



Mat á lífsgæðum og samanburður á sjálfsmatskvörðum

Arna Frímannsdóttir

**Lokaverkefni til MS-gráðu
Sálfræðideild
Heilbrigðisvísindasvið**



HÁSKÓLI ÍSLANDS

Mat á lífsgæðum og samanburður á sjálfsmatskvörðum

Arna Frímannsdóttir

Lokaverkefni til MS-gráðu í sálfræði

Leiðbeinandi: Fanney Þórsdóttir

Sálfræðideild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2014

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

© Arna Frímannsdóttir 2014

Prentun: Háskólaprent ehf.

Reykjavík, Ísland 2014

ÞAKKARORÐ

Fyrst og fremst vil ég þakka leiðbeinandanum mínum, Fanneyju Þórsdóttir, fyrir einstaklega góða leiðbeiningu. Hún hvatti mig til sjálfstæðra verka en var þó alltaf til staðar. Hinir ótalmörgu fundir á skrifstofu hennar voru blanda af grúski, þælingum, fróðleik og skemmtilegheitum. Einnig langar mig að þakka Antoni og Arndísi, doktorsnemunum, fyrir fræðandi og ánægjulegt samstarf.

Að síðustu vil ég þakka fjölskyldu minni fyrir mikla og einlæga aðstoð. Foreldrum mínum fyrir yfirlestur og barnapössun, tengdaforeldrum mínum fyrir að einfalda heimilishaldið með því að bjóða okkur ítrekað í mat og strákunum mínum: Gaua, Nökkva og Sölva Kára fyrir að halda mér á tánum við eitthvað allt annað en lærdóm.

EFNISYFIRLIT

FORMÁLI	6
Heimildir	11
The moderating effect of Protestant work ethic on the relationship between employment status and material well-being.....	14
Abstract	14
Introduction.....	15
Method	17
Sample	17
Measures.....	17
Results.....	19
Discussion	21
References.....	22
Five- or Seven-point Rating Scales: The Interplay between Verbal Labels and Number of Response Options.....	25
Abstract.....	25
Introduction.....	26
Background	27
Summary and hypotheses.....	32
Method	34
Participants	34
Measurements.....	34
Procedure.....	36
Results.....	36
Descriptive statistics.....	36
Hypotheses testing.....	37
Discussion	39
Implications	40
References.....	42
Appendix A	54

FORMÁLI

Meistaraverkefni þetta samanstendur af tveimur handritum (*manuscripts*) ætluðum til birtingar í tímariti. Fyrri greinin var unnin í samstarfi við Arndísi Vilhjálmsdóttur, doktorsnema, og þær báðar undir leiðsögn Fanneyjar Þórsdóttir, lektors. Fyrri greinin er á sérsviði Arndísar og fjallar um samband atvinnustöðu og vellíðunar og hugsanlega samvirkni milli vinnusiðferðis og atvinnustöðu á vellíðan. Eftir efnahagshrunið 2008 jókst atvinnuleysi til muna, bæði hérlandis sem og í öðrum vestrænum löndum. Þegar slíkar breytingar verða er mikilvægt að kanna hverjar hugsanlegar afleiðingar eru og hverjir eru líklegir til að verða fyrir þeim, svo hægt sé að bregðast við með viðeigandi hætti. Með vitneskju um jákvæð tengsl tekna og vellíðunar (Diener, Inglehart og Tay, 2013; Frey og Stutzer, 2002) er eðlilegt að vænta að þeir sem eru atvinnulausir séu ósáttari við sín efnislegu lífsgæði en þeir sem eru á vinnumarkaði. Rannsóknin varpaði einnig fram þeirri tilgátu að áhrif atvinnustöðu, það er að segja hvort viðkomandi er atvinnulaus eða í vinnu, á efnisleg lífsgæði yltu á vinnusiðferði. Það er að þeir sem eru atvinnulausir og hefðu veikt vinnusiðferði væru óánægðari með efnisleg lífsgæði sín en þeir sem eru atvinnulausir og hefðu sterkt vinnusiðferði. Þessi tilgáta var studd með vísan í að þeir sem eru annars vegar með lágt vinnusiðferði og hins vegar hátt vinnusiðferði meta rétt sinn á atvinnuleysisbótum mismikinn. Þeir sem eru með lágt vinnusiðferði ættu að meta rétt sinn á atvinnuleysisbótum sterkari en þeir sem eru með hátt vinnusiðferði og það hversu mikinn rétt viðkomandi telur sig hafa á atvinnuleysisbótum getur verið notað sem viðmið þegar efnisleg lífsgæði eru metin. Vinnusiðferði (*Protestant work ethic*) (Weber, 2003) felur í sér að vinnan göfgi manninn (Furnham, 1984) og að mikilvægt sé að reiða á sjálfan sig (Miller, Woehr og Hudspeth, 2002) og því skuli allir megnugir menn sjá fyrir sér sjálfir (Ciulla, 2000).

Tilgátur voru kannaðar með gögnum úr íslenska úrtaki Evrópsku gildakönnunarinnar frá 2008. Alls voru þátttakendur 397 en útreikningar miðuðust við þátttakendur á aldrinum 31 til 60 ára þar sem talið var að sá aldurshópur endurspeglaði fullorðið fólk, með tilliti til ábyrgðar og þar af leiðandi þörfum á tekjum. Vinnusiðferði var mælt með fimm atriðum, öll metin með fimm punkta kvarða (*strongly agree* til *strongly disagree*). Efnisleg lífsgæði voru mæld með tveimur atriðum sem bæði voru metin með tíu punkta kvarða (*dissatisfied* til *satisfied*).

Niðurstöður studdu fyrri tilgátuna, það er að tengsl væru milli atvinnustöðu og mats á efnislegum lífsgæðum ($b=4,72, p<0,001$). Stuðningur fékkst einnig við seinni tilgátunni, það er að samvirkni væri milli atvinnustöðu og efnislegra lífsgæða ($b=-0,74, p<0,001$). Með því að kanna samvirknina betur kom í ljós að atvinnustaða spáði ekki fyrir um efnisleg lífsgæði meðal þeirra með hátt vinnusiðferði ($b=0,02, p=0,93$). Því er ekki hægt að segja að atvinnulausir hafi verið ósáttari með efnisleg lífsgæði sín en þeir sem voru í vinnu, meðal þeirra sem hafa hátt

vinnusiðferði. Þegar sambandið var skoðað meðal þeirra með lágt vinnusiðferði kom í ljós að þar spáði atvinnustaða fyrir um lífsgæði ($b=9,98, p<0,001$). Þeir sem eru atvinnulausir eru því ósáttari með efnisleg lífsgjör sín en þeir sem eru í vinnu, meðal þeirra með lágt vinnusiðferði. Þessar niðurstöður fela í sér einskonar mótsögn hjá þeim sem hafa lágt vinnusiðferði (lítil atvinnuskuldbinding) og upplifa óánægju með efnisleg lífsgæði sín (sterkur hvati til að vinna).

