



Félagslega æskileg svörun í persónuleikamælingum

Svarskekkja eða breytileiki milli einstaklinga?

Eva Dögg Steingrímsdóttir

Lokaverkefni til BS-gráðu
Sálfræðideild
Heilbrigðisvísindasvið



HÁSKÓLI ÍSLANDS

Félagslega æskileg svörun í persónuleikamælingum Svarskekja eða *breytileiki milli einstaklinga?*

Eva Dögg Steingrímsdóttir

Lokaverkefni til BS-gráðu í sálfræði
Leiðbeinendur: Fanney Þórssdóttir og Vaka Vésteinsdóttir

Sálfræðideild
Heilbrigðisvíndasvið Háskóla Íslands
Júní 2013

Ritgerð þessi er lokaverkefni til BS gráðu í sálfraði og er óheimilt að afrita ritgerðina á nokkurn hátt nema með leyfi rétthafa.

© Eva Dögg Steingrímsdóttir 2013

Prentun: Háskólaprent
Reykjavík, Ísland 2013

Formáli

Ritgerð þessi er lokaverkefni til BS gráðu í sálfræði og skiptist í two hluta. Í fyrri hlutanum verður fjallað um félagslega æskilega svörun (*socially desirable responding*) og áhrif hennar á réttmæti sjálfsmatslista. Sérstaklega verður fjallað um þann vanda sem svarskekkjan skapar í persónuleikamælingum. Greint verður frá aðferðum til þess að bregðast við þessum vanda með áherslu á kvarða sem greina og mæla félagslega æskilega svörun. Í því samhengi verður deilunni um hvort slíkir kvarðar mæli í raun svarskekkju eða breytileika milli einstaklinga gerð skil.

Seinni hluti verkefnisins er rannsóknargrein sem skrifuð er á ensku. Rannsóknin er framlag til lausnar á fyrrnefndri deilu og benda niðurstöður hennar til þess að ótímabært sé að afskrifa félagslegan æskileika sem vanda í persónuleikamælingum.

Efnisyfirlit

Inngangur	4
Aðferðir til þess að fást við félagslegan æskileika	6
Marlowe og Crowne kvarðinn	7
Félagslega æskileg svörun í persónuleikamælingum	8
Mat maka sem ytra viðmið	11
Svarskekkja eða breytileiki milli einstaklinga?	11
Heimildir.....	13
Abstract	19
Introduction	20
Method	24
Participants	24
Measures	24
Personality measure	24
Social desirability measure	26
Procedure	26
Results.....	26
Discussion	28
References	33

Sálfræðileg próf byggð á sjálfsmati eru víða notuð í bæði hagnýtum og fræðilegum tilgangi. Á þeim fjölda ára sem liðin eru síðan sjálfsmatslistar voru fyrst notaðir við mælingar á eiginleikum fólks hafa verið borin kennsl á ýmsar svarskekkjur (*response bias*) sem draga úr réttmæti slíkra lista. Svarskekkjur koma fram þegar atriðum er svarað á öðrum grundvelli en efnislegu inntaki (*content*) þeirra eða því sem atriðinu er ætlað að mæla (Paulhus, 1991). Þar á meðal eru samsinnishneigð (*acquiescence bias*) og handahófssvörum (*random responding*) en hér verður einkum fjallað um félagslegan æskileika (*social desirability*), þá svarskekkju sem mest hefur verið rannsokuð (Beretvas, Meyers og Leite, 2002).

Í öllum samfélögum eru viðmið (*norms*) um hvernig æskilegt sé að haga sér (Crawford og Ostrom, 1995). Flestir myndu til dæmis telja æskilegt að vera samviskusamur, heiðarlegur og stundvís. Þegar fólk metur eigin persónuleika, viðhorf eða hegðun gefur það stundum félagslega æskileg svör í stað þeirra sem í raun eiga við, annað hvort vísvitandi eða vegna þess að það hefur skekkt a mynd af sjálfu sér (Paulhus, 2002). Félagslegur æskileiki er því tilhneiting fólks til þess að haga samskiptum sínum við aðra þannig að það komi vel fyrir (Nederhof, 1985).

Paulhus (2002) gerir greinarmun á svörunarhætti (*response set*) og svarstíl (*response style*) í umfjöllun sinni um félagslegan æskileika. Svörunarháttur er tímabundin skekkja og kemur fram sem viðbragð við tilteknum atriðum eða aðstæðum. Fólk er þannig líklegra til þess að svara atriðum um viðkvæma hegðun á félagslega æskilegan hátt og sýna rannsóknir til dæmis að fólk segist drekka minna áfengi en það í raun gerir (Tourangeau og Yan, 2007). Auk þess er fólk líklegra til þess að veita félagslega æskileg svör í aðstæðum þar sem mikilvægt er að koma vel fyrir, eins og þegar persónuleikapróf er hluti af atvinnuumsókn (Christiansen, Goffin, Johnston og Rothstein, 1994), en ólíklegra til þess þegar svarað er undir

nafnleynd (Paulhus, 1984). Þegar skekkjan nær yfir mismunandi aðstæður, atriði og spurningalista er um svarstíl að ræða. Svarstíll er því hneigð sumra til þess að meta sjálfa sig á félagslega æskilegan hátt (Paulhus, 2002).

Þegar atriðum sjálfsmatslista er svarað á félagslega æskilegan hátt dregur úr réttmæti niðurstaðna, þar sem atriði listans mæla ekki einvörðungu það sem þeim er ætlað að mæla (Paulhus, 2002). Niðurstöður sjálfsmats eru oft notaðar við mikilvæga ákvarðanatöku og þá getur félagslega æskileg svörun haft slæmar afleiðingar. Notkun persónuleikaprófa við ráðningu starfsfólks hefur til dæmis verið gagnrýnd á þeim grundvelli að félagslega æskileg svörun dragi úr forspárréttmæti (*predictive validity*) slíkra prófa hvað varðar frammistöðu í starfi (Christiansen o.fl., 1994). Utan vinnumarkaðar eru niðurstöður sjálfsmats meðal annars notaðar við mat á geðrænum vanda og árangri meðferðar, í faraldsfræðilegum rannsóknum (*epidemiologic research*) og við val á nemendum til náms. Auk þess byggir opinber stefnumótun í málefnum eins og áfengis- og vímuþörnum að hluta til á niðurstöðum sjálfsmats (Murphy og Davidshofer, 2005).

