problem i urvalssituationer och i andra situationer där testrespondenter tävlar med varandra och vill göra ett så gott intryck som möjligt. Målet blir alltså att konstruera ett test som upplevs som relevant av respondenterna men inte är alltför genomskinligt (Mabon, 2004).

Kriterierelaterad validitet delas upp i två olika former; samstämmig validitet och prediktiv validitet. Samstämmig validitet mäts genom att en respondent får fylla i två test som antas mäta samma fenomen och resultaten sedan korreleras. Metoden är beroende av att det redan finns mätinstrument med hög validitet att jämföra det nya mätinstrumentet med (Clark-Carter 1997). Inom exempelvis intelligenstestning finns det väl etablerade standardiserade test som kan användas som benchmarktest (Kline, 2000).

Inom personlighetspsykologin är det tyvärr ont om ordentliga benchmarks för olika variabler. Ett av de få benchmarktest som finns inom personlighetspsykologin är Eysenck Personality Questionnaire (EPQ) konstruerat av Hans Eysenck (Kline, 2000). Eysencks test har funnits länge och har använts och utvecklats så länge att det anses vara en standard på de variabler som testet mäter (Kline, 2000).

Eftersom det oftast inte finns någon väl etablerad benchmark så korreleras testet med flera olika test som antas mäta samma eller liknande variabler. Detta brukar dock ge lägre korrelationer runt +0,4-0,5 och är i sig inget tillfredställande sätt att validera testet utan används som ett komplement till andra validitetsstudier (Kline, 2000).

Prediktiv validitet mäts genom att man sätter upp ett framtida tillstånd eller situation som testresultatet sedan skall förutsäga. Ett exempel kan vara att ett högt resultat på en viss personlighetsdimension skall förutsäga framtida framgång i arbetslivet. Eftersom testresultaten endast är en av de många faktorer som påverkar den framtida situationen förväntas enligt Kline (2000) endast medelstora korrelationer på mellan +0,3-0,4. Att sätta upp ett kriterie för prediktion är dock väldigt svårt för personlighetstest då de är bredare och diffusare än begåvningsstest. Detta leder till att även denna form av validitet används tillsammans med andra validitetsmått för att ge en heltäckande bild (Kline, 2000).

Begreppsvaliditet är ett mått på hur väl testet mäter den konstruktion eller begrepp som testet avser att mäta. Detta kräver att begreppen definieras explicit och har en klar, tydlig och väl avgränsad definition. Exempel på begrepp som används inom psykologin är intelligens och extroversion. Den psykologiska definitionen av dessa begrepp skiljer sig från de definitioner som används till vardags då de vardagliga definitionerna tenderar att vara för breda och vaga (Clark-Carter, 1997).

Begreppsvaliditeten delas in i konvergent och divergent begreppsvaliditet. Konvergent begreppsvaliditet innebär att testet skall konvergera med test som mäter relaterade begrepp. Om det t ex antas att intelligens är en generell förmåga så bör en persons resultat på ett verbalt intelligenstest och ett numeriskt intelligenstest uppvisa liknande mönster. Divergent begreppsvaliditet mäts genom att korrelera testet med andra test som mäter andra orelaterade begrepp. Resultatet bör då divergera eller skilja sig från de andra testen. Ett exempel på detta kan vara om ett test utvecklas som mäter läsningsförmåga. Detta test bör då divergera med intelligenstest då det enbart skall mäta läsningsförmåga (Clark-Carter, 1997).

### *Kritik av psykometrin*

Michell har i en rad artiklar (1997; 2000; 2003; 2004) och i en bok (1999) kritiserat psykometrins grundförutsättningar. Michell (2000; 2004) kallar psykometri för en patologisk vetenskap, vilket innebär att en av grundförutsättningarna för psykometrin inte har testats och att detta sedan har "gömts" undan i senare forskning inom området. Den grundförutsättning inom psykometrin som Michell inriktat sig på är hypotesen om att vissa psykologiska attribut innehar en kvantitativ struktur.

När ett test utvecklas antas det helt enkelt dels att det teoretiska attributet faktiskt har en kvantitativ struktur men också att resultaten från testet är kvantitativt relaterade till attributet. Båda dessa antaganden är empiriska antaganden men trots detta så testas de inte utan antas helt enkelt vara sanna (Michell, 2000).

Kline (2000) redovisar Michells argument och instämmer med dem i stora drag. Till skillnad från Michell så förkastar inte Kline hela fältet utan anser att dessa grovmaskiga verktyg bör användas tills bättre mätverktyg kan utvecklas. Eftersom det inte finns någon ersättning för de psykologiska test som används idag anser Kline (2000) att dessa verktyg bör användas även i fortsättningen då de visat sig användbara inom en mängd olika områden, bl a urval och rekrytering, klinisk psykologi och inom rådgivning. Både Kline (2000) och Michell (1999) markerar dock att psykometrin inte ännu kan ses som en vetenskaplig mätmetod.

### *Ipsativitet*

Då ett antal av de populäraste personlighetstesten på den svenska marknaden har ett ipsativt upplägg ges här en djupare förklaring av vad ipsativitet innebär och vad det har för konsekvenser.

Ett test är ipsativt om frågorna är vad som kallas för "forced choice" där varje val poängsätts och den totala summan av alla val alltid blir densamma. Ett vanligt upplägg är att respondenten får välja mellan exempelvis fyra olika alternativ och måste välja det som stämmer in bäst respektive sämst på den egna personligheten. Eftersom denna sortens val är ett såkallat preferensval visar resultatet från ett ipsativt test bara de olika dimensionernas styrka i relation till varandra. Det ger alltså inget absolut mått för någon av de dimensioner som mäts (Johnson, Wood & Blinkhorn, 1988).

Ett exempel på detta är om en person som dricker mycket sällan och en alkoholist får fylla i ett ipsativt test där de måste rangordna vilka alkoholsorter som de föredrar. Resultaten skulle då kunna visa att de båda respondenterna har samma preferenser men att utifrån det dra slutsatsen att de skulle ha liknande konsumtionsvanor skulle vara grovt felaktigt (Kline, 2000).

Det är även viktigt att påpeka att de faktorer som mäts i ett ipsativt test inte är oberoende av varandra på grund av den ipsativa konstruktionen. Eftersom respondenten hela tiden väljer och aktivt utesluter vissa alternativ leder detta till att en ökning i en variabel orsakar en minskning i någon eller några av de andra variablerna. Detta leder också till att om resultatet på tre skalor för ett ipsativt test med fyra skalor är känt så kan resultatet för den fjärde skalan förutsägas enbart på grund av ipsativiteten. Eftersom skalorna är beroende av varandra så kan detta bli vanskligt om man påstår sig försöka mäta oberoende variabler (Meade, 2004).

Meade (2004) påpekar även i likhet med många andra före honom att korrelationerna i ett ipsativt test grupperar sig på ett visst sätt med en tendens till negativa korrelationer mellan skalorna ännu en gång enbart på grund av hur testet är konstruerat. Detta kan leda till att variabler som egentligen är högt korrelerade inom en population blir negativt korrelerade i ett ipsativt test då respondenten måste välja bort ett av alternativen. Den negativa korrelationen beror då enbart på den ipsativa konstruktionen, inte på det egentliga sambandet mellan de två variablerna.

Den enda sortens normer som kan samlas in på ipsativa test är rangordningar. Exempelvis skulle det vara meningsfullt att argumentera för att en majoritet av män föredrar öl framför cider och andra alkoholsorter men även detta är av tveksamt värde då avståndet mellan olika ranger och deras styrkeskillnader är okänt. Dessa resultat skulle alltså vara tämligen ytliga då allt framgår av en sådan studie är just att män föredrar öl framför cider och andra alkoholsorter (Kline, 2000).



För att resultatet från ett ipsativt test verkligen skall framstå som meningsfullt krävs det att resultatet diskuteras med respondenten. Det är först i en diskussion med respondenten som innebörden bakom preferensvalen verkligen kan komma fram. Detta innebär att ipsativa test fungerar bäst i de sammanhang där resultaten endast skall användas som diskussionsunderlag där kvantifiering och jämförelse mellan respondenter inte är viktigt. Trots detta förekommer ipsativa test i rekryteringssammanhang där syftet ofta är just att jämföra olika kandidaters lämplighet för ett jobb (Kline, 2000).