Hvatinn að seinni greininni fólst í vinnu minni á niðurstöðum fyrri greinarinnar. Eins og fyrr segir voru spurningarnar um vinnusiðferði metnar með fimm punkta kvarða og spurningarnar um efnisleg lífsgæði með tíu punkta kvarða. Þetta ósamræmi kveikti áhuga minn á mikilvægi þess að nota *réttan* kvarða við gagnaöflun. Fanney, leiðbeinandi minn og sérfræðingur á þessu sviði, upplýsti mig um að gögn sem aflað er með ólíkum kvörðum, bæði hvað varðar lengd og orðagildanotkun, hafa mismikil gæði. Þar sem flestar rannsóknir í sálfræði, og víðar, styðjast við svarkvarða varð mér ljóst hve mikilvægar samanburðarrannsóknir á kvarðalengdum eru. Upplýsingar um hvaða kvarða beri að nota til að draga úr líkum á svarskekkju, og afla þannig gagna af meiri gæðum, eru gríðarlega mikilvægar fyrir alla rannsakendur er nota svarkvarða í sínum rannsóknum. Þrátt fyrir að rannsóknir hafi sýnt að gögn aflað með fullmerktum kvörðum (*fully labelled rating scales*) innihaldi meiri gæði en gögn sem aflað er með númerískum kvörðum (Alwin og Krosnick, 1991; Krosnick og Berent, 1993; Zaller, 1988; Dickinson og Zellinger, 1980; Andrews, 1984; Wedell, Parducci, og Lane, 1990), hafa númerískir kvarðar verið nær eingöngu notaðir í fyrri samanburðarrannsóknum. Á fullmerktum kvörðum er hver svarmöguleiki merktur með orðagildi en á númerískum kvörðum eru einungis svarmöguleikarnir á hvorum enda merktir með orðagildi. Þær samanburðarrannsóknir sem notað hafa fullmerkta kvarða hafa gefið misvísandi niðurstöður (Boote, 1981; Masters, 1974; McKelvie, 1978). Þegar þær rannsóknir voru skoðaðar nánar kom í ljós að ólík orðagildi voru notuð milli rannsókna, en hægt er að búa til mjög ólíka kvarða með þeim orðagildum sem til eru. Svokallaðar gildarannsóknir (*scaling studies*), þar sem þátttakendur eru beðnir um að gefa ólíkum orðagildum tölugildi eftir því hversu sterku viðhorfi þau lýsa, gefa upplýsingar sem hægt er að nota til að setja saman ólíka kvarða. Þannig getur kvarði bæði verið þröngur og víður, með tilliti til viðhorfavíddarinnar. Séu orðagildi sem lýsa mjög sterku viðhorfi (fá hátt tölugildi í gildarannsóknum) notuð á endana telst kvarðinn vera víður en séu orðagildi sem lýsa veikara viðhorfi merkt á enda kvarðans er sá kvarði þrengri með tilliti til viðhorfavíddarinnar. Sýnt hefur verið fram á að best sé að nota kvarða sem spanni sem víðasta bilið á viðhorfavíddinni (Jones og Thurstone, 1955). Eins og fyrr segir hafa fyrri samanburðarrannsóknir, sem stuðst hafa við fullmerkta kvarða, gefið misvísandi niðurstöður. Allar báru þær saman fimm og sjö punkta kvarða með tilliti til

gagnagæða. Masters (1974) komst að því að meiri gæði fengust með sjö punkta kvarða en fimm punkta kvarða. Niðurstöður Boote (1981) sýndu hins vegar að meiri gæði fengust þegar fimm punkta kvarði var notaður fremur en sjö punkta kvarði. Þegar orðagildin, sem notuð voru í þessum samanburðarrannsóknum, eru skoðuð sést að sjö punkta kvarðinn sem Masters (1974) notaði var víðari en fimm punkta kvarðinn en fimm- og sjö punkta kvarðarnir í rannsókn Boote (1981) voru jafnvíðir. Það er því hugsanlegt að skýra megi þessar ólíku niðurstöður með vísan í ólík orðagildi sem valda því að kvarðarnir eru misvíðir. Þessi möguleiki varð kveikjan að tilgátum seinni rannsóknarinnar en markmið hennar var að skoða hlutverk orðagilda og hugsanlegt samspil þeirra og fjölda svarmöguleika á gæði gagna.

Ákveðið var að bera saman fimm og sjö punkta kvarða, þar sem þær kvarðalengdir hafa verið hvað mest rannsakaðar og eru mikið notaðar. Tvö pör af fimm og sjö punkta kvörðum voru notuð en í öðru parinu voru kvarðarnir jafnvíðir (fimm punkta víður og sjö punkta finn) en í hinu var sjö punkta kvarðinn víðari en sá fimm punkta (fimm punkta þröngur og sjö punkta langur). Með vísan í ólíkar niðurstöður fyrri samanburðarrannsókna (Boote, 1981; Masters, 1974) og vísbendingar um að gögnum afluð með víðari kvörðum innihaldi meiri gæði en gögn sem afluð er með þröngri kvörðum (Jones og Thurstone, 1955) var því spáð að gagnagæðin myndu aukast þegar kvarðinn væri lengdur, það er þegar fimm punkta þröngur og sjö punkta langur eru bornir saman, en að það drægi úr gagnagæðum þegar kvarðinn var gerður finn, það er þegar fimm punkta víður og sjö punkta finn eru bornir saman.