Áhrif félagslegs æskileika á réttmæti sjálfsmatslista hefur lengi verið áhyggjuefni. Sem dæmi fullyrtu Meehl og Hathaway (1946) fyrir nær 70 árum að fá persónuleikapróf væru nothæf í klínískum tilgangi vegna þess hve auðvelt væri að veita röng svör. Síðan þá hafa niðurstöður fjölda rannsókna á félagslegum æskileika vakið efasemdir um réttmæti sjálfsmatslista (Bäckström, 2007; Edwards, 1953, 1957; Klassen, Hornstra og Anderson, 1975; Phillips og Clancy, 1970, 1972; Linehan og Nielsen, 1981, 1983), en við því hefur verið brugðist með ýmsum hætti.

Aðferðir til þess að fást við félagslegan æskileika

Þegar brugðist er við félagslegum æskileika eru ýmist notaðar aðferðir sem koma í veg fyrir eða draga úr félagslega æskilegri svörun, eða aðferðir sem greina skekkjuna og mæla (Paulhus, 1991). Dæmi um aðferðir sem falla í fyrri flokkinn er notkun hlutlausra atriða hvað varðar æskileika (*neutral items*) og að láta svarendur velja á milli tveggja atriða sem talin eru jafn eftirsóknarverð (*forced-choice items*). Einnig hefur fólki verið talin trú um að hægt sé að sjá á svörum þeirra hvort það segi satt frá (*the bogus pipeline*). Yfirleitt er ómögulegt að koma alfarið í veg fyrir félagslega æskilega svörun og því er oft brugðist við skekkjunni með notkun kvarða sem greina hana og mæla (Nederhof, 1985).

Mælingar á félagslegum æskileika eru oftast notaðar til þess að sýna fram á sundurgreinandi réttmætti (*discriminant validity*) sjálfsmatslista (Paulhus, 1991). Sá listi sem verið er að athuga er þá lagður fyrir ásamt mælingu á félagslegum æskileika og fylgni þeirra reiknuð. Lág fylgni bendir þá til þess að mælitækin tvö mæli ólíka eiginleika og niðurstöður sjálfsmatsins því ekki skekktar af félagslega æskilegum svörum. Sambærileg aðferð er að þáttagreina niðurstöður beggja mælitækja og athuga hvort atriði þeirra mæli aðgreinda þætti (Beretvas o.fl., 2002). Mælist áhrif félagslegs æskileika á niðurstöður er hægt að bregðast við með því að leiðréttu svör þeirra sem mælast hátt eða eyða þeim úr gagnasafninu (Nederhof, 1985).

Fjöldi kvarða hafa verið hannaðir til þess að mæla félagslega æskilega svörun og áhrif hennar á niðurstöður sjálfsmatslista. Þeir algengustu eru MMPI lygakvarðinn (Meehl og Hathaway, 1946), *Edwards Social Desirability Scale* (SD; Edwards, 1957), *Marlowe-Crowne Social Desirability Scale* (MCSD; Crowne og Marlowe, 1960) og *Balanced Inventory of Desirable Responding* (BIDR; Paulhus,

1984). Þar sem Marlowe og Crowne kvarðinn er sú mæling á félagslegum æskileika sem mest er notuð (Beretvas o.fl., 2002) mun umfjöllun beinast að honum.

Marlowe og Crowne kvarðinn

Kvarði Marlowe og Crowne samanstendur af 33 staðhæfingum um hversdagslega hegðun sem annað hvort er æskileg en óalgeng, eða óæskileg en algeng. Sá sem svarar gefur til kynna hversu vel hver staðhæfing eigi við um hann með því að velja „Satt“ eða „Ósatt“. Dæmi um staðhæfingu sem felur í sér óæskilega en algenga hegðun er „Það kemur fyrir að mér finnist gaman að slúðra“ og „Ég man eftir að hafa gert mér upp veikindi til að sleppa við eitthvað“. „Ósatt“ við þessum atriðum gefur til kynna sterkari tilhneigingu til þess að svara á félagslega æskilegan hátt en „Satt“. Dæmi um staðhæfingar sem fela í sér æskilega en óalgenga hegðun er „Áður en ég geng til kosninga kanna ég rækilega hæfni allra frambjóðenda“ og „Ég er alltaf kurteis, meira að segja við fólk sem mér líkar ekki við“. Á þessum atriðum gefur „Satt“ til kynna sterkari tilhneigingu til þess að svara á félagslega æskilegan hátt en „Ósatt“ (Crowne og Marlowe, 1960).

Marlowe og Crowne segja kvarðann mæla persónueinkenni sem þeir kalla „þörf fyrir samþykki“ (*need for approval*) og þeir sem svari á félagslega æskilegan hátt hafi þannig meiri þörf fyrir samþykki en þeir sem gera það ekki. Þar sem flestir haga sér ekki alltaf og í öllum kringumstæðum á félagslega æskilegan hátt, bendir hærri stigafjöldi á kvarðanum til meiri þarf fyrir samþykki (Crowne og Marlowe, 1960).

Rannsóknir hafa sýnt að Marlowe og Crowne kvarðinn hafi góða eða viðunandi próffræðilega eiginleika. Innri áreiðanleiki kvarðans hefur mælst á bilinu 0,72-0,96 (Crowne og Marlowe, 1960; Fischer og Fick, 1993; Loo og Thorpe, 2000;

Strahan og Gerbasi, 1972) og endurtektaráreiðanleiki (*test-retest*) 0,84 eftir eina viku (Fisher, 1967) og 0,89 eftir mánuð (Crowne og Marlowe, 1960). Innri áreiðanleiki styttri útgáfu (*short form*) hefur mælst á bilinu 0,68-0,94 fyrir 20 atriða kvarða, á bilinu 0,62-0,89 fyrir 13 atriða kvarða og á bilinu 0,42-0,88 fyrir tíu atriða kvarða (Ballard, 1992; Barger, 2002; Fischer og Fick, 1993; Loo og Thorpe, 2000; Reynolds, 1982). Rannsóknir á þýddum útgáfum kvarðans hafa ýmist sýnt að próffræðilegir eiginleikar þeirra séu sambærilegir upprunalegu útgáfunni (Collazo, 2005) eða aðeins slakari (Verardi o.fl., 2010).

Marlowe og Crowne kvarðinn hefur mikið verið notaður í rannsóknum og fylgni hefur mælst á milli hans og margvíslegra sjálfsmatsmælinga. Þannig hefur til dæmis mælst neikvæð fylgni við einkenni slæmrar andlegrar heilsu (Klassen o.fl., 1975) og áfengis- og eiturlyfjaneyslu (Welte og Russel, 1993) en jákvæð fylgni við lífsánægju (Carstensen og Cone, 1983). Fylgni kvarðans við aðrar mælingar á félagslegum æskileika, eins og SD og MMPI lygakvarðann, er veik eða í meðallagi (Crowne og Marlowe, 1960).

