



PSYKOLOGISK METOD AB



Ökad testvaliditet genom korrektion
för skönmålning

Lennart Sjöberg

Forskningsrapport 2011: 1

2011-07-06

UPP-testet är godkänt av STP (Stiftelsen för Tillämpad
Psykologi, Sveriges Psykologförbund) enligt de
europeiska EFPA-kriterierna

Psykologisk Metod L Sjöberg AB arbetar med utveckling och användning av psykologiska test samt undersökningar av attityder och riskuppfattningar, andra psykologiska utredningar och tillämpad forskning.

Vår affärsidé är att bedriva arbetet i nära anslutning till den aktuella forskningen inom psykologin.

Skrifter utgivna av Psykologisk Metod AB

- Sjöberg, L. Bortom Big Five: Konstruktion och validering av ett personlighetstest. Rapport 2008:1.
- Sjöberg, L., & Möller, K. (2009). Sociala arbetsfunktioner och personlighet. Rapport 2009:1.
- Sjöberg, L. (2009). *UPP*-testet: Kriterierelaterad validitet. Rapport 2009:2.
- Sjöberg, L. (2009). *UPP*-testet: Korrektion för skönmålning. Rapport 2009:3.
- Sjöberg, L. (2010). *UPP*-testet: Tredje generationens personlighetstest. Rapport 2010: 1.
- Sjöberg, L. (2010). *UPP*-testet. Manual, reviderad version, februari 2010 .
- Sjöberg, L. (2010). *UPP*-testet: Mångfald och jämställdhet gynnas av korrektion för skönmålning. Rapport 2010: 2.
- Sjöberg, L. (2010). A third generation personality test. Rapport 2010:3.
- Sjöberg, L. (2010). Emotionell intelligens och social förmåga hos ungdomar. Rapport 2010:4.
- Sjöberg, L. (2010). Faktorstrukturen hos *UPP*-testet. Rapport 2010:5.
- Sjöberg, L. (2010). *UPP*-testet: Användarhandbok, april 2010.
- Sjöberg, L. (2010). Teknisk manual, april 2010.
- Sjöberg, L. (2010). *UPP*-testet och kundservice: Kriteriestudie. Rapport 2010:6.
- Sjöberg, L. (2010). *UPP/Screen*: Ett screeningtest för personlighet och begåvning. Rapport 2010:7.
- Sjöberg, L. (2010). Personlighetsdimensioners validitet i arbetslivet: teorier och empiri. Rapport 2010:8
- Nilheim, H. (2010). Internetplattformen för *UPP*-testet. Rapport 2010:9.
- Sjöberg, L. (2010). Prognos av riksdagsvalet 2010. Rapport 2010: 10.
- Sjöberg, L. (2010). Skönmålning på *UPP* hos chefskandidater. Rapport 2010: 11.
- Sjöberg, L. (2011). Ökad testvaliditet genom korrektion för skönmålning. Rapport 2011:1.

Ökad testvaliditet genom korrektion för skönmålning

Lennart Sjöberg

Sammanfattning

I denna studie undersöktes om skönmålning fungerar som en suppressorvariabel till relationen testskalor – kriterier. Detta är en teoretisk förutsättning för den metodik för att ”rensa” testskalorna från skönmålning som används inom *UPP*-testet. Suppressoreffekter föreligger när (a) skönmålning och testskalorna är positivt korrelerade, (b) testskalorna är positivt korrelerade med kriterierna, och (c) korrelationen mellan skönmålning och kriterierna är noll eller negativ. Om dessa villkor är uppfyllda bör korrektionen för skönmålning medföra ökad validitet (ökad korrelation mellan testskalor och kriterier) hos testet, och stöd erhålls för modellen för skönmålning och på den grundad metodik för korrektion. Data i denna studie var *UPP*-testningar med 166 chefer på mellannivå inom Stockholms Läns Landsting, samt s.k. 360-gradersbedömningar av dem, utförda av deras medarbetare, kollegor och överordnade chefer. Bedömningarna användes som kriterier för validering av testet. Resultaten visade, för alla tre grupperna av bedömare, att villkoren för suppressoreffekter uppfylldes och att validiteterna ökade efter korrektion för skönmålning. Validiteten hos *UPP*-testet mot chefers effektivitet var i denna studie omkring 0.5, efter korrektion för beskuren spridning i testvariablerna och mätfel i kriterierna. Annan forskning har i stället gett små sänkningar av validiteten med liknande metodik. Skälet till denna skillnad diskuteras och en fallstudie presenteras, där en person gått in för att aktivt ”lura” testet, men misslyckades fullständigt. Korrektionen för skönmålning fungerade.