Fyrri rannsóknir hafa flestar notað áreiðanleika sem mat á gæðum gagna en vegna þess að svarskekkja getur aukið áreiðanleika var ákveðið að nota svokallaða svarstíla til þess að meta gagnagæðin. Svarskekkja er skilgreind sem kerfisbundinn tilhneiging til að svara á öðrum grunni en innihalds spurningarinnar/staðhæfingarinnar (Cronbach, 1946; Lentz, 1938; Paulhaus, 1991). Þegar svarskekkja verður stöðug yfir tíma og aðstæður kallast hún svarstill (Jackson og Messick, 1958; Wiggins, 1973). Svarstílarnir jáhneigð og neihneigð voru valdir vegna þess að þeir eru meðal mest rannsökuðu svarstíla (Baumgartner og Steenkamp, 2001; Paulhus, 1991). Jáhneigð er skilgreind sem tilhneiging svaranda til þess að vera sammála staðhæfingu og neihneigð er andstæðan, eða að vera ósammála staðhæfingu. Svarstílarnir hafa verið tengdir við alls kyns bakgrunnspætti, til dæmis hefur jáhneigð verið tengd við lægra menntunarstig (Greenleaf, 1992) og lægri tekjur (Krosnick, 1999). Beiting svarstíla getur dregið mjög úr gæðum gagnanna, til dæmis við skalapróunarvinnu þar sem hún eykur fylgni, ranglega, milli atriða sem eru orðuð í sömu átt á kostnað fylgni milli atriða, sem eru orðuð í sitt hvora áttina, en tengjast þó. Helsta ástæðan fyrir já- og neihneigð er talin vera uppfylling (*satisficing*) (Krosnick 1991) sem á sér stað þegar svarandinn fer í gegnum stig svarferlisins (Tourangeau, 1984).

Svarferlið skiptist í fjögur stig, í því fyrsta túlkar þátttakandinn spurninguna og í því næsta reynir hann að rifja upp upplýsingar er tengjast efni spurningarinnar. Þriðja stigið samanstendur af samþættingu á öllum upplýsingunum í eina niðurstöðu og í fjórða stiginu verður þátttakandinn að umbreyta niðurstöðu sinni í eitt svar, eða eins og í tilfelli svarkvarða: í einn svarmöguleika (Tourangeau, 1984). Krosnick (1991) taldi að svarendur færu ekki allir fullkomnlega í gegnum öll fjögur stigin í svarferlinu heldur að sumir beittu uppfyllingu, veikri eða sterkri. Veik uppfylling ætti sér stað þegar svarandinn færi í gegnum öll fjögur stigin en ekki eins vel og þegar um fullkomið svar (*optimize*) er að ræða. Í sterkri uppfyllingu, hins vegar, er stigum tvö og þrjú sleppt úr svarferlinu (Krosnick, 1991). Samkvæmt uppfyllingarkenningunni (*satisficing theory*) hefur erfiðleikastig verkefnisins mikil áhrif á það hvort að svarandi svari fullkomnlega eða beiti uppfyllingu. Til að heimfæra þessar upplýsingar á rannsóknina sem um ræðir hér má ætla að sú kvarðagerð sem veitist svarendum erfiðust ætti að mælast með hæsta magn af já- og neihneigð.

Rannsóknin var unnin upp úr gögnum sem safnað var af Capacent á Íslandi. Um var að ræða þrjá spurningalista sem lagðir voru fyrir alla þátttakendur (N=1473) en þátttakendum var þó skipt í fjóra hópa. Hver hópur svaraði með einni kvarðagerð, það er fimm punkta þröngum, sjö punkta löngum, fimm punkta víðum eða sjö punkta finum. Kvarðagerðunum fjórum var svo skipt í tvö pör sem samanstóðu annars vegar af fimm punkta þröngum og sjö punkta löngum og hins vegar fimm punkta víðum og sjö punkta finum. Ólíkt sjö punkta langa kvarðanum spannaði fimm punkta þröngi kvarðinn ekki alla viðhorfavíddina og viðbótarsvarmöguleikarnir í sjö punkta löngum bættust því utan á endagildi fimm punkta þrönga kvarðans. Fimm punkta víði og sjö punkta fini kvarðinn spönnuðu báðir alla viðhorfavíddina en viðbótarsvarmöguleikar sjö punkta fina kvarðans skipuðu sér því inná milli annara svarmöguleika fimm punkta víða kvarðans og gerðu sjö punkta fina kvarðan því finari en þann fimm punkta víða.

Svarstílnir voru metnir með aðferð sem kallast misræmi í svörum (*misresponse*). Þannig fengust stig fyrir að vera (ó)sammála tveimur staðhæfingum sem voru orðaðar í sitt hvora áttina. Stigin voru vigtuð og fengust því fleiri já- eða neihneigðarstig eftir því sem svarandi tók sterkari afstöðu. Niðurstöður sýndu mun á já- og neihneigð í lengingarparinu en ekki finpússunarparinu. Það mældist meiri jáhneigð í gögnunum sem aflað var með fimm punkta þröngum heldur en sjö punkta löngum. Að sama skapi mældist meiri neihneigð í gögnunum sem aflað var með fimm punkta þröngum heldur en sjö punkta löngum. Ekki var hægt að staðfesta mun á milli svarstílamagns í gögnum sem annars vegar var aflað með fimm punkta víðum og hins vegar sjö punkta finum. Niðurstöðurnar studdu því aðra af tveimur tilgátum rannsóknarinnar, það er að það dragi úr svarstílsnotkun við það að lengja kvarða, það

er að fara úr fimm punkta þröngum í sjö punkta langan. Ekki fékkst stuðningur við seinni tilgátunni, það er að svarstílsnotkun aukist við að fínþússa kvarðann, að fara úr fimm punkta víðum í sjö punkta finan.

Niðurstöður þessar benda til þess að ástæðuna fyrir misvísandi niðurstöðum fyrri samanburðarrannsókna á fullmerktum kvörðum megi rekja til lengdar kvarðanna á viðhorfavíddinni. Það styður tilvist samvirkni milli orðagilda og punktalengdar á gagnagæði, það er að hvort betra er að nota fimm eða sjö punkta kvarða geti oltið á orðagildunum sem notuð eru á kvarðana. Rannsakendur sem stunda samanburðarránsóknir verða að taka þessa samvirkni með í reikninginn í leit sinni að hinum fullkomna kvarða.

Heimildir

- Alwin, D. F. og Krosnick, J. A. (1991). The reliability of survey attitude measurement. The influence of question and respondent attributes. *Sociological Methods & Research*, 20, 139–181.
- Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48, 409-442.
- Baumgartner, H. og Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Boote, A. S. (1981). Reliability testing of psychographic scales: five-point or seven-point? Anchored or labeled? *Journal of Advertising Research*, 21, 53-60.
- Ciulla, J. B. (2000). *The promise and betrayal of modern work*. New York: Three Rivers Press.
- Cronbach, L. J. "Response sets and test validity." *Educational and psychological measurement* 6.4 (1946): 475-494.
- Dickinson, T. L. og Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating mixed standard scale formats. *Journal of Applied Psychology*, 65, 147-154.
- Diener, E., Inglehart, R., and Tay, L. (2013). Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, 112, 497-527.
- Frey, B. og Stutzer, A. (2002). What can economists learn from happiness research ? *Journal of Economic Literature*, 40, 402–435.
- Furnham, A. (1984). The protestan work ethic: A review of the psychological literature. *European Journal of Social Psychology*, 17, 87–104.