Félagslega æskileg svörun í persónuleikamælingum

Í persónuleikamælingum skapar félagslega æskileg svörun þann vanda að ekki er greint á milli þeirra sem raunverulega búa yfir tilteknu persónueinkenni og þeirra sem ýkja svör sín. Meehl og Hathaway (1946) voru með þeim fyrstu til þess að leggja áherslu á þróun árangursríkra aðferða til þess að bregðast við áhrifum félagslegs æskileika á réttmæti persónuleikaprófa. Þrátt fyrir að vandinn hafi þá þegar verið vel þekktur höfðu fáar tilraunir verið gerðar til þess að bregðast við honum (Edwards, 1953) og málið yfirleitt afgreitt með staðhæfingum um það að þátttakendur væru *taldir* svara af hreinskilni. Auk þess beindust þær aðferðir sem

voru í notkun, eins og MMPI lygakvarðinn, einkum að þeim sem vísvitandi gáfu röng svör (Meehl og Hathaway, 1946).

Meehl og Hathaway (1946) gerðu mælingu á tilhneigingu fólks til þess að gefa jákvæða mynd af sér á MMPI prófinu. Kvarðinn var kallaður „K þátturinn“ og var talinn mæla svarskekkju sem kæmi í veg fyrir réttmætar ályktanir af niðurstöðum prófsins. Fylgni K kvarðans við persónuleikamælingar var talin sýna að hve miklu leyti niðurstöður þeirra væru skekktar.

Edwards (1957) tók undir áhyggjur Meehl og Hathaway (1946) og hannaði kvarðann *Edwards Social Desirability Scale* (SD) til þess að mæla áhrif félagslegs æskileika á niðurstöður MMPI. Líkt og Meehl og Hathaway túlkaði Edwards fylgni SD kvarðans við persónuleikamælingar sem vísbendingu um óréttmæti prófsins. Fylgnirannsóknir sýndu að fjöldi persónuleikamælinga var undir áhrifum svartskekkjunnar (Edwards, 1957) og Jackson og Messick (1958) fullyrtu að líklega mældu MMPI og sambærileg próf að mestu svarstíl.

Margir töldu þannig ljóst að félagslegur æskileiki skapaði vanda í persónuleikamælingum sem bregðast þyrfti við á einn eða annan hátt (Cronbach, 1946; Edwards, 1953, 1957; Meehl og Hathaway, 1946; Wiggins, 1968). Það varð því að vana við notkun persónuleikaprófa að mæla einnig félagslegan æskileika og athuga samband mælinganna. Há fylgni var þá talin benda til þess að mælitækin væru að hluta til að mæla það sama, tilhneigingu fólks til þess að svara á félagslega æskilegan hátt. Persónuleikamælingin var þá álitin skekkt og fjöldi aðferða var fundinn upp til þess að fjarlægja þá dreifingu sem rakin var til skekkjunnar. Ætlað var að leiðréttar niðurstöður væru nær raunverulegum eiginleikum fólks, því sem prófinu var ætlað að mæla, en skekkjan skyggði á (Paulhus, 1991).

Þessu voru þó ekki allir sammála og Rorer (1965) fullyrti, þvert á móti því sem almennt var talið, að svarskekkjur og þar á meðal félagslegur æskileiki væru ekki vandamál í persónuleikamælingum. Taldi hann ekkert benda til þess að spurningalistum væri svarað á öðrum grundvelli en efnislegu inntaki þeirra. Raddir heyrðust sem töldu fylgni mælinga á félagslegum æskileika við persónuleikamælingar stafa af því að verið væri að mæla að hluta til það sama, tiltekin persónueinkenni. Kvarðar sem gerðir höfðu verið til þess að mæla félagslegan æskileika væru því ekki að mæla svarskekkju heldur raunverulegan breytileika milli einstaklinga. Þar af leiðandi væri rangt að túlka fylgni persónuleikaprófs við mælingu á félagslegum æskileika sem vísbendingu um óréttmæti niðurstaðna (Dicken, 1963; Furnham, 1986; Heilbrun, 1964; McCrae, 1986; McCrae og Costa, 1983; Rorer, 1965).

Flestir voru þó sammála um að ekki væri hægt að leysa deiluna um það hvað kvarðarnir mældu í raun með sjálfsmatsrannsóknum (Furnham, 1986). Vandinn við slíkar rannsóknir er að erfitt er að segja til um hvort þeir sem mælist hátt á félagslegum æskileika hafi ýkt svör sín eða séu einfaldlega óvenju samviskusamir, heiðarlegir og samvinnuþýðir (Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist og Verkasalo, 2007). Ein lausnin er að bera niðurstöður sjálfsmats saman við hlutlægt ytra viðmið (*external criteria*) og athuga hvaða áhrif það hefur á sambandið að leiðréttta fyrir félagslegum æskileika. Mæli kvarðarnir svarskekkju, sem er raunverulegum persónueinkennum óviðkomandi, ætti samræmið á milli sjálfsmats og ytra viðmiðs að aukast þegar villuþátturinn er fjarlægður (Smith og Ellingson, 2002).

Mat maka sem ytra viðmið

Í rannsókn, sem síðar átti eftir að hafa mikil áhrif á túlkun félagslegs æskileika, notuðu McCrae og Costa (1983) mat maka sem ytra viðmið. Fylgni var reiknuð á milli sjálfsmats og mats maka á NEO-PI-R prófinu, bæði fyrir og eftir að sjálfsmatið var leiðrétt fyrir félagslegum æskileika sem mældur var með Marlowe og Crowne kvarðanum. Niðurstæðan varð sú að leiðréttning skilaði ekki meira samræmi á milli sjálfsmats og mats maka, heldur stóð fylgnin ýmist í stað eða lækkaði. Þetta túlkuðu McCrae og Costa þannig að Marlowe og Crowne kvarðinn mældi raunverulegan breytileika milli einstaklinga en ekki svarskekkju. Fylgni kvarðans við persónuleikaprófið stafaði því af efnislegri skörun atriða þeirra og vafasamt væri að leiðréttta niðurstöður fyrir félagslegum æskileika.