Kline (2000) själv anser att det egentligen inte finns några bra anledningar för att använda ipsativa test då även normativa test kan användas som diskussionsunderlag och kan dessutom jämföras sinsemellan.

Det har dock förekommit en debatt fram och tillbaka mellan de som förespråkar ipsativa test och de som likt Kline anser att ipsativa test är vanskliga. Johnson et al (1988) tar upp många av de problem som har nämnts ovan och ondgör sig över att denna form av personlighetstest florerar på den brittiska marknaden. De påpekar även att den tidens psykometriska litteratur tenderar att ignorera de problem som finns med ipsativa test och när det väl tas upp tenderar det att endast vara mycket ytligt och kortfattat.

De tar själva upp fem olika problem med ipsativa test (Johnson et al, 1988):

1. De kan inte, som nämnts tidigare, jämföras skala mot skala.
2. Korrelationer mellan ipsativa skalor kan inte faktoranalyseras legitimt på det vanliga sättet.
3. Reliabiliteten i ipsativa test tenderar att överskattas jämfört med deras verkliga reliabilitet.
4. Av samma anledning tenderar även validitetssiffror att bli för höga.
5. Medelvärde, standardavvikelser och korrelationer från ipsativa test är inte oberoende och kan därför inte tolkas på det vanliga sättet.

Saville och Willson (1991) har försökt besvara den kritik som riktats mot ipsativa test. I deras artikel användes simulerade data där simulerade normativa och ipsativa testresultat jämfördes med hypotetiska "riktiga" resultat. I denna studie visade de, tvärtemot vad som tidigare skrivits i ämnet, att det gick att jämföra normativa och ipsativa resultat då de simulerade dataresultaten gav höga korrelationer mellan de normativa och ipsativa resultaten. Dessutom gav inte den ipsativa datan några överdrivna reliabilitetsvärden och de ipsativa resultaten korrelerade i vissa fall bättre med de hypotetiska "riktiga" poängen.

Cornwell och Dunlap (1994) ger svar på tal och påvisar en mängd brister i Saville och Willsons studie. Bl a skall Saville och Willson bortsett från de negativa korrelationer som uppstår i den ipsativa datan och som inte finns i den normativa. Deras försök att replikera faktoranalysen gav inte heller samma resultat som Saville och Willsons studie och det är därför tveksamt om de slutsatser som dras i deras studie verkligen är befogade.

Saville och Willson (1991) tar även upp några fördelar med ipsativa test. Den första fördelen som tas upp är att ipsativa test skall vara svårare att se igenom och därigenom "fuska" på. Salze et al (1962) har visat att det går att fejka svar på ett ipsativt test om personer instrueras. Martin, Bowen och Hunt (2002) har i sin studie jämfört fejkning på ipsativa och normativa test och anser att deras studie visar att det är betydligt svårare att fejka på ett ipsativt test. I experimentet fick deltagarna fylla i antingen ett normativt test eller ett ipsativt test tillsammans med ett frågeformulär om deras idealbild av en chef.

Bland dem som fick fylla i det normativa testet var det signifikanta skillnader mellan den grupp som ombads fylla i ärligt och de som fick instruktioner om att försöka fejka ett resultat som motsvarade deras ideal av en chef. De som fejkade hade resultat som stämde väl överens med deras ideal. I den gruppen som fick fylla i ipsativa test visade det sig däremot att de respondenter som ombads fejka sitt ideal inte kom markant närmre än de som fyllde i testet ärligt (Martin et al, 2002).

Studien undersökte även om det ideal som testrespondenterna hade om en chef stämde väl överens med det ideal som högt uppsatta HR-chefer har om chefsjobbet. Korrelationen mellan respondenternas ideal och chefernas ideal visade sig vara väldigt hög och statistiskt signifikant med en korrelationskoefficient på +0,86. Detta tycks alltså innebära att i de fall där normativa test används i rekryteringssammanhang så kan det finnas en relativt stor risk att respondenterna kan fejka det resultat som organisationen vill ha (Martin et al, 2002).

Martin et al (2002) studie har dock kritiserats av Meade (2004) då idealet som respondenterna ombads sätta upp gjordes med hjälp av normativa skalor. Respondenterna ombads betygsätta olika egenskaper från en skala på 1 till 10 där 10 är en egenskap som de anser vara väldigt viktig för jobbet. Det var dessa resultat som sedan jämfördes med respondenternas riktiga testresultat.

Det var alltså teoretiskt möjligt att sätta 10 på alla egenskaper vilket gör att resultatet från ett ipsativt test inte kan matcha detta då vissa egenskaper antingen kommer att bli låga eller dras mot mitten av skalan. Meade (2004) anser det vara förvånansvärt att Martin et al (2002) inte verkar ha tänkt på detta och istället konstruerat ett ipsativt idealformulär för de som sedan fyllde i de ipsativa testen. Den jämförelse mellan avstånden som gjordes blir alltså felaktig på grund att man försöker jämföra två helt olika sorters skalor som om de var ekvivalenta.

Meade (2004) påpekar dock att ipsativa test skyddar mot den enklaste formen av fejkning som innebär att man sätter det högsta poängalternativet på varje fråga i ett normativt test så stoppar detta inte respondenter från att försöka lista ut vilka alternativ som företaget eller organisationen ser som mest värdefulla och helt enkelt välja det alternativ som de tror stämmer bäst överens med detta.

Denna typ av fejkning är mycket svårare då respondenten först måste konstruera korrekta hypoteser om vad organisationen verkligen vill ha och sedan försöka svara på ett sådant sätt att resultatet från testet återger den bilden. Eftersom många test är tidsbegränsade på ett eller annat sätt och anställningssituationen ofta är väldigt nervös eller stressig gör detta till en avancerad och svår strategi för de som desperat vill göra ett gott intryck (Meade, 2004).

Ipsativa skalor undviker även en mängd biases som förekommer i normativa test tack vare sin konstruktion. I normativa test förekommer något som kallas för central tendency bias vilket innebär att vissa respondenter tenderar att konsekvent välja antingen mittenalternativet eller om ett sådant inte finns, välja det alternativ som ligger närmast. Den motsatta tendensen förekommer också där respondenten istället enbart väljer extremalternativen (Saville & Willson, 1991).

Saville och Willson (1991) påpekar också att ipsativa test, precis som livet, handlar om val. De hävdar att många av de val som människor är tvungna att göra i livet handlar om att välja mellan olika alternativ, det går inte att välja allihop och på det sättet är många av våra livssituationer ipsativa. De påpekar även att många personlighetsteorier, däribland Jungs typeteori, argumenterar för en ipsativ personlighetsstruktur. I Jungs typeteori som används i Myers Briggs test (MBTI) ställs personlighetsdragen upp som motsatspar, till exempel introversion-extraversion. Det är alltså inte möjligt att vara båda två samtidigt.

Som tidigare har nämnts har Michell (1997; 1999; 2000; 2003) i sina verk pekat på att hypotesen om den kvantitativa strukturen i personlighetsdrag aldrig har testats utan bara förutsatts trots att detta är en empirisk fråga. Alla normativa personlighetstest är beroende av att denna hypotes är sann, i alla fall för de personlighetsdrag som testet påstår sig testa. Om en metod för att testa detta skulle utvecklas och det visar sig att de personlighetsdrag som testas faktiskt inte har en kvantitativ struktur skulle detta vara en stor kris för alla normativa test.

Ipsativa test undviker dock detta dilemma då preferensval i sig inte är beroende av att de underliggande fenomenen verkligen har en kvantitativ struktur. Det skulle alltså mycket väl kunna vara så att även om det inte är meningsfullt att säga att jag har värdet X i personlighetsdraget Y så föredrar jag personlighetsdraget T framför Y. Ett enklare exempel skulle vara att bara för att jag föredrar Mercedes framför Volvo så behöver detta inte betyda att det är meningsfullt att säga att Mercedes är mera bil än Volvo.

### *Person Profil Analys – Historik, utformning och teori*

Person Profil Analys (hädanefter PPA eller thomastestet) utvecklades av Thomas Hendrickson år 1958 baserat på boken *Emotions of normal people*, skriven av William Moulton Marston (Marston, 1989). Hendrickson arbetsanpassade även verktyget för att det skulle kunna användas som ett analysverktyg i yrkesmässiga situationer. Marston utvecklade i sin bok *Emotions of normal people* en teori om hur människor agerar i samspel med sin omgivning. Denna teori delar upp mänskligt agerande i fyra olika beteendetendenser (SLG, 2004).