- Greenleaf, E. A. (1992). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research*, 29, 176-88.
- Jones, L. V. og Thurstone, L. L. (1955). The psychophysics of semantics: An experimental investigation. *The Journal of Applied Psychology*, 39, 31-36.
- Jackson, D. N. og Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243-252.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Krosnick, J. A. and Berent, M. K. (1993). Comparisons of party identification and policy preferences: the impact of survey question format. *American Journal of Political Science*, 37, 941-964
- Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin*, 35, 659.
- Masters, J. R. (1974). The relationship between number of response categories and reliability of likert-type questionnaires. *Journal of Educational Measurement*, 11, 49-53.
- McKelvie, S. J. (1978). Graphic rating scales-How many categories. *British Journal of Psychology*, 69, 185-202.
- Miller, M., Woehr, D. J. og Hudspeth, N. (2002). The meaning and measurement of work ethic: Construction and initial validation of a multidimensional inventory. *Journal of Vocational Behavior*, 60, 451-489.

- Paulhus, D. L. (1991) Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, and L. S. Wrightman (Eds.), *Measures of Social Psychological attitudes Series* (pp. 17-59). San Diego, CA: Academic Press.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, and R Tourangeau (Eds.), *Cognitive aspects of survey methodology: building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.
- Weber, M. (2003). *The protestant work ethic and the spirit of capitalism*. Mineola: Dover Publications.
- Wedell, D. H., Parducci, A. og Lane, M. (1990). Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and types of anchors. *Journal of Personality and Social Psychology*, 58, 319-329.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading MA: Addison-Wesley.
- Zaller, J. (1988). "Vague Minds vs. Vague Questions: An Experimental Attempt to Reduce Measurement Error." Kynnt á árlegum fundi American Political Science Association, Washington, DC.

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE
RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

**The moderating effect of Protestant work ethic on the relationship between
employment status and material well-being**

Arndís Vilhjalmsdóttir^a

Arna Frimannsdóttir^a

Fanney Thorsdóttir^a

^aUniversity of Iceland, Faculty of Psychology, Oddi, 101 Reykjavík, Iceland

Abstract

This study focuses on the relationship between employment status and material well-being and the moderating effect of Protestant work ethic (PWE) on this relationship. Data from the 2009 European Value Study was used to explore this issue. The findings show that material well-being decreases when people are unemployed and also that the relationship between employment status and material well-being depends on PWE. Implications of these findings are discussed.

Keywords: Protestant work ethic, Unemployment, Material well-being, European Values Study

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

Introduction

Material well-being is an important predictor of life satisfaction (Diener, Ingelhart & Tay, 2013) and the importance of identifying factors that affect material well-being has been recognised for some time. Following the recent economic recession, unemployment became a serious problem in many developed countries where unemployment rates rose considerably (Guichard & Rusticelli, 2010; OEDC, 2009). Exactly how such changes affect people's sense of material well-being is not clear. It has been suggested that income determines material satisfaction, through access to socially defined necessities (Diener et al., 1999; Hobfoll et al., 1996; Ullah, 1990). Since unemployment benefits are usually low, leading to restricted access to important resources, a decrease in material satisfaction when people become unemployed might be expected. To our knowledge, no studies have focused on the direct relationship between unemployment and material well-being. However, studies within a given country at a given point in time have repeatedly shown that general well-being depends on absolute income, people with higher income reporting more well-being (Diener et al., 2013; Frey & Stutzer, 2002). Thus, there is indirect evidence for a link between employment status and material well-being. The aim of the present study was to establish this relationship more directly. Since the unemployment benefits are generally low in a given country, people who are unemployed should be less satisfied with their material situation than people who are employed. We therefore hypothesize that:

(H1) Unemployed people are less satisfied with their material well-being than employed people.

As discussed above, empirical research on the relationship between income and general well-being within a given country suggests that people with higher income report more well-being

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

(Diener et al., 2013; Frey & Stutzer, 2002). Interestingly though, studies looking at well-being at the macro level have not found relationship between well-being in general and GDP (gross domestic product) (Easterlin et al., 2010). These findings suggest that much of general well-being is due to relative income rather than absolute income; people with higher income are more satisfied with life because they have higher income than most people around them. For material well-being, these findings imply that when people can judge their financial situation as better than a standard financial situation, they are more materially satisfied.

This presumed relativity of income may prove to be important when the relationship between unemployment and material well-being is explored. Clearly since unemployment benefits are usually lower than the income of most people, we would still expect unemployment to significantly reduce material well-being.

However, when people are unemployed, income can be relative in another sense, namely the entitlement to receive unemployment benefits may be used as a standard when financial situation is being judged.

Thus, the negative impact of unemployment on material well-being may depend on a normative disposition regarding the unemployment situation. More specifically, those who are unemployed and have a weak adherence to Protestant work ethic (PWE) may be less satisfied with their material well-being than those who strongly endorse PWE. The PWE (Weber, 2003) entails conformity to the importance of work, frugality and thrift (Furnham, 1984) thus stressing self-reliance (Miller et al., 2002) where all able bodied men should provide for themselves (Ciulla, 2000). With regard to employment status and material well-being this suggests that PWE's emphasis on self-reliance would spur those high on PWE to be less willing to consider public financial support (i.e. unemployment benefits) to be their legitimate right thus appraising their material situation as good „considering their situation“ and even feeling grateful for the support provided. By contrast those who have a weak PWE

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

may regard social benefits as rightfully theirs regarding it societies' responsibility to provide for their material needs and feeling no need to feel especially grateful for their material support. There is another reason to expect PWE to moderate the link between employment status and material satisfaction; since those who endorse strong PWE emphasize the importance of frugality, they may better be able to cope with the reduction in income associated with unemployment because of its ascetic nature (Bouma, 1973; Furnham, 1984; Schaltegger & Torgler, 2010). Those who have a strong PWE may therefore be more materially satisfied during unemployment than those who do not (strongly endorse PWE). Both theoretical accounts imply that:

(H2) The relationship between employment status and material well-being is moderated by PWE.