McCrae og Costa (1983) gerðu ráð fyrir því að mat maka væri hlutlægt ytra viðmið en sýndu ekki fram á það á neinn hátt. Ef mat maka er skekkt af félagslegum æskileika, líkt og sjálfsmat, er hugsanlegt að gagnstæðar niðurstöður hefðu fengist ef mat maka hefði einnig verið leiðrétt fyrir félagslegum æskileika. Því er vafasamt að álykta út frá þessum niðurstöðum að félagslegur æskileiki sé ekki vandi í persónuleikamælingum. Nauðsynlegt er að útiloka aðrar jafn líklegar útskýringar, til dæmis þá að mat maka sé í raun ekki hlutlægt ytra viðmið.

Svarskekkja eða breytileiki milli einstaklinga?

Margir telja að niðurstöður McCrae og Costa (1983) hafi sýnt í eitt skipti fyrir öll að félagslegs æskileikakvarðar mæli raunveruleg persónueinkenni en ekki svarskekkju. Vísað hefur verið í greinina oftar en 500 sinnum og oft til stuðnings þeirri staðhæfingu að félagslega æskileg svörun sé ekki vandi í persónuleikamælingum (Borkenau og Ostendorf, 1989, 1992; Kurtz, Tarquini og Iobst, 2008; Nicholson og

Hogan, 1990; Ones, Viswesvaran og Reiss, 1996; Piedmont, McCrae, Riemann og Angleitner, 2000; Smith og Ellingson, 2002).

Þessi ályktun er, eins og fyrr var sagt, mjög vafasöm og ekki aðeins vegna þess að niðurstöður McCrae og Costa (1983) leyfi ekki slíkar ályktanir. Þó sýnt væri með óyggjandi hætti að þeir kvarðar sem gerðir hafa verið til þess að mæla félagslegan æskileika mæli persónueinkenni í stað svarskekkju, þá væri ekki þar með sagt að félagslegur æskileiki sé ekki vandi í persónuleikamælingum. Það að ekki hafi tekist að mæla áhrif skekkjunnar útilokar ekki tilvist hennar. Marlowe og Crowne kvarðinn hefur til dæmis verið nokkuð gagnrýndur (Barger, 2002) og margir telja nauðsynlegt að hanna betri mælingar á félagslegum æskileika (Schinka, Kinder og Kremer, 1997). Holden (2007) bendir á það að hugtakaréttmæti kvarðanna sé oft óviðunandi, sem eigi þátt í því að fjöldi rannsókna sýni að félagslegur æskileiki dragi ekki úr réttmæti persónuleikaprófa. Telur hann ótímabært að afskrifa félagslegan æskileika sem vanda í persónuleikamælingum og fleiri eru á sama máli (Bäckström, 2007; Ellingson, Sackett og Hough, 1999; Linehan og Nielsen, 1981, 1983; Paulhus, 1991, 2002; Zerbe og Paulhus, 1987).

Þrátt fyrir að sumir segi deiluna um félagslegan æskileika löngu leysta (Nevid, 1983; McCrae og Costa, 1983; McCrae o.fl., 1989; Ones o.fl., 1996), þá er enn margt sem nauðsynlegt er að skoða betur. Til dæmis þarf að endurtaka rannsókn McCrae og Costa (1983) og leiðréttta einnig fyrir félagslegum æskileika í mati maka. Fyrir um 70 árum kölluðu Meehl og Hathaway (1946) eftir áhrifaríkum aðferðum til þess að bregðast við félagslegum æskileika, sem þeir töldu áður hafa verið afgreiddan með innistæðulausum staðhæfingum. Það að vitna í McCrae og Costa (1983) og telja sig þannig hafa afgreitt félagslegan æskileika er ekki skref í rétta átt.

Heimildir

- Bäckström, M. (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment*, 23(2), 63-70.
- Ballard, R. (1992). Short forms of the Marlowe–Crowne Social Desirability Scale. *Psychological Reports*, 71, 1155–1160.
- Barger, S. D. (2002). The Marlowe-Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment*, 79(2), 286-305.
- Beretvas, S. N., Meyers, J. L., og Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570-589.
- Borkenau, P., og Ostendorf, F. (1989). Descriptive consistency and social desirability in self and peer reports. *European Journal of Personality*, 3, 31–45.
- Borkenau, P., og Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, 6(3), 199-214.
- Carstensen, L. L., og Cone, J. D. (1983). Social desirability and the measurement of psychological well-being in elderly persons. *Journal of Gerontology*, 38(6), 713-715.
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., og Rothstein, M. G. (1994). Correcting the Sixteen Personality Factors tests for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology*, 47(4), 847-860.

- Collazo, A. A. (2005). Translation of the Marlowe-Crowne Social Desirability Scale into an equivalent Spanish version. *Educational and Psychological Measurement*, 65(5), 780-806.
- Crawford, S. E., og Ostrom, E. (1995). A grammar of institutions. *American Political Science Review*, 89, 582-600.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475-494.
- Crowne, D. P., og Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354.
- Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement*, 23, 699-720.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37(2), 90-93.
- Edwards, A. L. (1957). The social desirability variable in personality assessment and research. New York: Dryden.
- Ellingson, J. E., Sackett, P. R., og Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84(2), 155-166.
- Fischer, D. G., og Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 53(2), 417-424.

- Fisher, G. (1967). Normative and reliability data for the standard and the cross validated Marlowe-Crowne Social Desirability Scale. *Psychological Reports*, 20(1), 174-174.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400.
- Heilbrun, A. B. (1964). Social-learning theory, social desirability, and the MMPI. *Psychological Bulletin*, 61(5), 377-387.
- Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science*, 39(3), 184-201.
- Jackson, D. N., og Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55(4), 243-252.
- Klassen, D., Hornstra, R. K., og Anderson, P. B. (1975). Influence of social desirability on symptom and mood reporting in a community survey. *Journal of Consulting and Clinical Psychology*, 43(4), 448-452.
- Kurtz, J. E., Tarquini, S. J., og Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences*, 45(1), 22-27.
- Linehan, M. M., og Nielsen, S. L. (1981). Assessment of suicide ideation and parasuicide: Hopelessness and social desirability. *Journal of Consulting and Clinical Psychology*, 49(5), 773-775.
- Linehan, M. M., og Nielsen, S. L. (1983). Social desirability: Its relevance to the measurement of hopelessness and suicidal behavior. *Journal of Consulting and Clinical Psychology*, 51, 141-143.