- Dominance (dominans) – Ett aktivt pådrivande beteende i fientliga miljöer.
- Inducement (inflytande) – Ett aktivt, socialt och ofta övertalande beteende i gynnsamma miljöer.
- Submission (stabilitet) – Ett avvaktande, förstående beteende i gynnsamma miljöer.
- Compliance (anpassning) – En tendens till regel- och normstyrt beteende i fientliga miljöer.

Testet har utvecklats under sina år på marknaden vilket lett till att begreppen Submission och Compliance bytt namn för att få bort den negativa klangen i orden. Dessa namnbyten har gjorts dels för att tydliggöra men även för att undvika att personer skall tolka dessa dimensioner som negativa när de egentligen skall vara så värdeneutrala som möjligt.

Om begreppen tolkas som negativa av respondenterna kan detta leda till att de aktivt försöker svara på ett sätt som ger en bild av dem där dessa tendenser spelas ned. Denna form av förvrängning förekommer alltid i någon form men det är såklart önskvärt att den är så låg som möjligt. Det kan nu vara på sin plats att beskriva de fyra dimensionerna mer ingående.

#### **Dominans (D)**

Dominansfaktorn indikerar en persons handlingskraft och initiativrikhet, framförallt i spända situationer och fientliga miljöer. En person med högt D anses trivas i miljöer där mycket konkurrens förekommer och är väldigt måldriven.

Personen agerar oftast rättfram och direkt gentemot andra människor och kan bli emotionell i situationer där åsikter går isär. Personen kan brusa upp i sådana situationer och skälla ut andra människor men sedan glömma detta vilket kan leda till sårade känslor hos människor i omgivningen.

En person med hög dominansfaktor är ofta intresserad av det ovanliga och riskfyllda. Denna person trivs bäst i föränderliga miljöer och är anpassningsbar, dock på sina egna villkor. Om dominansfaktorn är låg tenderar personen istället att vara mildare och mindre konkurrensdriven i sitt beteende. Det är då inte lika viktigt för personen att hävda sig själv i alla situationer och personen kan istället uppfattas som tveksam och osäker.

### Inflytande (I)

Inflytandefaktorn visar på en persons utåtriktade och sociala tendenser. Denna dimension är dock inte dominant på samma sätt utan visar mer på en övertalande och sällskaplig beteendeform. Övertalningen av andra görs på ett vänligt sätt och till skillnad från någon med hög dominansfaktor så används inte ren kraft i övertalningen.

Personer med högt inflytande är ofta intresserade av människor med allt vad det innebär och är ofta hjälpsam och främjar andra människor i sin omgivning. Denna person är lätt att umgås med och gör ett öppet intryck på andra människor. En person med högt på inflytandefaktorn har ofta en stor bekantskapskrets utöver de närmre vännerna. En person med högt I kan verka något ytlig och uppvisa ett något överdrivet känslomässigt beteende. En person med låg inflytandefaktor tenderar att vara något tillbakadragen, reserverad och tystlåten. Lågt inflytande indikerar även mer faktabaserat och logiskt tänkande.

### Stabilitet (S)

Som namnet på dimensionen antyder är detta behärskade personer som sällan brusar upp och som är tillmötesgående. En person med hög stabilitet bygger oftast endast upp en liten bekantskapskrets med nära vänner och är försiktig och tålmodig. Dessa personer är väldigt hjälpsamma mot dem som betraktas som vänner.

I sociala situationer försöker personen ofta uppehålla "status quo" då personen är försiktig med plötsliga, oväntade förändringar. Personen har ett stort ägandebehov och utvecklar ofta starka anknytningar till sin arbetsgrupp, vänner och familj. Personen tänker också på andra människor när hon fattar beslut och vill gärna förankra sina beslut innan hon handlar.

Om stabilitetsfaktorn är låg tenderar personen att vara mer impulsiv, rastlös och kanske även något missnöjd. Låg stabilitet kan även antyda nervositet eller ångest. En person med låg stabilitet tenderar att vara betydligt mer förändringsbenägen och är inte heller lika bunden till familj och vänner på samma sätt.

### Anpassning (C)

Denna dimension hänvisar till ett beteende som anpassar sig efter regler och normer. Personer med hög anpassningsfaktor vill gärna följa regler och riktlinjer som en garant för kvalitet. Personen är ödmjuk, lojal, saklig och kan framstå som något kylig. I anpassningsdimensionen finns tendenser till försiktighet och konservatism (Marston, 1928).

Personen vill gärna ta gott om tid på sig för att fatta beslut men uppvisar ofta en skarpsinnighet i att fatta rätt beslut vid rätt tidpunkt. Personen är detaljriktad och håller sig helst till metoder som visat sig effektiva tidigare i arbetet eller livet. En person med låg anpassningsfaktor föredrar istället autonomi och kan ha en viss svårighet med att följa regler. Personen har ofta starka åsikter och en stark vilja men kan vara något oförsiktig.

### *PPA:s utformning*

Thomas PPA är ett ipsativt test som består av 24 rader med fyra stycken adjektiv på varje rad. Respondenten skall på varje rad välja ut det adjektiv som hon tycker stämmer in bäst på henne i hennes nuvarande arbetssituation och det som stämmer in sämst. Om personen inte har ett jobb när hon fyller i testet ombeds hon tänka på hur hon var i sitt senaste arbete och om hon ej har jobbat skall hon tänka på hur hon är privat. Resultatet räknas sedan ut till tre olika profiler där profil 1 visar hur personen fungerar i sin nuvarande arbetssituation, profil 2 visar personens beteende under press och profil 3 visar personens självbild (SLG, 2004). I denna studie har vikten lagts vid profil 3.

### *PPA:s psykometriska egenskaper*

Forskning på PPA har främst utförts av SLG International själva och den granskning som Stiftelsen för Tillämpad Psykologi (hädanefter STP) har gjort är den första externa granskning som utförts på instrumentet i Sverige. De psykometriska data som redovisas nedan kommer alltså från SLG:s egna användarmanual där de hänvisar till interna studier som har gjorts. Det är dock oftast sammandrag av resultaten från de olika studierna som visas i manualen vilket kan göra det svårt att bedöma dem ordentligt.

I en studie utförd av SLG (2004) i Sverige mellan 2000-2001 med en population på 131 undersöktes test-retest reliabilitet. Alla respondenter var verksamma i arbetslivet och testningarna gjordes med tre till fyra veckors mellanrum. Notera att denna tidsperiod mellan testningen är markant lägre än den som rekommenderas av Kline (2000).

Test-retest reliabiliteten ligger på +0,7-0,9 för alla variablerna med ett undantag för C som låg på +0,69 på profil II. I denna studie är korrelationskoefficienten för C konsekvent lägre än de andra variablerna men det påpekas att detta inte förekommit i tidigare studier (SLG, 2004).

I STP:s granskning av Thomas PPA redovisas resultaten från en studie av testets interna homogenitet som beräknas med Cronbachs alfa. Resultaten varierar mellan +0,51-0,72. Även i detta fall är det C som står för det lägsta värdet (+0,51) medan D har det högsta (+0,72). S och I hamnar båda två på +0,62. Enligt STP:s egna normer bör den interna konsistensen ligga över +0,6 vilket alltså innebär att C ligger under deras norm för acceptabla värden (STP, 2004).

Det har även gjorts en studie på den interna konsistensen i PPA i Sverige under 2000-2001. Precis som i den andra studien var alla som ingick i studien yrkesverksamma och testningen gjordes på 127 personer. I detta fall kunde inte testningen göras på alla tre profilerna eftersom profil III räknas ut med hjälp av de andra två profilerna. De data som presenteras här gäller alltså bara för profil I och II. I denna studie varierar korrelationskoefficienterna mellan +0,73 och +0,51 där C ännu en gång står för de lägsta värdena. Det påpekas här att de låga siffrorna för C har förekommit även i andra studier som gjorts och detta antyder att denna dimension är något diffus och dåligt definierad (SLG, 2004). Överlag ligger dock reliabiliteten i studierna på fullt acceptabla nivåer.