Method

Sample

Data was taken from the Icelandic sample of the 2008 European Values Study. In total, 397 participants were regarded as belonging to the labour force. Our analysis was restricted to the age range 31-60 years ($m^{age}=46.1$) to better accommodate those who can be considered a typical workers bearing the full burden of adult life and cost of living.

Measures

PWE was represented by five items commonly used for this purpose (Lepianka et al., 2010). They were; *job needed to develop talents*, *humiliating receiving money without working*, *people become lazy by not working*, *work is duty towards society* and *work always*

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

comes first, all measured on the scale 1 (strongly agree) to

5 (strongly disagree). Scoring was reversed so lower values represented a weak work value and higher values strong work value. Principal axis factor analysis indicated that one latent factor, that explained approximately 45% of the items variance should be extracted. Factor loadings were 0.58-0.71 and Cronbach's $\alpha=0.70$. Item responses were aggregated to form a scale score for the PWE with the minimum of 5 and maximum of 25, average of 13.75 and standard deviation of 3.40.

Material well-being was measured with two items; "*In general, how satisfied or dissatisfied are you with your standard of living?*" (85B) and "*How satisfied or dissatisfied are you with the financial outcome of your household?*" (85D). Both items were rated on a 1-10 scale with 1 representing "*dissatisfied*" and 10 "*satisfied*". The item responses were aggregated to form a scale score for material well-being, with the minimum of 2 and maximum of 20, average of 14.87 and standard deviation of 3.30. The correlation between the two items was $r=0.72$, $p<0.0001$.

To create the variable *Employment status*, respondents were split into two groups on the bases of their answers to the question "*Are you yourself gainfully employed at the moment or not?*" (Q111). Those who reported working more than 30 hours per week, less than 30 hours per week, self-employed, or in military service coded as 1 (employed) ($n=382$). Those who reported being unemployed were coded as 0 ($n=15$). All other values (retired or pensioned; doing housework; student and disabled) were coded as missing.

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

Results

First the relationship between employment status and material well-being was tested (hypothesis 1). The findings showed that the relationship was statistically significant ($b=4.72$, $p<.001$). Employed respondents were more satisfied with their material well-being than those who were unemployed. To test for the presence of a moderated relationship between employment status and material satisfaction, PWE and the product of PWE and employment status were entered into the regression equation. PWE scale scores were centered on the scale average (13.75) and multiplied with employment status to represent an interaction variable. The results showed significant interaction effects, $b=-0.74$; $p<0.01$. The slope for employment status was statistically significant ($b=5.07$, $p<.001$). Thus, subjects with an average PWE were less satisfied with their material status when they were unemployed than when they were employed (see Figure 1). To better explore the nature of the interaction effect, the PWE variable was centered on a value two standard deviations above the average value of the original scale (20.55). The new PWE variable (PWE_{high}) and the product of PWE_{high} and employment status were entered into the regression equation along with the employment status variable. Now the slope for the employment status was close to zero and not statistically significant ($b=0.02$, $p=.93$). Unemployed subjects with high PWE were therefore equally satisfied with their material well-being as employed subjects with high PWE (see Figure 1). Finally, the same calculations were done after centering the PWE variable on a value two standard deviations below the average value of the original scale (6.96). The regression analysis showed that the slope for employment status was $b=9.98$ and statistically significant¹ $p<.001$. Employed respondents with low PWE were less satisfied with their material well-being than unemployed respondents (see Figure 1).

¹ When household income was controlled for, the interaction between employment status and the PWE remained significant, $\beta=-0.50$; $p<0.05$

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE
RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

[Figure 1]

Discussion

The purpose of this study was twofold. First to directly test the relationship between employment status and material satisfaction. The results support previous assumptions about the negative effect of unemployment on material well-being. The second purpose was to test the moderating effect of PWE on the relationship between employment status and material satisfaction. The hypothesis regarding the buffering effect of PWE is also supported. Interestingly, there is an indication of *no difference* between the employed and unemployed for those who endorse strong PWE. Being committed to work has traditionally been regarded as a motivation to work (Paul & Moser, 2006). However the psychological cost of being highly committed to work with regards to material incentives on reemployment initiative and agency remains unclear. The material cost of being unemployed can be regarded as a strong motivational force for job search and thus the results invoke a motivational question regarding the normative pressure to work versus the material. There is an interesting motivational paradox inherent in the absence of difference in material well-being depending on one's employment is the case of those endorsing a very strong PWE. Similarly there seems to be a paradox applicable to those with low PWE (weak work commitment) and low material well-being (strong incentive to work).

Questions regarding work motivation among the unemployed are certainly legitimate, especially during periods of high unemployment (OECD, 2013) and when the psychological cost of being unemployed may not be as high as in times of low unemployment (Kolm, 2005). This paper thus provokes questions regarding motives for labour market choices.

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE
RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

References

- Bouma, G. D., 1973. A critical review of the recent “Protestant ethic” research. *Journal for the Scientific Study of Religion*, 12, 141–155.
- Ciulla, J. B., 2000. *The promise and betrayal of modern work*. New York: Three Rivers Press.
- Diener, E., Inglehart, R., and Tay, L., 2013. Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, 112, 497–527.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L., 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125, 276–302.
- Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S., 2010. The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 22463–22468.
- Frey, B., & Stutzer, A., 2002. What can economists learn from happiness research ? *Journal of Economic Literature*, 40, 402–435.
- Furnham, A., 1984. The protestan work ethic: A review of the psychological literature. *European Journal of Social Psychology*, 17, 87–104.
- Guichard, S., & Rusticelli, E., 2010. Assessing the impact of the financial crisis on structural unemployment in OECD countries.
- Hobfoll, S., Freedy, J., Green, B., & Solomon, S., 1996. Coping in reaction to extreme stress: The roles of resource loss and resource availability. In M. Zeidner & N. Endler (Eds.), *Handbook of coping: Theory, research, applications* (pp. 322–349). New York: Wiley.
- Kolm, A.-S., 2005. Work norms and unemployment. *Economics Letters*, 88, 426–431.

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

Lepianka, D., Gelissen, J., & van Oorschot, W., 2010. Popular Explanations of Poverty in Europe: Effects of Contextual and Individual Characteristics across 28 European Countries. *Acta Sociologica*, 53, 53–72.

Miller, M., Woehr, D. J., & Hudspeth, N., 2002. The meaning and measurement of work ethic: Construction and initial validation of a multidimensional inventory. *Journal of Vocational Behavior*, 60, 451–489.

OECD, 2013. OECD employment outlook 2013.

OECD, 2009. The jobs crisis: What are the implications for employment and social policy? (pp. 17–115). OECD employment outlook.