- Loo, R., og Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-Crowne Social Desirability Scale. *The Journal of Social Psychology*, 140(5), 628-635.
- Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., og Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality*, 75(2), 291-322.
- McCrae, R. R. (1986). Well-being scales do not measure social desirability. *Journal of Gerontology*, 41(3), 390-392.
- McCrae, R. R., og Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882- 888.
- McCrae, R. R., Costa, P. T., Dahlstrom, W. G., Barefoot, J. C., Siegler, I. C., og Williams, R. B. (1989). A caution on the use of the MMPI K-correction in research on psychosomatic medicine. *Psychosomatic Medicine*, 51(1), 58-65.
- Meehl, P. E., og Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30(5), 525-563.
- Murphy, K. R., og Davidshofer, C. O. (2005). *Psychological testing* (6. útgáfa). New Jersey: Prentice Hall.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263-280.
- Nevid, J. S. (1983). Hopelessness, social desirability, and construct validity. *Journal of Consulting and Clinical Psychology*, 51, 139-140.
- Nicholson, R. A., og Hogan, R. T. (1990). The construct validity of social desirability. *American Psychologist*, 45, 290-292.

Ones, D. S., Viswesvaran, C., og Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81(6), 660-679.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. Í J. P. Robinson, P. R. Shaver og L. S. Wrightsman (ritstjórar), *Measures of personality and social psychological attitudes* (bls. 17-59). San Diego: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. Í H. I. Braun, D. N. Jackson og D. E. Wiley (ritstjórar), *The role of constructs in psychological and educational measurement*, (bls. 49-69). New Jersey: Erlbaum.

Phillips, D. L., og Clancy, K. J. (1970). Response biases in field studies of mental illness. *American Sociological Review*, 35, 503-515.

Phillips, D. L., og Clancy, K. J. (1972). Some effects of “social desirability” in survey studies. *American Journal of Sociology*, 77, 921-940.

Piedmont, R. L., McCrae, R. R., Riemann, R., og Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78(3), 582-593.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 38(1), 119-125.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63(3), 129-156.

Schinka, J. A., Kinder, B. N., og Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment*, 68(1), 127-138.

Smith, D. B., og Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87(2), 211-219.

Strahan, R., og Gerbasi, K. C. (1972). Short homogenous version of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28(2), 191–193.

Tourangeau, R., og Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859-883.

Verardi, S., Dahourou, D., Ah-Kion, J., Bhowon, U., Tseung, C. N., Amoussou-Yeye, D., o.fl. (2010). Psychometric properties of the Marlowe Crowne Social Desirability Scale in eight African countries and Switzerland. *Journal of Cross-Cultural Psychology*, 41(1), 19-34.

Welte, J. W., og Russell, M. (1993). Influence of socially desirable responding in a study of stress and substance abuse. *Alcoholism: Clinical and Experimental Research*, 17(4), 758-761.

Wiggins, J. S. (1968). Personality structure. *Annual Review of Psychology*, 19(1), 293-350.

Zerbe, W. J., og Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, 12, 250-264.

Whether or not socially desirable responding (SDR) is a cause for concern in personality assessment has long been debated. For many researchers, McCrae and Costa (1983) laid the issue to rest when they showed that correcting for SDR in self-reports did not improve the agreement with spouse ratings on the NEO Personality Inventory. However, their findings rest on the assumption that observer ratings in general, and spouse ratings in particular, are an unbiased external criterion. If spouse ratings are also susceptible to SDR, correcting for the bias in *both* self-reports and spouse ratings might be necessary to enhance the validity of personality measures. In the present study, McCrae and Costa's influential study was replicated with the exception of correcting for SDR, measured with the Marlowe-Crowne Social Desirability Scale, in both self-reports and spouse ratings. Analyses were based on responses from 63 couples who had lived together for at least six months. The results showed that correcting for SDR in both self-reports and spouse ratings increased their agreement on most traits. Thus indicating, contrary to McCrae and Costa's conclusion, that SDR does affect the validity of personality measures and that SDR scales can be used to rectify the problem.

For half a century, there has been considerable debate on whether socially desirable responding (SDR) is a cause for concern in personality assessment (e.g. Holden, 2007; Nevid, 1983; Rorer, 1965). SDR, defined as the tendency to give overly positive self-descriptions (Paulhus, 2002), has traditionally been regarded as a response bias that poses a threat to the validity of personality scales (Cronbach, 1946; Edwards, 1953, 1957; Meehl & Hathaway, 1946). Under this assumption, an observed correlation between a personality scale and a scale designed to measure SDR indicates that the personality measures are confounded with error variance (Paulhus, 1991). Those who argue that SDR scales measure a substantive trait have challenged this interpretation and maintained that an observed correlation is the result of content overlap between the instruments. SDR scales should thus not be regarded as indicators of response bias (Furnham, 1986; McCrae & Costa, 1983; Nicholson & Hogan, 1990, Smith & Ellingson, 2002).

More specifically the debate centers on whether SDR is a response style or a substantive personality trait, and whether SDR variance should be removed from personality scales. Because self-report studies cannot disentangle substance from response style, one strategy to resolve the debate is the comparison of self-reports and external criteria (Furnham, 1986). If SDR scales are indicators of response style then the removal of variance associated with SDR should enhance the validity of self-reported personality traits on the account of removing extraneous variance. Correcting for SDR should therefore improve agreement between self-reports and an external criterion, showing that SDR acted as a suppressor variable (e.g. Ganster, Hennessey & Luthans, 1983).

In an influential study, McCrae & Costa (1983) used spouse ratings as an external criterion of personality traits in the domains of Neuroticism, Extraversion

and Openness to Experience (NEO). Correcting for SDR in self-reports with the Marlowe and Crowne Social Desirability Scale (MCSD; Crowne & Marlowe, 1960) did not improve the agreement with spouse ratings, and for many traits lowered the agreement. Therefore, the authors concluded that SDR scales should be given substantive rather than artifactual interpretations.

Because correcting for SDR failed to enhance validity in McCrae and Costa's study (1983), and numerous others that used observer ratings as external criteria (Borkenau & Ostendorf, 1992; Dicken, 1963; Kozma & Stones, 1987; Kurtz, Tarquini & Iobst, 2008; McCrae et al., 1989; Pauls & Stemmler, 2003; Piedmont, McCrae, Riemann & Angleitner, 2000), some researchers have concluded that SDR scales are not useful for increasing the validity of personality scales (e.g. Borkenau & Ostendorf, 1992) or what's more dubious, that SDR is not a cause for concern (e.g. Piedmont et al., 2000). As Ones, Viswesvaran and Reiss (1996) note, for many researchers the SDR issue is a "methodological dead horse" (Nevid, 1983, p. 139), laid to rest by McCrae and Costa (1983).