I manualen ges inga resultat för face validitet men instrumentet påstås dock ha en god innehållsvaliditet då adjektiven som används är lätta att förstå och används i dagligt tal för att beskriva beteende på arbetet. Även instruktionerna anses ge tyngd åt innehållsvaliditeten då de ombeds tänka sig in i arbetsrelaterade situationer när de väljer mellan de fyra adjektiven. Det som i manualen kallas för begreppsvaliditet har tidigare omnämnts i denna uppsats som samstämmig validitet och undersöker alltså om testet korrelerar väl med andra test som man antar mäter samma eller liknande variabler. Resultaten som presenteras i denna del är något knapphändiga och det ges bara två exempel på korrelationer med OPQ, Occupational Personality Questionnaire. D-faktorn korrelerade positivt med konkurrens (+0,56) och beslutsam (+0,53). I-faktorn korrelerade positivt med sällskaplig (+0,64) (SLG, 2004).

Precis som det har påpekats tidigare uppstår det dock problem så fort ipsativa test korreleras med ett oberoende kriterium (exempelvis framgång i arbetslivet) då konstruktionen gör att summan av korrelationerna för de fyra olika dimensionerna måste bli 0 eller väldigt nära 0. Detta gör att det blir svårt att ta ställning till vad korrelationerna verkligen innebär då det mycket väl kan vara så att en egenskap som egentligen korrelerar högt med en utomstående faktor får ett lågt resultat p g a testkonstruktionen.

I en studie på kriterierelaterad validitet som utfördes 1994 med en population på 228 stycken försäljningsmäklare presenteras fullständiga resultat som visar upp det symptom som beskrivs ovan. Populationen delades upp i tre olika grupper beroende på deras försäljningsresultat och medelvärdena för de olika faktorerna jämfördes sedan. Det visade sig i studien att den grupp som hade uppnått bäst försäljningsresultat hade hög D- och I-faktor vilket även är ett väntat resultat (SLG, 2004).

Det är dock väldigt svårt att säga hur stor del av dessa resultat som är beroende av testets konstruktion eftersom summorna på varje rad mycket riktigt blir 0 eller väldigt nära 0. Det är inte osannolikt att höga D- och I-faktorer korrelerar positivt med försäljningsframgång men det är svårt att utröna på den ipsativa konstruktionen.

I STP:s granskning av Thomas PPA påpekas att instrumentets teoretiska förankring och bevis rörande validitet, reliabilitet och användbarhet är något svaga. STP anmärker också på att i de studier som redovisas så ges ofta inte en fullständig redovisning av data utan bara korta sammandrag. STP anser att SLG bör redovisa tydligare hur de grundläggande dimensionerna i DISC har tagits fram och hur de olika adjektiven för varje dimension har framkommit (STP, 2004).

STP (2004) anser även att sättet som respondentens svar tolkas är anmärkningsvärt. De adjektiv som respondenten anser stämmer bäst in på henne reflekterar hur hon fungerar i sin nuvarande arbetssituation (profil I), de som stämmer in sämst hur hon fungerar under press (profil II) och den tredje profilen (profil III) som består av skillnaden mellan profil I och profil II anses representera respondentens självbild. Enligt STP (2004) finns det ingen redogörelse för någon vetenskaplig forskning eller teoribildning som stödjer denna tolkning av respondentens svar och tenderar att försvaga testets trovärdighet. Det påpekas även att det finns andra test på marknaden som använder sig av DISC-ansatsen som tolkar respondentens svar på andra sätt.

Det har inte heller gjorts några begreppsvaliditetsstudier med de andra testen som använder sig av DISC-ansatsen. Om det är så att dessa test mäter samma sak borde resultaten korrelera högt med varandra. Istället har SLG valt att korrelera PPA med 16PF och OPQ (STP, 2004).

### *Experimentets syften och arbetshypoteser*

Vårt experiment hade två syften. Dels att mäta om testanvändarna var bättre än slumpen på att bedöma om testet de hade framför sig tillhörde personen som fick det återkopplat. Hälften av testerna roterades inom gruppen så att fem personer per användare fick fel test återkopplat. Liknande studier har vi inte kunnat finna i våra efterforskningar. Testanvändarnas förutsättningar har inte varit optimala. Se vidare information under *Avgränsningar*. Det är dock rimligt att anta att testanvändarna har mer än 50% rätt i sina bedömningar då de använde sig av ett personlighetsinventerande instrument i kombination med sina yrkeserfarenheter. Följande arbetshypotes för denna del (A) av experimentet har alltså förelagat:

$H_{1A}$ : Testanvändarnas förmåga att avgöra försökspersonernas gruppstillhörighet är statistiskt signifikant bättre än slumpen.

Emellertid går studiens andra syfte (B) i Gordons och Trankells fotspår. Försökspersonerna var inte informerade om att möjligheten fanns att de skulle få någon annans test återkopplat.

Efter återkopplingen mättes hur väl de tyckte att återkopplad profil överensstämde med deras självbild och efter att de informerats om experimentets syfte fick de avgöra om de fått rätt eller fel test återkopplat. Således var andra syftet med experimentet att mäta i vilken mån en duperingseffekt kunde skönjas och om försökspersonerna var kapabla att avgöra om deras testresultat verkade rimliga. De studier som presenterats i introduktionen talar för att det föreligger en duperingseffekt när resultat av personlighetstest återkopplas till försökspersonen och således har följande arbetshypotes utformats:

H<sub>1B</sub>: Det föreligger en statistiskt signifikant skillnad i vilken utsträckning grupperna 1 och 2 upplever att de fått rätt test återkopplat.

## Metod

För att kunna testa hypoteserna valdes en kvantitativ ansats. Som nämnts i det inledande kapitlet fann vi inte tidigare forskning kring testanvändares träffsäkerhet varför studien/experimentet också bör anses ha ett explorativt inslag. Nedan kommer termen *experiment* användas istället för *studie* då effekten av en medveten manipulation studerades och inte en naturligt förekommande företeelse. Att pröva testanvändarnas träffsäkerhet utan att kunna jämföra med tidigare forskning får också anses vara ett experiment.

### Försökspersoner

För att kunna genomföra experimentet gjordes bekvämlighetsurval. Frivilliga eftersöktes i experimentledarnas bekantskapskretsar och på studenthemmen i Vänersborg. Inga kriterier för deltagande ställdes upp förutom att ha möjlighet att avsätta ca en timmas tid för återkoppling av Thomas PPA vid tillfälle. Som framgår av Tabell 1 blev könsfördelningen totalt sett jämn medan övervägande majoriteten av förklarliga skäl kom att utgöras av högskolestudenter med varierande utbildningsinriktningar. Det bör dock påpekas att samtliga deltagare hade någon form av arbetslivserfarenhet. Ingen av deltagarna hade tidigare genomfört Thomas PPA.

Tabell 1

#### Fördelning av försökspersoner

	Totalt		30	
	Göteborg	Vänersborg		Andel av totalt antal försökspersoner
Antal	10	20		
Yrkesarbetande	5	2		23 %
Högskolestudenter	5	18		77 %
Kvinnor	4	11		50 %
Män	6	9		50 %
	Åldersfördelning			Andel av totalt antal försökspersoner
	Göteborg	Vänersborg		
18-20 år	0	2		6,7 %
21-25 år	4	15		63,3 %
26-30 år	5	2		23,3 %
31-35 år	1	0		3,3 %
36-40 år	0	1		3,3 %

### *Testanvändare*

Även de tre testanvändarna framskaffades genom bekvämlighetsurval. Användarna var certifierade på Thomas PPA och nyttjade instrumentet som samtalsunderlag i urvals- och utvecklingssammang i sitt nuvarande yrke. De hade alla olika lång erfarenhet av att använda testet men ingen av dem kunde anses vara nybörjare. De informerades om experimentets syfte och tyckte det skulle bli spännande att delta och få sätta sina färdigheter och Thomas PPA på prov.

### *Instrument*

”Penna-och-papperversjonen” på svenska av Thomas PPA ifylldes av samtliga försökspersoner. Efter återkopplingarna av testerna fick deltagarna en enkät med frågan: ”Hur väl anser du att den profil du fick presenterad för dig överensstämmer med din självbild?” Svaret kryssades i på en skala från 0-10 där 0 stod för ”inte alls”, 5 för ”varken eller” och 10 för ”helt och hållet”.