Paul, K. I., & Moser, K., 2006. Incongruence as an explanation for the negative mental health effects of unemployment: Meta-analytic evidence. *Journal of Occupational and Organizational Psychology*, 79, 595–621.

Schaltegger, C. A., & Torgler, B., 2010. Work ethic, Protestantism, and human capital. *Economics Letters*, 107, 99–101.

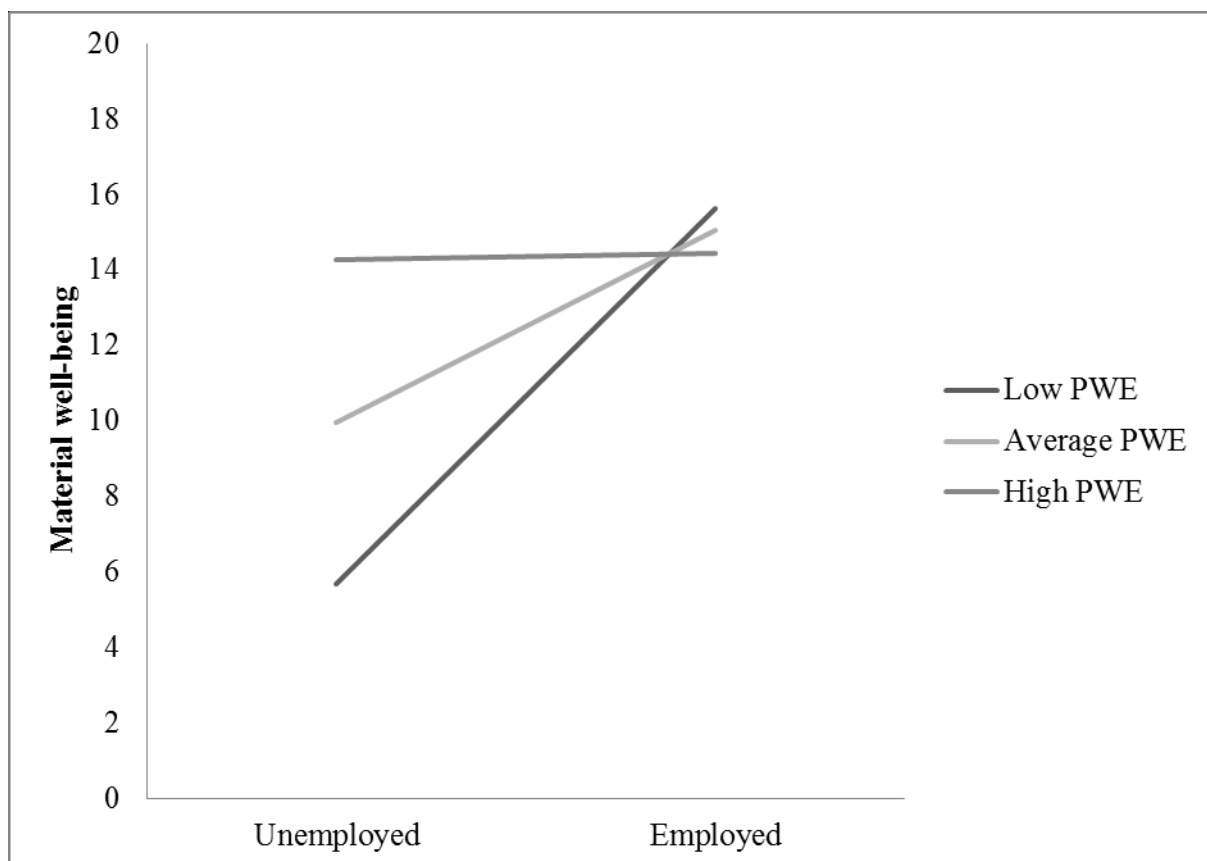
Ullah, P., 1990. The association between income, financial strain and psychological well-being among unemployed youths. *Journal of Occupational Psychology*, 63, 317–330.

Weber, M., 2003. The protestant work ethic and the spirit of capitalism. Mineola: Dover Publications.

THE MODERATING EFFECT OF PROTESTANT WORK ETHIC ON THE RELATIONSHIP BETWEEN EMPLOYMENT STATUS AND MATERIAL WELL-BEING

Figure Caption

Figure 1. The relationship between employment status and material well-being by levels of PWE.



**Five- or Seven-point Rating Scales: The Interplay between Verbal Labels and
Number of Response Options**

Arna Frimannsdóttir^{a1}

Fanney Thorsdóttir^a

^aUniversity of Iceland, Faculty of Psychology, Oddi, 101 Reykjavík, Iceland

Abstract

Researchers, using Likert rating scales, must decide on the number of response options to use. Response biases were used as a measure of data quality and two pairs of five and seven point scales were compared. The hypothesis was that when a five point scale is lengthened, in terms of the attitude response continuum, into a seven point scale the response bias reduces but when a five-point scale is transformed into a finer seven-point scale, with uneven intervals between points, the response bias increases. Results from a total of 1473 participants, each answering a questionnaire on only one of the four formats, partially supported the hypothesis. The response bias was reduced when the scale was made longer but no difference was found when the scale was made finer. Additional information did though indicate that the task difficulty increased when the five-point scale was transformed into a finer seven-point scale. Researchers must take this interplay between verbal labelling and number of response options into account when selecting a rating scale.

Keywords: Rating scales, number of response options, verbal labels, acquiescence, disacquiescence.

¹**Corresponding author:**

Arna Frimannsdóttir
University of Iceland
Faculty of Psychology
Oddi
101 Reykjavík
Iceland
Tel: +354 847 1117
e-mail: arf3@hi.is

Fully labelled bipolar rating scales also called Likert-type of scales are often used to measure subjective phenomena such as attitudes or personality traits in survey research. Since the format of such scales can affect the quality of data obtained (Andrews, 1984; Krosnick and Berent, 1993; Scherpenzeel and Saris, 1993; Weijters, Cabooter, and Schillewaert, 2010) survey researchers must make several decisions regarding the rating scale format. One major decision is on the number of response options.