These conclusions are an overgeneralization from McCrae and Costa's (1983) results. The respondents were disinterested volunteers who answered anonymously and were thus not motivated to distort their answers. SDR might still be a problem in situations where motivation to distort is high, for example in personnel assessments, as McCrae & Costa (1983) fully admit but some researchers have ignored. In addition, a number of studies have shown that SDR degrades the construct validity of personality measure (e.g. Ellingson, Sackett & Hough, 1999; Holden, 2007; Viswesvaran & Ones, 1999). Topping and O'Gorman (1997), for example, reported lowered self-other agreement on the NEO-FFI scales under instructions to "fake good".

Previous objections notwithstanding, McCrae and Costa's (1983) findings rest on the assumption that observer ratings in general, and spouse ratings in particular, are an unbiased external criterion. It has been implicitly assumed that only self-reports are susceptible to SDR, but as Funder and Colvin (1997, p.625) suggest, the "self"-enhancement bias may be poorly named. Similar enhancement effects were found when acquaintances ratings were compared to ratings by strangers as when self-reports were compared to stranger ratings. These findings may be explained by the Self-Evaluation Maintenance Model (SEM) which assumes that peoples self-evaluations are partly determined by a process whereby the good qualities of close others are perceived as reflecting something about oneself (Tesser, Pilkington & McIntosh, 1989). McCrae and Costa (1983) recognize that spouse ratings may also contain some bias but claim them to be an unbiased external criterion. On the contrary to that assertion, Murray (1999) argues that seeing your partner in the most positive light contributes to the relationships quality. This process might be reflected in SDR by the spouse. It is thus highly possible that in order for SDR scales to improve agreement between self-reports and spouse ratings, variance associated with SDR has to be removed from *both* measures.

Although researchers have generally overlooked this possibility, it was recently tested in a study using peer-reports as an external criterion (Konstabel, Aavik & Allik, 2006). The Big-Five dimensions were measured with the NEO-PI-R (Costa & McCrae, 1992) and SDR with the Social Desirability Index (SDI). While correcting for SDR in self-reports lowered the agreement with peer-reports on all dimensions, correcting for SDR in both self-reports and peer-reports improved the agreement. The bipartial correlations were in all cases higher than the zero-order

correlations, showing the usefulness of the SDI in enhancing the validity of the NEO-PI-R.

Despite these results, researchers continue to cite McCrae and Costa (1983) as support for the conclusion that SDR is not an issue in personality assessment. This conclusion was recently reaffirmed in a study that once again overlooked the possibility that variance associated with SDR has to be removed from both observer ratings and self-reports (Kurtz et al., 2008). Why Konstabel et al's. (2006) findings have been disregarded in the SDR literature is unclear. One suggestion is that because the study used a student sample, McCrae and Costa's (1983) results are considered more relevant to the general population. In addition, Konstabel et al. (2006) state that the SDI is conceptually less dependent on the content of any personality scale than the SDR scales (p. 553). Thus, it could be argued that their findings do not apply to the more traditional measures of SDR, such as the Marlowe-Crowne Social Desirability Scale (MCSD), used in McCrae and Costa's study. Whatever the reason, it can be said that "conceptual habits die hard", as McCrae and Costa (1983, p. 887) stated regarding the interpretation of correlations with SDR scales as evidence of the invalidity of personality measures.

Because of the great influence McCrae and Costa's (1983) study had, and still has, on the interpretation of SDR, the purpose of the present study is to replicate it precisely, with the exceptions of controlling for SDR in both self-reports and spouse ratings. It is hypothesized that because spouse ratings may also be biased by SDR, removing the bias from both self-reports and spouse ratings will improve their agreement. Thus showing, contrary to the findings reported by McCrae and Costa, that SDR does affect the validity of personality scales and that SDR scales can be used to rectify the problem.

Method

Participants

Seventy-two couples who had lived together for at least six months, and nearly half (44%) of whom were married, participated in this study. Nine couples who indicated they had discussed some of the questionnaires items with their spouses were excluded from all analysis. Therefore, data from 63 women and 63 men, aged 21 to 59 ($M = 34$, $SD = 9.91$) were analyzed. Participants were volunteers and did not receive compensation for study participation. Approximately 70% of respondents returned completed questionnaires.

Measures

Personality measure. The NEO-PI-R inventory is a 240-item measure of the Big Five personality traits: Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness. Each dimension has six facets consisting of eight items rated on a five-point scale ranging from “strongly disagree” to “strongly agree” (Costa and McCrae, 1992). In accordance with McCrae and Costa (1983), the self-report form (Form S) and the observer rating form (Form R) were used to measure the first three dimensions (Neuroticism, Extraversion and Openness to Experience).

Neuroticism contains 48 items that measure the predisposition to experience psychological distress. The six facets are Anxiety, Angry Hostility, Depression, Self-Consciousness, Impulsiveness and Vulnerability. Examples of items include “I am not a worrier” and “I’m an even-tempered person” (Costa & McCrae, 1992). For the Neuroticism domain, reported internal consistency of the Icelandic self-report

form is 0.91 (Jónsson & Bergþórsson, 2004), which is similar to the original version (0.92) (Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form (0.90) is also similar to the original version (0.91) (McCrae & Terracciano, 2005). In this study, Cronbach's alphas for Neuroticism were 0.93 (Form S) and 0.96 (Form R).

Extraversion contains 48 items used to measure, for example, friendliness and preference for the company of others. The six facets are Warmth, Gregariousness, Assertiveness, Activity, Excitement-Seeking and Positive Emotions. Examples of items include "I really like most people I meet" and "I shy away from crowds of people" (Costa & McCrae, 1992). For the Extraversion domain, reported internal consistency of the Icelandic self-report form is 0.88 (Jónsson & Bergþórsson, 2004), which is similar to the original version (0.89) (Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form is equal to the original version (0.91) (McCrae & Terracciano, 2005). In this study, Cronbach's alpha for Extraversion was 0.92 (Form S and Form R).

Openness to Experience contains 48 items used to measure, for example, aesthetic sensitivity and intellectual curiosity. The six facets are Fantasy, Aesthetics, Feelings, Actions, Ideas and Values. Examples of items include "I have a very active imagination" and "Watching ballet or modern dance bores me" (Costa & McCrae, 1992). For the Openness domain, reported internal consistency of the Icelandic self-report form is 0.87 (Jónsson & Bergþórsson, 2004), which is equal to the original version (Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form is also equal to the original version (0.88) (McCrae & Terracciano, 2005). In this study, Cronbach's alphas for Openness to Experience were 0.82 (Form S) and 0.85 (Form R).