### *Tillvägagångssätt*

När vi var på jakt efter försökspersoner kunde vi givetvis inte berätta vad experimentet gick ut på men informerade deltagarna om att de skulle få göra ett personlighetstest som var ett av de tre största på svenska marknaden och sedan uttala sig om hur väl de tyckte det stämde in på deras självbild.

Deltagarna skulle fylla i testet och sedan vid ett senare tillfälle få det återkopplat av en certifierad testanvändare. Testerna gjordes i enlighet med SLG:s instruktioner. Dock genomfördes vissa testningar med flera personer i samma rum medan andra gjorde testerna på sina studentrum. Experimentledarna var dock noggranna med att hålla rekommenderade tidsramar och tillse att formulären fylldes i under tysthet. Testerna genomfördes i april och maj 2005 på studenthemmen i Vänersborg och på universitetsbiblioteket i Göteborg.

Efter samtliga tester var ifyllda, räknade experimentledarna ut respektive individs poäng på de olika dimensionerna (se teoriavsnittet) och ritade upp de profiler som hör till Thomas PPA och finns på baksidan av testet. Detta gjordes inte när försökspersonerna var närvarande. Varje test och således indirekt också varje försöksperson tilldelades en siffra för identifiering. Namn och siffra matades in i programmet Excel. Med hjälp av programmets slumpgenerator fick deltagarna ytterligare en siffra, ett eller två. Ett innebar att personen ifråga vid återkopplingstillfället skulle få sitt rätta test återkopplat medan siffran två således innebar att någon annans test skulle återkopplas till aktuell försöksperson.

För att hålla ordning på testnumren i Excel och för att inte göra användarna misstänksamma, användes den ursprungliga identifieringssiffran som markering på de profiler testanvändarna erhöll av oss. En separat kolumn i Excel nyttjades och döptes till ”erhållet test” där det stod vems test försökspersonen blivit tilldelad och vilket nummer det hade. Före slumpning skedde togs hänsyn till bostadsort av praktiska skäl. Således kom de tjugo personer som hade sin hemvist i Vänersborg respektive de tio boende i Göteborg att slumpas inom separata grupper.

När gruppstillhörigheten ”rätt eller fel test” väl var fastställd kontaktades testanvändarna för att boka datum och tid för återkopplingarna. Även här fick hänsyn tas till vad som kunde genomföras praktiskt. Försökspersonerna slumpades alltså inte till testanvändarna utan istället togs kontakt med den person som stod näst på tur i listan i Excel, förutsatt att denne hade rätt gruppstillhörighet (rätt eller fel test) så att användaren fick återkoppla fem rätt och fem fel test. För att öka testanvändarnas chanser att gissa rätt på om det var ”testets ägare” som fick det återkopplat roterades testen för de fem som fått ”gruppstillhörighet två” inom gruppen.



Testanvändarna behövde inte lämna svar till oss vilka de trodde hade fått rätt eller fel test återkopplat förrän de genomfört samtliga tio återkopplingar och således fanns möjlighet att känna igen en tveksam profil i någon av de andra försökspersonerna. Det skall tilläggas att användarna inte var medvetna om hur stor andel felaktiga återkopplingar som skulle ske och inte heller att de var roterade inom den egna gruppen.

Samtliga återkopplingar genomfördes under maj månad i för syftet lånade rum på ett av studenthemmen i Vänersborg och i ett samtalsrum på universitetsbiblioteket i Göteborg. Återkopplingarna gick till på följande sätt: Försökspersonen träffade testanvändaren i lånad lokal och fick i korthet Thomas PPA presenterat för sig, för att sedan få sin testprofil återkopplad. Testanvändaren berättade således vad Thomas PPA visade och förde kompletterande resonemang och ställde frågor efter eget huvud.

Således använde testanvändarna Thomas PPA på samma sätt i detta sammanhang som de var vana att göra i nuvarande yrke, det vill säga som samtalsunderlag kring personligheten. Försökspersonerna var fria att ställa frågor. Tiden för återkopplingarna varierade beroende på användare från en halv till en timma. Det lades ingen vikt vid att standardisera återkopplingarna utan lät tillvägagångssättet vara upp till användarna. Exempelvis nyttjade en av testanvändarna ett datorprogram som ingår i Thomas-systemet, för att automatiskt generera en slags berättelse om personen ifråga genom att mata in dennes värden på de olika dimensionerna. Syftet med experimentet var emellertid inte att utröna vilken återkopplingsmetod som var effektivast varför hänsyn till användarnas enskilda arbetsgångar inte tagits.

Efter varje återkoppling fick försökspersonen tidigare nämnd enkät av användaren och ombads att fylla i den och lämna till experimentledaren. Vid något tillfälle glömde testanvändaren säga åt försökspersonen att fylla i enkäten i enskildhet vilket kan ha påverkat svaret. Efter försökspersonen gått var användaren fri att göra noteringar kring intrycket av personen och om profilen verkade stämma in. Således tog användarna inte det slutliga beslutet om vilka personer de trodde fått fel test återkopplat förrän alla tio återkopplingarna var genomförda. Samtliga användare valde att genomföra sina tio återkopplingar uppdelat på två tillfällen/dagar. Vid sista tillfället fick användarna en stund på sig att fatta sina beslut kring vilka som fått rätt och fel test återkopplat och överlämnade sedan sina svar till oss och fick samtidigt rätt svar av oss.

När alla trettio återkopplingar var genomförda togs telefonkontakt med försökspersonerna för att berätta om experimentets egentliga syfte. Vid tillfället ställdes också frågan om de nu när de var medvetna om syftet trodde att de fått rätt profil återkopplad. Huruvida rätt profil återkopplats berättades sedan och återkoppling av rätt profil erbjöds de berörda.

### *Statistisk analys*

För att bearbeta den insamlade datan användes SPSS for Windows version 12.0.1. Ett flertal binomialfördelningstabeller användes i studien och ligger med som bilaga 1. Binomialfördelningstabellerna räknades fram i Excel.

### *Avgränsningar*

I denna studie har inte hänsyn kunnat tagas till de eventuella likheter som kan ha uppstått bland de test som återkopplades till fel person. Det kan vara så att något eller några av de test som byttes ut och återkopplades till fel person var väldigt snarlikt det riktiga resultatet.

## Resultat

Experimentets två huvudsyften har varit att undersöka om återkopplaren är kapabel att bedöma om testresultatet verkar rimligt eller ej och undersöka huruvida de personer som personlighetstestas är kapabla att bedöma om deras testresultat verkar rimliga eller ej och om.

### *Testanvändarnas skattningsförmåga*

De tre testanvändarna gjorde 10 återkopplingar var och skulle avgöra om det var rätt eller fel test som återkopplades. Eftersom fördelningen är diskret (endast alternativen stämmer/stämmer ej) och sannolikheten för att gissa rätt i varje situation är fast (50%) och oberoende av föregående gissning är detta en så kallad binomialfördelning. Beräkningar för sannolikheten att få det antal rätt som erhöles beräknades dels för varje testanvändare separat men resultaten slogs även ihop för att undersöka om det blev någon signifikansskillnad. Se bilaga 1 för sannolikhetstabeller för binomialfördelningen med  $n=10$ ,  $n=30$ .

Konfidensintervallen är markerade i båda tabellerna för att göra det enklare för läsaren.

Testanvändare 1 fick 7 rätt, testanvändare 2 fick 5 rätt och testanvändare 3 fick 7 rätt. Totalt ger detta 19 rätt av 30. Samtliga resultat faller inom konfidensintervallet då det för 10 försök ger ett konfidensintervall på 3-7 och för 30 ett intervall på 11-19. Se även tabell 2 för en sammanfattning av dessa data.