A wide variety of research has been done on the optimal number of response options (Bendig, 1954; Birkett, 1986; Boote, 1981; Champney and Marshall, 1939; Givon and Shapira, 1984; Green and Rao, 1970; Jacoby and Matell, 1971; Jenkins and Taber, 1977; Kieruj og Moors, 2013; Lehmann and Hulbert, 1972; Lissitz and Green, 1975; Martin, 1973, 1978; Masters, 1974; McKelvie, 1978; Muniz, Garcia-Cueto and Lozani, 2005; Preston and Colman, 2000; Revilla, Saris og Krosnick 2014; Rosenstone, Hansen and Kinder, 1986; Schuman and Presser, 1981; Smith, 1994; Srinivasan and Basu, 1989; Warr, Barter and Brownridge, 1983; Watson, 1988; Wedell and Parducci, 1988: Wedell, Parducci and Lane, 1990). However, despite the large number of studies on this topic, most of them have focused on numerical rating scales. Moreover, findings from the few studies that have focused on comparing fully labelled scales have been mixed.

McKelvie (1978) compared five- and seven-point fully labelled scales and found no significant effect on test retest reliability estimate. It should though be mentioned that unipolar scales were used in McKelvie's study. Boote (1981) also compared five- and seven-point fully labelled scales. His results demonstrated that it is better to use a five-point scale than a seven-point scale, when reliability was measured either with the correlations of mean scale ratings, between test and retest, or with the proportion of completely consistent responses, between test and retest. Masters (1974) compared five and seven point fully labelled scales. His results, on the other hand, showed that when the total scores had low variation the data obtained with a seven-point scale had marginally higher internal consistency reliability than the data obtained with a shorter scale.

Evaluating the optimal number of response options becomes more complex when the rating scale is fully labelled than when the scale is numerical because of the interplay between the number of response options and the meaning of the verbal labels attached to the options. Increasing the number of response options can either produce a finer or a longer scale depending on the verbal labels that are attached to the new response options. Results from scaling studies (see for example Cliff, 1959; Jones and Thurstone, 1955; Spector, 1976),

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

where participants are asked to assign a numerical value to a verbal label, allow researchers to select appropriate verbal labels for a rating scale. If a five-point rating scale is transformed into a seven-point scale by adding two new verbal labels outside the scope of the five-point scale then the increase in the number of response options produces a longer scale. One can also make a five-point scale and a seven-point scale that cover the same width but then the two extra points make the seven-point scale a finer one.

Thus, one possible explanation for the inconsistent results in studies on the optimal number of response options on fully labelled scales is the interplay between the number of response options and verbal labels attached to the options. The seven-point scales in the studies conducted by McKelvie (1978) and Boote (1981) were finer than the five-point scales whereas the seven-point scales were longer than the five-point scales in Masters' (1974) study. Consequently, the meaning of the verbal labels used in these studies may have affected the relationship between number of response options and data quality. Since this may call into question the generalizability of studies focusing on the optimal number of response options on fully labelled scales, it is important to explore this issue further.

The main objective of the present study was to compare the quality of data obtained with a fully labelled five- and seven-point scales when the seven-point scale is either finer or longer in terms of the attitude response continuum. Previous research on this issue has evaluated data quality in terms of reliability. Reliability is a reasonable measure of data quality when only random errors are associated with the measurements. However, since measurements obtained with Likert-type of scales are susceptible to a host of systematic response biases, this assumption is questionable when such scales are used. Among the most well-known response biases associated with Likert-type of scales is the acquiescence response bias. Acquiescence is the tendency to agree or disagree (then called disacquiescence) with survey items irrespective of their content and research has shown that even small levels of acquiescence can have considerable effects on substantive findings (McCrae, Herbst, and Costa, 2001; Rammstedt, Goldberg, and Borg, 2010; Rammstedt and Kemper, 2011; Rammstedt, Kemper, and Borg, 2013; Soto, John, Gosling, and Potter, 2008; Ware, 1978; Winkler, Kanouse, and Ware, 1982). Thus, instead of focusing on reliability across scales, the present study will compare data quality in terms of acquiescence response bias.

Background

Initially, when researchers started to suspect that data quality depended on the number of response options, longer scales were expected to be better than shorter ones (Garner and

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Hake, 1951; Shannon and Weaver, 1949). These expectations were based on information theory, which argues that the amount of information a researcher can obtain, regarding respondents' subjective states is only limited by the number of response options provided. Later in an influential article, Miller (1956) combined results from a variety of research, exploring the number of unidimensional physical stimuli people can discriminate between, and came to the conclusion that increasing the number of stimuli increases the amount of information transmitted up to a certain degree. In most of these sensory studies this level-off point was around seven (Eriksen, 1954; Eriksen and Hake, 1955; Garner, 1953; Pollack, 1952). And Miller's (1956) conclusion was, based on peoples' span of absolute judgment, that it is possible to perceive and process 7 ± 2 stimuli. Miller (1956) further states that it is not only useless to add more stimuli after the level-off point, it is also damaging because it increases measurement error.

All studies in Miller's review used sensory stimuli, for example weight or brightness, and soon there were doubts as to whether they were generalizable when rating scales are used to measure more complex stimuli such as attitudes or personality (for an overview see Cox, 1980). Since, studies on the optimal number of response options for complex stimuli do not have information on the magnitude of the stimuli they focused on reliability. Thus, the question whether the reliability of the data is dependent on the number of response options became essential (Bendig, 1954; Birkett, 1986; Champney and Marshall, 1939; Givon and Shapira, 1984; Jacoby and Mattell, 1971; Jenkins and Taber, 1977; Komorita and Graham, 1965; Lissitz and Green, 1975; Martin, 1978; Master, 1974; Matell and Jacoby, 1971; Srinivasan and Basu, 1989).

Most of these studies compared numerical scales presumably because it was implicitly assumed that results on numerical scales also applied to measurements obtained with fully labelled scales. Later however, studies revealed that data, obtained with fully labelled scales, was of more quality than data obtained with numerical scales (Alwin and Krosnick, 1991; Krosnick and Berent, 1993; Zaller, 1988; Dickinson and Zellinger, 1980; Andrews, 1984; Wedell et al, 1990). It must therefore be presumed that the response process - the cognitive process that underlies selection of the response options - is different when using a fully labelled scale as opposed to a numerical one. The results from comparison studies using numerical scales may therefore not apply to fully labelled scales.