Social desirability measure. The Icelandic version of the Marlowe-Crowne Social Desirability Scale (MCSD) was used as a measure of socially desirable responding (SDR). The MCSD is a true-false questionnaire consisting of 33 statements of either uncommon but socially desirable behavior or common but undesirable behavior (Crowne & Marlowe, 1960). In this study, Cronbach's alphas were 0.77 and 0.73 for the self-report and the observer rating form respectively.

Procedure

Each couple received an envelope containing the self-report and observer rating form of the NEO-PI-R inventory followed by the Marlowe-Crowne Social Desirability Scale. Participants filled out the questionnaires in their own homes under the instructions to answer one of the two forms independently of their spouses. As a check on these instructions, an item was added at the end of both inventories asking if any items had been discussed with their spouse. Those who answered "yes" were excluded from all analysis. Upon completion, participants returned the questionnaires in an envelope labeled only with identification numbers to insure complete anonymity. Data collection started in 2012 and ended in 2013.

Results

Because the purpose of this study was to replicate McCrae and Costa's (1983) study with the exception of correcting for SDR in both self-reports and spouse ratings, results are presented the same way the authors presented them in their article.

Table 1 shows correlations between self-reports and spouse ratings on the dimensions of Neuroticism, Extraversion and Openness to Experience (NEO), together with partial correlations correcting for SDR.

Table 1

Correlations Between Self-Reported and Spouse-Rated NEO Scales correcting for the Marlowe-Crowne Social Desirability (MCSD) Scale

Trait	Present study		McCrae & Costa (1983)	
	Simple correlation	Correcting for MCSD ^a	Simple correlation	Correcting for MCSD ^b
Anxiety	0.61***	0.56***	0.59***	0.58***
Hostility	0.59***	0.64***	0.53***	0.50***
Depression	0.66***	0.66***	0.51***	0.48***
Self-consciousness	0.47**	0.47***	0.37***	0.35***
Impulsiveness	0.46***	0.38**	0.47***	0.41***
Vulnerability	0.58***	0.58***	0.45***	0.44***
Neuroticism	0.65***	0.66***	0.58***	0.54***
Warmth	0.57***	0.61***	0.48**	0.47***
Gregariousness	0.57***	0.61***	0.58**	0.58***
Assertiveness	0.59***	0.62***	0.54***	0.54***
Activity	0.65***	0.67***	0.56***	0.57***
Excitement seeking	0.54***	0.67***	0.51***	0.50***
Positive emotions	0.41**	0.44**	0.35***	0.35***
Extraversion	0.67***	0.71***	0.61***	0.61***
Fantasy	0.31*	0.31*	0.37***	0.34***
Aesthetics	0.56***	0.59***	0.58***	0.58***
Feelings	0.41**	0.47***	0.25***	0.25***
Actions	0.44***	0.46***	0.36***	0.35***
Ideas	0.72***	0.72***	0.49***	0.49***
Values	0.08	0.05	0.50***	0.49***
Openness	0.54***	0.61***	0.50***	0.49***

Note. From “Social Desirability Scales: More Substance Than Style” by R. McCrae and P. Costa, 1983, *Journal of Consulting and Clinical Psychology*, 51(6), p. 885.

* $p < .05$, ** $p < .01$, *** $p < .001$.

^aPartial correlation correcting for SDR in both self-reports and spouse ratings.

^bPartial correlation correcting for SDR in self-reports only.

In the present study, correcting for SDR in both self-reports and spouse ratings increased their agreement in most cases, showing that removing variance associated with SDR enhances the self-reported validity of most traits. Corrected correlations

were lower than the simple correlations for just three traits. These findings are inconsistent with the results obtained by McCrae and Costa (1983) which are also shown in Table 1. Correcting for SDR in only self-reports decreased their agreement with spouse ratings in most cases, and increased agreement on only one trait.

Discussion

The present study replicated an influential study by McCrae and Costa (1983) with the exception of correcting for SDR in both self-reports and spouse ratings. In accordance with our hypotheses, removing variance associated with SDR from both self-reports and spouse ratings improved their agreement on most traits. These results are in contrast with the findings reported by McCrae and Costa who arrived at the opposite conclusion. Correcting for SDR in only self-reports lowered their agreement with spouse ratings on most traits. Because of this, the authors concluded that the attempt to enhance validity by correcting for SDR was unsuccessful.

The present study's results are however more consistent with the findings reported by Konstabel et al. (2006) that correcting for SDR in both self-reports and peer-reports improves their agreement on all Big five dimensions. These results indicate, contrary to the conclusion of McCrae and Costa, that SDR scales can be useful for increasing the validity of personality measures. This inconsistency is most likely due to the recognition that observer ratings may also be biased by SDR. It has been implicitly assumed that only self-reports are susceptible to SDR and that the only variance common to self-reports and observer ratings is real trait variance. The current results, along with the findings from Konstabel et al. (2006), suggest that the failure of previous studies to enhance validity by correcting for SDR at least

partly results from the choice of external criteria and from disregarding SDR in observer ratings (among others, McCrae & Costa, 1983; Kurtz et al., 2008; Piedmont et al., 2000; Pauls & Stemmler, 2003). An ideal way to resolve the SDR debate would be to compare self-reports to true personality scores. The impossibility of obtaining true personality scores is why researchers started to compare self-reports to observer ratings. Although a step forward from self-report studies, observer ratings are not as objective as, for example, observed behavior. External criteria that is more objective than peer and spouse ratings might be necessary to disentangle substance from style. Therefore it may be erroneous, when correcting for SDR in self-reports only, to interpret lower agreement between self-reports and observer reports as a failure of the method to enhance validity.

Because correcting for SDR failed to enhance validity in previous studies, some researchers have argued that SDR scales are not useful (e.g. Borkenau & Ostendorf, 1992; Piedmont et al., 2000) and Costa and McCrae (1997) refuse to add validity scales to identify SDR to the NEO-PI-R, despite popular demand. The authors claim that they remain unconvinced of the utility of such scales and in support of that opinion cite studies, including their own from 1983, showing that correcting for SDR in self-reports reduces their validity as assessed by observer ratings (p. 89). Although they claim that the omission of SDR scales from the NEO-PI-R is not an oversight, the susceptibility of observer ratings to SDR might be. Other researchers have argued that SDR does not affect the validity of personality measures (e.g. Smith & Ellingson, 2002). The present study suggests that both of these conclusions are unwarranted, although the small sample size and the use of only one measure of SDR are a limitation to that assertion.