Det är väl känt att skönmålning är ett stort problem vid alla former av personlighetstest som bygger på självrapporter. *UPP*-testet använder sig av särskilda skalor för att mäta tendensen att ge socialt önskvärda var, som används för att korrigera testvärdena så att de blir i görligaste mån fria från effekter av skönmålning¹. Dessa skalor är korrelerade med övriga testskalor i varierande grad. De korrigerade värdena är helt enkelt residualer i modeller för sambandet mellan testskalorna och skalorna för skönmålning; olika modeller för olika testskalor - allt beroende på det empiriska sambandet. Exempelvis är skalan för emotionell stabilitet ganska starkt korrelerad med social önskvärdhet, kreativitet knappast alls. Den här ansatsen har visat sig fungera i flera olika avseenden: man kan identifiera testade personer som "bluffat", fördelningarna i de korrigerade skalorna är mindre sneda och bättre differentierade än rådata, experimentella data stödjer förfarandet liksom skillnader mellan testningar i skarpa och oskarpa lägen. Dessutom blir det mera rättvisande och rättvisa jämförelser mellan män och kvinnor, mellan invandrare och testade av inhemsk härkomst.

Men en fråga återstår. Påverkas validiteten hos testet, dvs. sambandet med kriterier på arbetsresultat? I ett pågående projekt vid Karolinska Institutet har vi studerat sambandet mellan *UPP* och kriterier på chefsframgång, som insamlats via s.k. 360-gradersbedömningar. I studien medverkade 166 chefer på mellannivå inom SLL (Stockholms Läns Landsting). De tog *UPP*-testet och bedömdes av medarbetare, chefer och kolleger enligt den s.k. Farax-metoden för 360-gradersbedömningar.

Grundförutsättningen för ökad validitet genom korrektion för skönmålning är att måttet på skönmålning fungerar som en suppressorvariabel, dvs. en variabel som korrelerar svagt eller negativt med kriteriet men positivt med testet². Så ser våra data faktiskt ut. Se Tabell 1.

Tabell 1. Korrelationer mellan 360-gradersbedömningar och skönmålning.		
	Overt skönmålning	Kovert skönmålning
Struktur, bedömd av chef	-0.04	0.16
Relationer, bedömd av chef	-0.02	0.22
Förändring, bedömd av chef	-0.08	0.20
Struktur, bedömd av	-0.16	-0.17
Relationer, bedömd av kollegor	-0.17	-0.08
Förändring, bedömd av kollegor	-0.22	-0.02
Struktur, bedömd av underställda	-0.11	-0.03
Relationer, bedömd av underställda	-0.09	-0.00
Förändring, bedömd av underställda	-0.17	0.01

Enda undantaget är kovert skönmålning och bedömning av överordnad chef där korrelation är positiv (men svag). Uziel har sammanställt en mängd forskning som visar att skönmålning ("impression management") ofta samvarierar positivt med olika mått på social anpassning. Bedömningar av typ 360-gradersbedömningar är troligen bättre som kriterier av det skälet att de inte enbart är uttryck för ett globalt intryck, något som kanske kan manipuleras av dem som skönmålar i en testsituation. I en tidigare, mindre, validering av *UPP* fann vi just att globala bedömningar samvarierade positivt med skönmålning.