Tabell 2

*Sammanfattning av testanvändarnas träffsäkerhet tillsammans med en totaluträkning*

	Erhållet antal rätt	Förväntat antal rätt	p-värde
Återkopplare 1	7	5	0,12
Återkopplare 2	5	5	0,25
Återkopplare 3	7	5	0,12
Totalresultat	19	15	0,051

### *Försökspersonernas skattningsförmåga*

Försökspersonerna var uppdelade i två grupper, grupp 1 fick rätt test återkopplat och grupp 2 fick fel test återkopplat. Skattningarna från försökspersonerna kan också behandlas som en binomialfördelning där grupp 1 har gjort 15 oberoende försök och grupp 2 har gjort 15 oberoende försök. Precis som med återkopplarna så är sannolikheten för att gissa rätt 50 %. Antal personer som skattade att det var rätt test som återkopplades i grupp 1 var 14 stycken, en person skattade att det var fel test ( $p: 0,00$ ). I grupp 2 skattade 11 stycken att det var rätt test som återkopplades, 4 stycken skattade att det var fel test ( $p: 0,04$ ). Ett chi2-test utfördes också för att avgöra om skillnaden i skattning mellan grupperna var statistiskt signifikant. Resultaten från chi2-testet ( $fg=1$ ,  $chi2=2,16$ ,  $ns$ ) visade att skillnaden mellan grupperna inte var statistiskt signifikant. En övervägande majoritet har alltså skattat att det test som återkopplades till dem var det rätta, oberoende av om så var fallet eller inte.

Innan försökspersonerna fick ta ställning till om det var deras riktiga test fick de skatta hur väl de upplevde att resultatet stämde in med deras självbild på en skala från 0-10.

Medelvärdet i grupp 1 låg på 8,6 med en standardavvikelse på 0,8 medan medelvärdet i grupp 2 låg på 7,3 med en standardavvikelse på 2,8. Ett t-test utfördes för att utröna om det fanns en statistiskt signifikant skillnad mellan de två gruppernas skattningar. Så var inte fallet ( $t(28)=1,76$ ,  $ns$ ).

Effektstorleken för gruppernas självskattningar har också räknats ut och presenteras i resultattabellen. Effektstorlek är oberoende av stickprovsstorlek och kan därför användas för att jämföra resultaten mellan studier av olika storlekar. Formeln för beräkning av effektstorlek hämtas från Clark-Carter (1997).

Clark-Carter (1997) anger normer för hur effektstorlek skall bedömas. 0,2 representerar en liten effektstorlek, 0,5 en medelstor effektstorlek och 0,8 en stor effektstorlek. Resultatet av beräkningen ger en effektstorlek på 0,5. Detta innebär alltså att testbytet har en medelstor effekt på självskattningen. I tabell 3 nedan sammanfattas dessa resultat.

Tabell 3  
*Sammanfattning av gruppernas skattningar*

	Grupp 1	Grupp 2
Skattade att det var rätt test	14	11
Skattade att det var fel test	1	4
p-värde (binomialfördelning)	0,00	0,04
Medelvärde för självbildsskattning	8,6	7,3
Standardavvikelse för självbildsskattning	0,8	2,8
Effektstorlek på självskattningen	0,5	

### Diskussion

Det har, som nämnts i inledningen, riktats kritik mot den ibland magra utbildning som krävs för att vara behörig att använda vissa personlighetstest. Mot denna bakgrund gjordes den explorativa delen av experimentet eftersom att vi inte lyckades finna tidigare studier på området. Resultatet går inte att generalisera men kan förhoppningsvis inspirera till nya och större studier på området. Efter analys av insamlad data gick det inte att förkasta nollhypotesen ( $H_{0A}$ ). Således visade inte resultatet att testanvändarna var bättre än slumpen på att uttröna om försökspersonerna fick rätt eller fel test återkopplat.

De mindre fördelaktiga förutsättningarna som anges i avgränsningarna torde ha spelat in men samtidigt talar övriga förutsättningar i testanvändarnas favör. Vad som här avses är att testanvändarna fick se profilerna i god tid innan återkopplingarna skulle ske. Tidsåtgången för återkopplingarna var dessutom upp till användaren och de var fria att genomföra proceduren på det sätt de behagade. Exempel på detta är att en av testanvändarna valde att ta ut en datorgenererad berättelse om var och en av sina tio försökspersoner baserad på värdena i profilerna. Dessutom var sannolikheten att gissa rätt 50 % (rätt eller fel). Dock kände inte testanvändarna till hur stor andel av försökspersonerna som skulle få fel test återkopplat för att minimera risken för taktiskt handlande. Givetvis hade det varit intressant att intervjua testanvändarna kring hur de resonerade kring andelen rätt och fel test.

Testanvändarna blev som bekant satta på prov och det vore därför rimligt att tro att någon form av tankar kring hur många rätt respektive fel återkopplingar som genomförts uppenbarade sig. Beskedet om vilka återkopplingar som var gjorts till rätt respektive fel person behövde ju inte avlämnas förrän samtliga återkopplingar var genomförda och således verkar det sannolikt att testanvändarna frågade sig om antalet i respektive kategori var rimligt.

Vad som avses är skillnaden mellan att direkt efter en återkoppling avgöra gruppstillhörighet och att få avlämna besked när samtliga återkopplingar är genomförda. Det är dock inte säkert att det varit till fördel för testanvändarna att få avlämna alla svar på en gång då numeriska tankar kan ha gjort bedömningarna grumligare. Intressant hade därför varit att låta testanvändare återkoppla ett okänt antal test och avlämna svar efter varje session.

Det bör också påpekas att då testanvändarna kände till experimentets syften och kände till möjligheten av att det mycket väl kunde vara fel test som presenterats till dem. Detta bör ha gjort dem mer kritiska och uppmärksamma i sina granskningar och intervjuer. I den vanliga rekryteringssituationen finns bara möjligheten att testet gett fel utslag eller att testpersonen försökt förstå sig och det är inte säkert att de har lärt sig att kritiskt jämföra testresultaten med respondenten på samma sätt som de har gjort under experimentet.

De i inledningen nämnda studierna om Barnumeffekten talar för att det är relativt enkelt att dupera människor med personlighetstest. Resultatet från experimentet stöder denna tes. Med den data som samlats in går det inte att förkasta nollhypotesen ( $H_{0B}$ ) om att ingen skillnad mellan grupperna 1 och 2 föreligger vad gäller bedömning om huruvida man fått rätt eller fel test återkopplat till sig. Detta trots att experimentets syfte vid tidpunkten var känt och eventuella tveksamheter vid återkopplingstillfället kunde ha fått sin förklaring.

När dessa resultat behandlas som en binomialfördelning visar resultaten också att träffsäkerheten för grupp 1 och grupp 2 är signifikanta där grupp 1 träffsäkerhet är markant bättre än det förväntade värdet och grupp 2 träffsäkerhet är markant sämre. Det var endast fem personer av 30 som misstänkte att något inte stod rätt till. Fyra av dessa hade helt rätt i sina misstankar och lyckades upptäcka att de faktiskt fått fel test återkopplat till sig medan en person upplevde att hennes riktiga test var fel.

Samma tendens syns även vid skattning av överensstämmelsen mellan självbild och återkopplad profil. Dock talar den större standardavvikelsen i grupp 2 (fel test) för att vissa försökspersoner ställde sig kritiska och inte accepterade återkopplarens bild som spegling av den egna personligheten. Även om resultatet av denna studie inte visat på statistiskt signifikant skillnad mellan grupperna och således ger stöd åt tidigare liknande undersökningar, bör det understrykas att vissa skillnader kan skönjas.

Samplet innehöll endast 30 personer och ett större experiment hade kanske genererat utslag på de skillnader mellan grupperna som ändå gjorde sig gällande. Till sammanhanget hör också experimentets avgränsningar som i praktiken skulle kunna innebära att en felaktig men snarlik profil kan ha återkopplats till försökspersonen varför denne inte ställt sig kritisk till utlåtandet. Trots att experimentet inte kan anses vara optimalt utfört pekar dock resultaten på en duperingseffekt i linje med de tidigare studier som gjorts på Barnumeffekten.

Tämligen höga medelvärden för överensstämmelsen mellan självbild och presenterad profil har erhållits vilket kan tolkas som att de övervägande positiva omdömen återkopplaren vid ett dylikt sammanhang presenterar är något som försökspersonen vill ta till sig och anse utmärkande för den egna personen. Övervägande majoritet av försökspersonerna var i åldern 21-25 år och givetvis hade det varit intressant att undersöka om åldern i sig spelar någon roll för hur kritisk en försöksperson är i ett liknande experiment. Således ska inte möjligheten uteslutas att det identitetsskapande som studier, åldrande och yrkeserfarenhet utgör kunnat ge andra resultat.