Even though our findings were in the hypothesized direction, some might regard the differences between the simple correlations and the corrected correlations as trivial. This can at least partly be explained by the research context. Participants were anonymous volunteers and thus not motivated to distort their answers. Despite disinterested sample, the Marlowe and Crowne Social Desirability scale (MCSD) showed utility in enhancing the validity of self-reported personality traits. Studies testing SDR scales in clinical or selection settings might report larger validity increments when correcting for SDR in both self-reports and observer ratings. However, correcting for SDR is likely to produce minimal incremental validity within the range of typical validity coefficients due to mathematical constraints (Conger & Jackson, 1972). In addition, trivial increments on some traits might be the result of the scale used to measure SDR. Holden (2007) has argued that the effects of SDR on the validity of personality scales might be underestimated due to the fallibility of traditional SDR scales. Specifically, the MCSD has been criticized for the overqualified nature of many of the scales items (Kurtz et al., 2008) and for its multidimensional structure (Barger, 2002). Perhaps the future direction of SDR research is a better explication of the SDR construct and consequently the design of more effective SDR scales. Also, SDR scales could be designed for different contexts to maximize their effectiveness in detecting the sort of distortion expected in each situation. Different SDR scales could thus be used in personnel assessment than in clinical situations.

This discussion brings up an important point. Even if SDR scales such as the MCSD may measure more substance than style (or entirely substance and no style), it does not follow that SDR or faking is not a cause for concern in personality assessment. More effective measures of SDR might be needed. In fact, it has been

shown that faking is possible (e.g. Meehl & Hathaway, 1946), that it does degrade the construct validity of personality measures (among others, Ellingson et al., 1999; Rosse, Stecher, Miller & Levin, 1998; Topping and O'Gorman, 1997) and that SDR scales can identify fakers (e.g. Holden, 2007). In addition, the findings from the present study, along with the ones from Konstabel et al. (2006), offer support to the conclusion that SDR scales can be used to enhance the validity of self-reported personality traits.

Many issues regarding the nature of SDR are however still left unexamined, such as gender differences. Intuitively it seems plausible that what is considered socially desirable in a given context may vary by gender. Indeed, according to the artifact model (Feingold, 1994), the self-reports of men and women are differentially biased by SDR because men and women have different beliefs about which traits are desirable due to factors such as gender stereotyping. Women may thus overreport their level of nurturance while men may underreport their level of fearfulness. Studies using the MCSD report mixed results of gender differences. Some researchers have found none (e.g. Loo and Thorpe, 2000) while others have found that women score higher than men do (e.g. Hebert et al., 1997). Whether gender differences in SDR contribute to the lack of substantial suppressor effect in the personality literature is open to speculation.

Although some researcher may regard the issue of SDR as a “methodological dead horse” (Nevid, 1983, p. 139), we believe, among others (Bäckström, 2007; Holden, 2007; Linehan & Nielsen, 1983), that the horse still has a pulse. Our hope is that more effort will be put into investigation of the nature of SDR and to the improvement of SDR scales. The common practice of citing

McCrae and Costa's (1983) as support for the conclusion that SDR is not an issue in personality assessment is questionable.

References

- Bäckström, M. (2007). Higher-order factors in a five-factor personality inventory and its relation to social desirability. *European Journal of Psychological Assessment, 23*(2), 63-70.
- Barger, S. D. (2002). The Marlowe-Crowne affair: Short forms, psychometric structure, and social desirability. *Journal of Personality Assessment, 79*(2), 286-305.
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6*(3), 199-214.
- Conger, A. J., & Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement, 32*, 579-599.
- Costa, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five factor inventory (NEO-FFI). Professional manual. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: The revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68*(1), 86-94.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*(4), 475-494.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349-354.

- Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement*, 23, 699-720.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37(2), 90-93.
- Edwards, A. L. (1957). The social desirability variable in personality assessment and research. New York: Dryden.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84, 155–166.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116(3), 429-456.
- Funder, D. C., & Colvin, C. R. (1997). Congruence of other's and self-judgments of personality. In R. R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego: Academic Press.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400.
- Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, 26(2), 321-331.
- Hebert, J. R., Ma, Y., Clemow, L., Ockene, I. S., Saperia, G., Stanek, E. J., ... & Ockene, J. K. (1997). Gender differences in social desirability and social approval bias in dietary self-report. *American Journal of Epidemiology*, 146(12), 1046-1055.

Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science*, 39(3), 184-201.

Jónsson, F. H., & Bergþórsson, A. (2004). Fyrstu niðurstöður úr stöðlun NEO-PI-R á Íslandi. *Sálfræðiritið*, 9, 9-16.

Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality*, 20(7), 549-566.

Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: A systematic evaluation. *Journal of Gerontology*, 42, 56–59.

Kurtz, J. E., Tarquini, S. J., & Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences*, 45(1), 22-27.

Linehan, M. M., & Nielsen, S. L. (1983). Social desirability: Its relevance to the measurement of hopelessness and suicidal behavior. *Journal of Consulting and Clinical Psychology*, 51, 141-143.

Loo, R., & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-Crowne Social Desirability Scale. *The Journal of Social Psychology*, 140(5), 628-635.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882-888.

McCrae, R. R., Costa, P. T., Dahlstrom, W. G., Barefoot, J. C., Siegler, I. C., & Williams, R. B. (1989). A caution on the use of the MMPI K-correction in research on psychosomatic medicine. *Psychosomatic Medicine*, 51(1), 58-65.

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology*, 88(3), 547-561.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30(5), 525-563.

Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry*, 10(1), 23-34.

Nevid, J. S. (1983). Hopelessness, social desirability, and construct validity. *Journal of Consulting and Clinical Psychology*, 51, 139-140.

Nicholson, R. A., & Hogan, R. T. (1990). The construct validity of social desirability. *American Psychologist*, 45, 290-292.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81(6), 660-679.

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: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement*, (pp. 49-69). New Jersey: Erlbaum.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences*, 35(2), 263-275.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78(3), 582-593.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63(3), 129-156.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634-644.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, 87(2), 211-219.

Tesser, A., Pilkington, C. J., & McIntosh, W. D. (1989). Self-evaluation maintenance and the meditational role of emotion: The perception of friends and strangers. *Journal of Personality and Social Psychology*, 57, 442–456.

Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences*, 23, 117-124.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality assessment. *Educational and Psychological Measurement*, 59, 197-210.