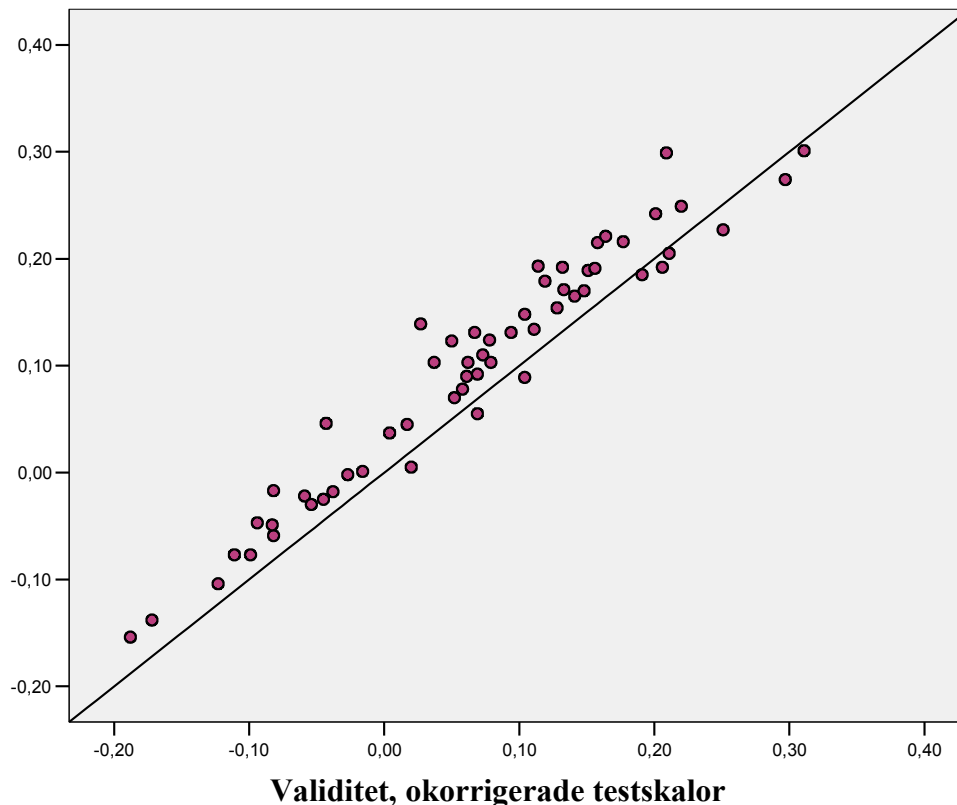
¹ Ca 90 % av effekten elimineras.

² En suppressor fungerar så att den s a s rensar bort en del av testskalornas varians som är utan värde för att predicera eller förklara kriteriet. Se http://en.wikipedia.org/wiki/Mediation_%28statistics%29

Eftersom den overta skönmålningen gav det tydligaste resultatet i Tabell 1 har enbart den skalan använts i det följande. Vi vet emellertid att användning av båda skalorna i praktiken ger ett resultat som ligger nära det man får med bara en skala.