Det går att tänka sig en mängd olika upplägg på den utförda studien som medför sina egna fördelar och nackdelar. I denna studie manipulerades inte testresultaten fick in då målet var att använda autentiska data i hela studien. Som det nämnts tidigare så medför detta att snarlika testresultat kan ha bytts ut vilket sätter experimentdeltagarna i en omöjlig situation eftersom resultatet då stämmer trots att ett byte har skett. Slump användes för att avgöra grupptillhörighet och rotation för testbyte för att minska risken för att detta skulle ske.

Av praktiska skäl fanns det tyvärr ingen möjlighet att göra dessa byten inom hela populationen vilket såklart hade varit att föredra. Detta var även till fördel för testanvändarna då de riktiga testen fanns med bland de personer som intervjuades. Eftersom alla valde att vänta med sina bedömningar till efter den sista intervjun så gick det rent teoretiskt att ”pussla” ihop det riktiga resultatet.

Om experimentet replikeras så är det helt klart önskvärt att dessa byten sker inom hela den population som får fel test återkopplat för att omöjliggöra eller i alla fall försvåra eventuell pussling. För att garanterat undvika att väldigt snarlika testresultat återkopplas kan man till skillnad från oss välja att manipulera de aktuella testresultaten. Manipulation av ipsativa data kan dock vara något omständligt eftersom förändring av en variabel leder till att även de andra variablerna påverkas.

Man kan antingen välja att bortse från detta och helt enkelt bara förändra en variabel men detta ger då orealistiska eller omöjliga testresultat vilket inte är önskvärt eller så kan man räkna ut hur en ny profil kan se ut om en variabel manipuleras. Detta kan bli ett problem eftersom det i praktiken inte går att manipulera endast en variabel i ett ipsativt test. Om ett normativt personlighetstest används uppstår inte detta dilemma vilket gör denna form av manipulation enklare och mindre problematiskt i sådana fall.

Det går även att styra bytet av test i den grupp som skall få fel test återkopplade. I detta fall skulle det då vara önskvärt att försäkra sig om att det är någorlunda stor skillnad mellan testet som skall återkopplas och originaltestet. I alla dessa fall när man väljer att manipulera data uppstår det problem med att de som utför experimentet kan anklagas för att på ett eller annat sätt styra resultatet genom sina manipulationer. Detta är en stor bidragande anledning till att slumpen styrde gruppstillhörighet och att vi även valde att använda oss av autentiska testdata för att resultaten som återkopplades skulle vara så realistiska som möjligt.

Det skulle vara intressant med nya studier på området. Exempelvis skulle det kunna röra sig om att replikera studien fast med ett annat personlighetstest som instrument. Dels är det självklart intressant att se om liknande resultat uppnås även med andra personlighetstest som återkopplingsverktyg och man skulle även kunna undersöka om det finns någon skillnad mellan ipsativa och normativa test på den här punkten.

Det är dock mindre intressant att replikera Barnumeffekten då den är så väletablerad och har studerats i en mängd olika rapporter. Mer intressant är att som Macdonald och Standing (2002) studera situationer där Barnumeffekten förstärks eller sätts ur spel. Med experimentets nuvarande upplägg är det dock en enkel procedur att samla in de data som behövs för att avgöra om Barnumeffekten påverkar testrespondenterna vilket leder till att det ändå är lämpligt att försöka replikera de tidigare studiernas resultat.

Tyvärr användes ett bekvämlighetsurval p g a tidsbegränsningar, det hade självklart varit önskvärt med ett ordentligt slumpmässigt urval. Studien skulle också kunna göras på enbart yrkesverksamma i något högre ålder än de nu deltagande försökspersonerna eller på beteendevetarstudenter som rimligtvis borde ställa sig mer kritiska än gemene man.

Det hade varit intressant att studera om det finns skillnad mellan olika yrkesgrupper av testanvändare. En möjlig hypotes hade kunnat vara att psykologer med relevant testutbildning bör vara markant bättre än slumpen på att avgöra om det är rätt test som återkopplas eller inte. Även de som är utbildade inom arbets- och organisationspsykologi kan kanske vara bättre utrustade än gemene man när det kommer till människobedömning av den här sorten.

Ett potentiellt problem med experimentet uppstår om det skall utföras i större skala med nuvarande upplägg som grund. Med nuvarande upplägg skulle det krävas att 300 personer personlighetstestades för att 30 stycken återkopplare skulle få varsin grupp på tio personer att göra återkopplingar till. Tio återkopplingar kan också vara något i underkant, det skulle vara önskvärt med fler återkopplingar per person för att resultaten skall bli säkrare. Om testanvändarna hade fått göra 20 återkopplingar istället för tio så hade återkopplare 1 och 3 fått statistiskt signifikanta resultat om deras träffsäkerhet hållit i sig. Det är dock omöjligt att utifrån dessa resultat säga om deras träffsäkerhet hade legat på samma nivå eller om den hade dragits mot 50 %.

Ett sätt att komma runt detta problem skulle kunna vara att filma intervjun och sedan låta de andra testanvändarna titta på intervjuerna på TV och försöka bedöma om det test som återkopplades i intervjun var den personens rätta test eller inte. På detta sätt kan varje testanvändare göra fler bedömningar utan att antalet försökspersoner behöver öka så dramatiskt. Det är såklart en fråga i sig om bedömning utifrån en filmad intervju kan likställas med bedömning utifrån en intervju man själv genomfört. I sådana fall skulle det vara önskvärt att standardisera intervjuformen för att bedömningsmaterialet skall bli någorlunda likvärdigt trots att det är olika personer som genomför intervjuerna.

I denna versionen av experiment lämnades metodval till testanvändaren vilket innebär att de kan ha använt olika intervjumetoder. Med hjälp av en videokamera hade testanvändarnas skattningar i experimentet kunnat ökas från tio per testanvändare till 30 utan att behöva ta in några nya försökspersoner. Dessa bedömningar kan då delas upp i två kategorier: bedömning av egna intervjuer och bedömning av videointervjuer.

Som det tidigare antytts skapar Barnumeffekten en svår situation för den person som utifrån ett personlighetstestsresultat försöker bedöma om resultaten är rimliga eller inte. Så länge materialet som återkopplas är positivt i någon mening är chansen väldigt stor att respondenten kommer att acceptera denna information. Det räcker ofta bara med att beskrivningen som ges är någorlunda vag och allmän för att denna beskrivning skall accepteras som en specifik beskrivning av den egna personligheten.

Det blir då väldigt svårt att bedöma vad som verkar rimligt och vad som inte stämmer om respondenten svarar ja på alltihop. Resultaten från experimentet pekar på att denna situation är så svår att testanvändarna inte lyckas bättre än om resultaten hade slumpats fram. Studiens begränsade storlek gör att det är vanskligt att generalisera utifrån dessa data men resultaten pekar ändå på att detta verkar vara en väldigt svår situation för den som vill veta om testresultatet verkligen stämmer.

Faktumet att testrespondenter tenderar att ge positiva utlåtanden oberoende av vilket resultat som presenteras till dem gör det problematiskt att mäta face validitet, i alla fall om begreppet skall ha någon form av diskriminerande funktion. För att face validitet skall kunna mätas ordentligt måste Barnumeffekten sättas ur spel eller försvagas på något sätt. Nevo (1985) föreslår en metod för att mäta face validitet men denna metod förutsätter mycket riktigt att respondenterna ger olika svar angående hur väl de tycker att testresultaten stämmer för att resultaten skall bli meningsfulla. De empiriska data som samlats in på testmetoden görs tyvärr bara med olika begåvningsstest, inte personlighetstest vilket gör det svårt att bedöma om metoden skulle ge meningsfulla resultat för personlighetstest.

I PPA-profilens fall används adjektiv som ett mått på olika personliga dimensioner vilket kanske inte är helt självklart. Det kan mycket väl vara så att en dominant (högt D) person väljer adjektivet "övertalande" före adjektivet "snäll" men denna koppling är inte självklar. Även om en person med högt D faktiskt väljer övertalande före snäll så behöver detta inte bero på den personliga egenskapen utan valet kan göras av helt andra anledningar.