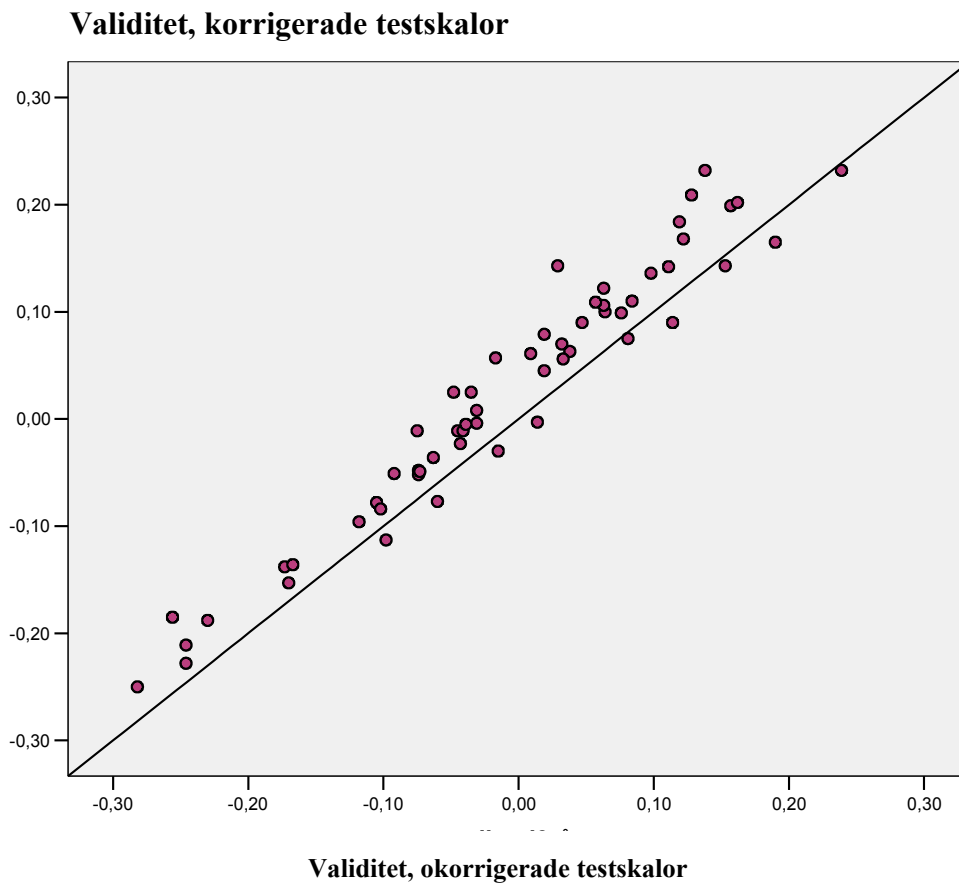
I Fig. 1 har jag plottat resultaten från analys av medarbetares bedömningar av chefernas effektivitet i skilda avseenden. På x-axeln validitet för varje skala och kriterium för okorrigerade data; på y-axeln motsvarande för korrigerade data. Varje punkt motsvarar en testskala (19 skalor) och 1 kriterieskala (3 skalor), alltså 57 punkter.

Validitet, korrigerade testskalor



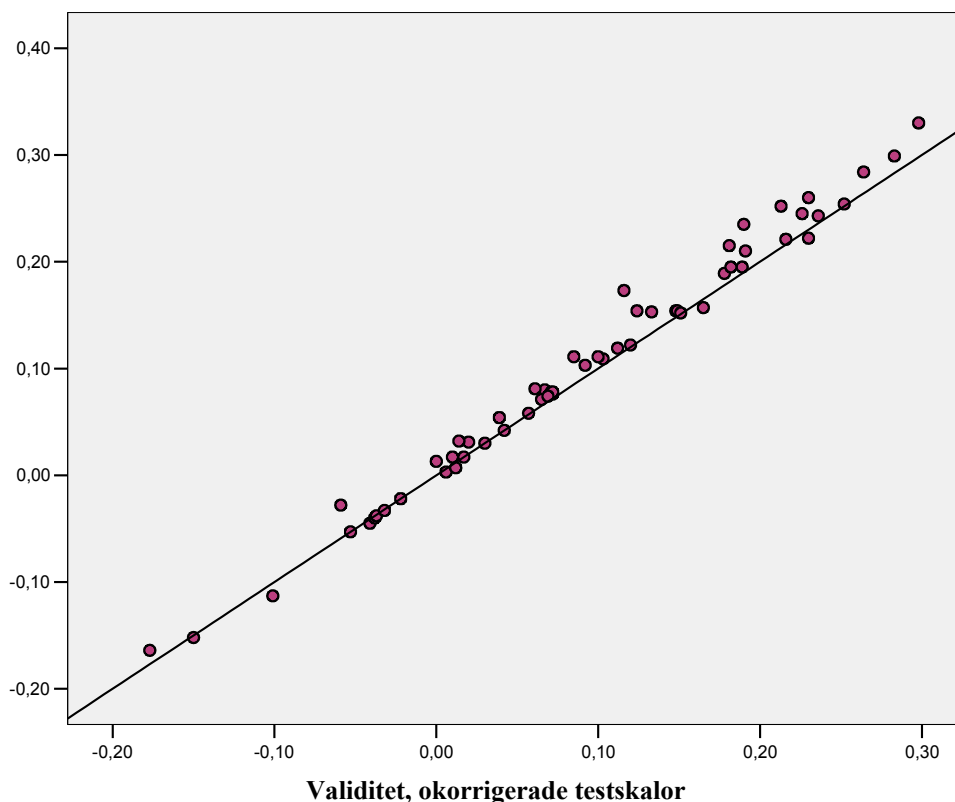
Figur 1. Validitet hos okorrigerade och korrigerade testskalor, medarbetares bedömningar som kriterium. Varje punkt motsvarar en testskala och en kriterieskala.

Validiteten ökade tydligt efter korrektion för skönmålning. Samma tendens kunde dokumenteras vid analyser av 360-gradersbedömningar utförda av kolleger och överordnade chefer. Korrektionen medförde en höjning av validiteten. Se Fig. 2 och 3.



Figur 2. Validitet hos okorrigerade och korrigerade testskalor, kollegers bedömningar som kriterium. Varje punkt motsvarar en testskala och en kriterieskala.

Validitet, korrigerade testskalor



Figur 3. Validitet hos okorrigerade och korrigerade testskalor, överordnade chefers bedömningar som kriterium. Varje punkt motsvarar en testskala och en kriterieskala.

Detta är sista pusselbiten i validering av korrektionsmetodik. Den fungerar, och det har nu dokumenterats i många avseenden. Eftersom det är ett mycket allvarligt problem för alla självrapporttest att somliga testade försöker bluffa sig till en tätposition på testet - är detta äntligen en lösning på problemet.

Slutligen ger jag resultaten av validering av *UPP* mot 360-gradersbedömningarna, se Tabell 2. Dessa data har korrigerats för beskuren spridning i testet och bristande reliabilitet i kriteriet enligt den "state-of-the-art"-metodik som nyligen publicerats, och som befunnits ge rättvisande resultat.

Tabell 2. Korrelationer mellan Faraxbedömningar (360 grader) och matchande <i>UPP</i> -skalor (validiteter) efter korrektion för mätfel i kriteriet och beskuren spridning i testvariablerna.			
Bedömare	Faraxbedömningar av		
	Relationer	Strukturering	Förändring
Överordnad chef	0.66	0.21	0.39
Kollegor	0.65	0.26	0.38
Underställda	0.66	0.40	0.51
Egenbedömning	0.52	0.37	0.64

Eftersom *UPP*-testets i dessa data når en validitet på ca 0.5 kan det hävdas att testet är nästan lika bra som kognitiva test för att predicera arbetsresultat. I andra studier har liknande resultat

erhållits . En kombination av *UPP* och ett test på g-faktorn är en idealisk grund för urval och rådgivning .

Det är väl känt att personlighetsdimensioner av typ ”Big Five” sällan eller aldrig når en högre validitet än 0.3 , och nästan alla moderna test av självrapporttyp är ”Big Five”-test. *UPP* har många unika fördelar, de viktigaste är kanske inriktningen på smala och fokuserade skalor i stället för de breda ”Big Five”, samt en validerad och väl fungerande metodik för att korrigera för skönmålning.

Resultaten i denna rapport avviker från en del annan forskning. En aktuell meta-analys av Li och Bagger visar på genomsnittliga, små, sänkningar av validiteten i stället för ökning som i denna rapport. Bakom detta resultat ligger positiva korrelationer mellan skönmålning och kriterier på ”performance” i arbetslivet. Ett skäl till skillnaden kan vara att de använde sig av andra skalor för att mäta skönmålning , men det är troligare att skillnaden beror på vilka kriterier som använts. Kriterier är ju ofta grundade på mer eller mindre globala bedömningar av chefer och förmän och det är möjligt att dessa låter sig påverkas av samma taktiska inställning som ligger bakom skönmålningen i testet. Fortsatt forskning är nödvändig för att pröva denna hypotes. Tills vidare är ju resultaten här helt klara och de talar, tillsammans med andra erfarenheter av vår teknik för korrektion för skönmålning, för att man får mera rättvisande resultat genom korrektion.

Betrakta följande fallstudie. En person som vi *vet*³ gick in för att ”lura” *UPP* fick skönmålningsvärdena 7 (overt) och 8 (kovert). Tabellerna 3-5 visar hans skalvärden före och efter korrektion, samt de absoluta måtten i procent av det maximalt möjliga i varje dimension. (Dessa tabeller är direkt tagna ur testrapporten).