Om det är så att det inte går förlita sig till de rimlighetsbedömningar som görs av testanvändarna och försökspersonerna blir vi tvungna att förlita oss till själva instrumentets träffsäkerhet och giltighet. Detta är inte en önskvärd situation då Michell (1997; 1999; 2000; 2003; 2004) visat att psykometrin har väldigt allvarliga problem att lösa innan den kan sälla sig till de andra vetenskaperna och frågan om vad som verkligen mäts med psykometriska test behöver besvaras på ett mer adekvat sätt än vad som gjorts hittills. Det skulle även behöva utföras forskning på de psykologiska personlighetsdragen för att avgöra om de verkligen har en kvantitativ struktur eller om denna intuition är felaktig.

Om det visar sig att personlighetsdragen inte har en kvantitativ struktur så är de normativa test som används idag inte ett meningsfullt sätt att mäta personlighetsdrag på. I sådana fall är den ipsativa mätmetoden att föredra trots de problem som finns med de ipsativa skalorna. Anledningen till detta är att den ipsativa mätmetoden inte nödvändigtvis förutsätter att de underliggande personlighetsdrag som mäts har en kvantitativ struktur. Det enda som förutsätts i detta fall är att personen har olika preferenser inom ett område som då kan mätas av ett psykologiskt test.

## Referenser

- Axelsson, M. Dalgren, M. & Lansinger, F. (2004). *Vad borgar för hög kvalitet i personlighetstest?* [Opublicerad B-uppsats i arbets- & organisationspsykologi], HTU: Institutionen för individ och samhälle, Göteborg.
- Boyce, T.E. & Geller, S.E. (2002). Using the Barnum effect to teach psychological research methods. *Teaching psychology*, Vol. 29 (4), 316-318.
- Clark-Carter, D. (1997). *Doing quantitative psychological research*. UK: Psychology press ltd.
- Cornwell, J & Dunlap, W. (1994). On the questionable soundness of factoring ipsative data: a response to Saville & Willson (1991). *Journal of occupational & organizational psychology*, Vol. 67 (2).
- Gordon, H. (1998). *Att använda test vid personbedömning*. 030305, [www.gordonconsulting.se](http://www.gordonconsulting.se).
- Gordon, H. (Ej känt). *Ljuger kameran? Betraktelser över användningen av psykologiska tester*. 030305, [www.gordonconsulting.se](http://www.gordonconsulting.se).
- Gordon, H. (Ej känt). *Att tolka karaktärer. Om konsten att dupera människor via personlighetstester*. 030305, [www.gordonconsulting.se](http://www.gordonconsulting.se).
- Friedman, H. & Schustack, M. (1999). *Personality, classic theories and modern research*. Boston: Allyn & Bacon.
- Hedlund, N. (2005). Så blir du en kritisk testköpare. *Personal & Ledarskap*, 2, 39-52.
- Johnson, C.E., Wood, R., Blinkhorn, S.F. (1988). Spuriouiser and spuriouiser: the use of ipsative personality tests. *Journal of occupational psychology*, 61, 153-162.
- Kline, P. (2000). *The handbook of psychological testing* (2:nd edition). London: Routledge.
- Mabon, H. (2004). *Arbetspsykologisk testning. Om urvalsmetoder i arbetslivet* (2:a upplagan). Stockholm: Psykologiförlaget AB.
- Macdonald, D.J. & Standing, L.G. (2002). *Social behaviour and personality*, 30 (6), 625-630.
- Martin, B.A., Bowen, C.C. & Hunt, S.T. (2002). How effective are people at faking on personality questionnaires? *Personality and individual differences*, Vol. 32 (2), 247-256.
- Marston, W. (1989). *Emotions of normal people*. Ormskirk: Lyster.
- Meade, A. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of occupational and organizational psychology*, 77, 531-552.



Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British journal of psychology*, 88, 355-383.

Michell, J. (1999). *Measurement in psychology : a critical history of a methodological concept*. Cambridge: University Press.

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory and psychology*, Vol. 10 (5): 639-667.

Michell, J. (2003). Pragmatism, positivism and the quantitative imperative, *Theory and psychology*, Vol. 13 (1): 45-52.

Michell, J. (2004). Item response models, pathological science and the shape of error. *Theory and psychology*, Vol. 14 (1): 121-129.

Nevo, B. (1985). Face validity revisited. *Journal of educational measurement*, Vol. 22 (4): 287-293

Trankell, A. (1961). *Magi och förnuft i människobedömning*. Stockholm: Bonnier.

Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of occupational psychology*, 64, 219-238.

SLG Thomas International. (2004). *Person Profil Analys manual*. Tryckort okänd.

Standing, L. & Keys, G. (1986). Computer assessment of personality: a demonstration of gullibility. *Social behaviour and personality*, 14 (2), 197-202.

STP, (2004). Granskningsrapport – Thomassystemet Personlig Profil Analys, Thomas PPA. *Stiftelsen för tillämpad psykologi*.

Bilaga 1

*Binomialfördelning vid diskret fördelning för n=10, n=15 och n=30*

n=10	P: 0,5	n=15	P: 0,5	n=30	P: 0,5
k	P (X=k)	k	P (X=k)	K	P (X=k)
0	0,000976563	0	3,05176E-05	0	9,31323E-10
1	0,009765625	1	0,000457764	1	2,79397E-08
2	0,043945313	2	0,003204346	2	4,05125E-07
3	0,1171875	3	0,013885498	3	3,78117E-06
4	0,205078125	4	0,041656494	4	2,55229E-05
5	0,24609375	5	0,091644287	5	0,000132719
6	0,205078125	6	0,152740479	6	0,000552996
7	0,1171875	7	0,196380615	7	0,001895986
8	0,043945313	8	0,196380615	8	0,005450961
9	0,009765625	9	0,152740479	9	0,013324572
10	0,000976563	10	0,091644287	10	0,027981601
		11	0,041656494	11	0,050875638
		12	0,013885498	12	0,080553093
		13	0,003204346	13	0,111535052
		14	0,000457764	14	0,13543542
		15	3,05176E-05	15	0,144464448
				16	0,13543542
				17	0,111535052
				18	0,080553093
				19	0,050875638
				20	0,027981601
				21	0,013324572
				22	0,005450961
				23	0,001895986
				24	0,000552996
				25	0,000132719
				26	2,55229E-05
				27	3,78117E-06
				28	4,05125E-07
				29	2,79397E-08
				30	9,31323E-10
Total:	1			Total:	1



## Vill du ställa upp i ett experiment?

Hej! Vi är två studenter på arbets- och organisationspsykologiska programmet som skriver C-uppsats nu. Inom ramen för uppsatsen kommer ett experiment utföras och vi är i stort behov av 30 stycken frivilliga deltagare.

Som deltagare kommer du att få göra ett personlighetstest. Resultatet av testet erhålls inte direkt efter du genomfört det utan istället kommer du att få det återkopplat av en certifierad testanvändare vid ett senare tillfälle.

Att skriva testet går fort och tar bara ca 10 min. Återkopplingen tar ca 30 min. *Det är av största vikt att du som genomför testet också kan ställa upp och få det återkopplat!* Sannolikt kommer testet att kunna genomföras på studenthemmet och återkopplingen i lokal på HTU Campus Vänersborg. Tyvärr kan ingen belöning för deltagande utgå förutom vår enorma tacksamhet!

Du som vill ställa upp skriver upp dig på listan och blir sedan kontaktad. Om du av någon anledning inte vill skylta med namn och kontaktuppgifter på listan går det bra att skicka e-post till nedanstående.

Tack!

//Magnus Axelsson & Marcus Dalgren  
Arbets- och organisationspsykologiska programmet 02

Kontakta oss? Frågor?

E-post: [magnus.axelsson@student.htu.se](mailto:magnus.axelsson@student.htu.se), [smurkas@hotmail.com](mailto:smurkas@hotmail.com)

## Enkät om överensstämmelse av återkopplad Thomas PPA-profil och deltagarens självbild

Deltagarens namn: \_\_\_\_\_

Datum: \_\_\_\_\_

Återkopplarens namn:

Hur väl anser du att den profil du fick presenterad för dig överensstämmer med din självbild?  
Markera med ett kryss.

0 1 2 3 4 5 6 7 8 9 10  
inte alls varken helt och hållet  
bra eller dåligt

Högskolan i Trollhättan / Uddevalla  
Institutionen för Individ och samhälle  
Box 1236  
462 28 Vänersborg  
Tel 0521-26 40 00 Fax 0521-26 40 99  
[www.htu.se](http://www.htu.se)