Tabell 3. Personlighetsdimensioner av särskild vikt i arbetslivet			
	Normerat värde, korrigerat för skönmålning	Normerat värde, ej korrigerat för skönmålning	Genomsnittlig bedömning, procent av max-värde efter vändningar
Samarbetsvilja	5	7	92%
Kreativitet	7	8	90%
Emotionell intelligens	3	5	80%
Social förmåga	5	7	90%
Grad av positiv grundattityd	6	8	96%
Uthållighet	7	8	93%
Självförtroende	6	7	93%
Perfektionism	5	5	71%

³ Han svarade på en ”utmaning” som vi hade publicerat på Internet. Den som lyckades lura testet utlovades ”en skön morgonrock och två kaffemuggar”.

	Normerat värde, korrigerat för skönmålning	Normerat värde, ej korrigerat för skönmålning	Genomsnittlig bedömning, procent av max-värde efter vändningar
Utåtvändhet	6	7	90%
Emotionell stabilitet	6	8	90%
Öppenhet	8	8	91%
Följsamhet	4	6	80%
Noggrannhet	5	6	78%

	Normerat värde, korrigerat för skönmålning	Normerat värde, ej korrigerat för skönmålning	Genomsnittlig bedömning, procent av max-värde efter vändningar
Arbetsvilja	4	6	83%
Arbetsintresse	6	7	83%
Arbetsstillfredsställelse	5	5	93%
Resultatorientering	4	6	82%
Förändringsvilja	4	5	85%
Balans arbete - övrigt liv	5	6	84%

Tabellerna visar att det endast var två skalor som behölls sitt värde efter korrektion. I samtliga övriga fall sjönk skalvärdet, ibland så mycket som två steg på stanineskalan, vilket motsvarar en hel standardavvikelse. En sådan effekt är ”stor” enligt gängse statistisk terminologi .

Referenser

- [1]. Arvonen, J. (2002). *Change, production and employees. An integrated model of leadership*. Stockholm: Department of Psychology, University of Stockholm.
- [2]. Cohen, J. (1988). *Statistical power analysis for behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- [3]. de Colli, D. (2011). *Ett nytt svenskt arbetspsykologiskt test och arbetsprestation inom polisen – samtidig validitet*: Mälardalens högskola, Akademin för hållbar samhälls- och teknikutveckling.
- [4]. Li, A., & Bagger, J. (2006). Using the BIDR to Distinguish the Effects of Impression Management and Self-Deception on the Criterion Validity of Personality Measures: A Meta-Analysis. *International Journal of Selection and Assessment*, 14, 131-141.
- [5]. Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social-psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- [6]. 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.
- [7]. 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(4), 827-868.

- [8]. Sjöberg, L. (2009). *UPP-testet: Korrektion för skönmålning. (The UPP test: Correction for impression management)*. *Forskningsrapport 2009:3*. Stockholm: Psykologisk Metod AB.
- [9]. Sjöberg, L. (2009). *UPP-testet: Kriterierelaterad validitet. Forskningsrapport 2009:2*. Stockholm: Psykologisk Metod AB.
- [10]. Sjöberg, L. (2010). *Personlighetsdimensioners validitet i arbetslivet: teorier och empiri* (SSE/EFI Working Paper Series in Business Administration No. 2010:6). Stockholm: Stockholm School of Economics.
- [11]. Sjöberg, L. (2010). *Skönmålning på UPP hos chefskandidater*. Stockholm: Psykologisk Metod AB.
- [12]. Sjöberg, L. (2010). *UPP-testet och kundservice: Kriteriestudie. Forskningsrapport 2010:6*. Stockholm: Psykologisk Metod AB.
- [13]. Sjöberg, L. (2010). *UPP-testet. Teknisk manual*. Stockholm: Psykologisk Metod AB.
- [14]. Sjöberg, L. (2010). *UPP-testet: Mångfald gynnas av korrektion för skönmålning. Forskningsrapport 2010: 2*. Stockholm: Psykologisk Metod AB.
- [15]. Sjöberg, L. (2010). *UPP/Screen: Ett screeningstest för personlighet och begåvning*. Stockholm: Psykologisk Metod AB.
- [16]. Uziel, L. (2010). Rethinking social desirability scales: from impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5(3), 243-262.