The few studies that have focused on the optimal number of response options on fully labelled scales have all compared five- and seven-point scales and have, up till now, produced mixed results. McKelvie (1978), who used the labels *not at all*, *barely*, *hard to say*, *quite*, and

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

highly, for the five-point scale and *not at all*, *barely*, *not very*, *hard to say*, *quite*, *very*, and *highly* for the seven-point scale found no differences in reliabilities across the two scale formats. Unfortunately, the sample in McKelvie's study was very small (30 undergraduate students) which might explain why his findings were inconclusive for the fully labelled scales. Boote (1981) used the verbal labels *extremely important*, *very important*, *somewhat important*, *slightly important*, and *not at all important* for the five-point scale and the verbal labels for the seven-point scale were: *extremely important*, *very important*, *quite important*, *somewhat important*, *moderately important*, *slightly important*, and *not at all important*. Boote (1981) results showed difference in reliability between the scales, in favour of the five-point scale. Finally, Master's (1974) used the verbal labels: *strongly agree*, *agree*, *undecided*, *disagree*, and *strongly disagree* for the five-point rating scale and for the seven-point scale the verbal labels chosen were: *very strongly agree*, *strongly agree*, *agree*, *undecided*, *disagree*, *strongly disagree*, and *very strongly disagree*. Masters' (1974) results showed a slightly higher reliability for the seven-point scale.

Clearly, the labels used in Boote's and Masters' studies are different in many respects. For instance, the labels in Boote's study define a unipolar scale whereas they define a bipolar scale in Masters' study and this may explain the contradictory results obtained. There is however, another possible explanation that is worthy of consideration. In Boote's study, the verbal labels attached to response options on the five- and the seven-point scales covered the same width on the response continuum. Thus, moving from five to seven options resulted in a finer scale. This was not the case in Masters' study where the response options on the five-point scale covered a narrower range of the response continuum; adding response options resulted in a longer scale. This may very well be the reason for the different results obtained in the two studies and if it is, the interplay between the number of response options and verbal labels must be taken into consideration in methodology research on the optimal number of response options on fully labelled scales.

In the studies discussed above, reliability estimates were used to assess data quality across scales. Implicit assumption in these studies is therefore, that as the reliability increases the data quality increases because of the reduction of the random measurement error. However, it is well known that response bias can enhance reliability of measurement (Gove and Geerken, 1977; Peer and Gamliel, 2011). An example of this is when a participant agrees with many of the study's statements without regard to their content, the reliability is inflated, given that the statements are all worded in the same direction. Thus, it is possible that measurements obtained with certain scales are more reliable, and therefore considered better,

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

when they simply produce more response bias. It is therefore, more suitable to use measures of response bias to evaluate data quality than reliability in methodological studies.

Response bias

A response bias is a systematic tendency to answer on a basis other than the content of items (Cronbach, 1946; Lentz, 1938; Paulhaus, 1991). When a response bias is stable over time and situations it is called response style but when it is a temporary reaction to situational demand it is considered a response set (Jackson and Messick, 1958; Wiggins, 1973). More specifically, response sets are defined as any tendency which causes respondents to consistently make different responses to survey questions than they would have made had the same content been presented in a question with a different form (Cronbach, 1946).

A major contributing factor in response set is task difficulty (Cronbach, 1946; Shulman, 1973). Thus, if question format increases the difficulty of respondents' task, response set will partly determine their response. If the scale's response options are too few, some participants may struggle with finding an option that corresponds to their attitude. This presumably increases the difficulty of the task of answering the questionnaire and therefore a response set is more likely to influence the response. The same applies if the response options are too many and the participants have difficulty with differentiating their meaning. This should increase the task difficulty and at the same time a response set is more likely to influence the response. When correspondence between the number of response options and the levels of attitude, or trait that respondents can discriminate between, is ensured the task of selecting a response option is as easy as possible. Under these circumstances the response bias should be reduced.

Many types of response biases have been identified, such as acquiescence/disacquiescence response bias, extreme response bias, social desirability response bias and neutral response bias (for an overview of response biases see Baumgartner and Steenkamp, 2001). One of the most studied response biases is the acquiescence/disacquiescence response bias (Baumgartner and Steenkamp, 2001; Paulhus, 1991).

Acquiescence/disacquiescence

Acquiescence is defined as the responder's tendency to agree to a statement or to answer a question with a "yes" unrelated to the question's content (Messick, 1991; Paulhaus, 1991;

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Ray, 1983). A variety of factors have been linked to acquiescence, such as personality traits (Couch and Kensiton, 1960) and demographic variables such as lower education level (Greenleaf, 1992) and lower income (Krosnick, 1999).

Acquiescence can have damaging consequences, that is bias the data which leads to incorrect results on the group difference which again reduces validity (Gove and Geerken, 1977). Acquiescence can also be damaging in scale development as it wrongfully enhances the relationship among items that are worded in the same direction to the disadvantage of the items that are worded in the opposite direction but are conceptually related (Bentler, Jackson and Messick, 1971; McCrae et al., 2001; Soto et al., 2008; Winkler, Kanouse and Ware, 1982); meaning that acquiescence can increase the magnitude of correlation for items that are positively related and decrease the relationship between items that are negatively correlated. Acquiescence can also be damaging when measuring predictive accuracy of a certain variable. If acquiescence is inherited in participant's answers for the predictor as well as the criterion then the predictive accuracy is inflated. These widespread damaging consequences underline the importance of assessing data quality with acquiescence.

As with other response biases there is disagreement regarding the reasons why participants show acquiescence in their answers. Some maintain that it is a reflection of a stable personality trait (Couch and Keniston, 1960) but studies have shown an inconsistency in acquiescence, either within the same questionnaire or between questionnaires which are either similar or dissimilar in content (Hui and Triandis, 1985; Husek, 1961; Nunnally and Husek, 1958; Ray, 1983). Another explanation for acquiescence is that the participant is satisficing (Krosnick 1991) when he goes through the stages of the response process (Tourangeau, 1984). In the first stage the participant interprets the question, in the second stage the participant searches his memory for information relating to the question's content. The third stage consists of intergration of all this information into one single conclusion and in the fourth, and final stage, the participant must translate his conclusion to a response, or as with rating scales to a single response option. Tourangeau (1984) assumes that all participants go through all four stages in order to optimize. Krosnick's (1991) satisficing theory challenges the assumptions that all participants optimize at all times. Krosnick (1991) introduced the possibility that some participants may only satisfice in their answers and he furthermore introduced a difference between a weak and strong satisficing. In weak satisficing the participants goes through all four stages of the response process (Tourangeau, 1984) but with less effectiveness than when optimizing occurs (Krosnick, 1991). When a participant processes a question he searches for reasons why the question is correct, or why he should

agree with it, and as most statements in a questionnaire are reasonable it is likely that the participant will easily come up with reasons to agree to the statement (Krosnick, 1991). Krosnick (1991) assumes that acquiescence can also be a strong satisficing where stages two and three, in the answering process, are skipped by the respondent but in those stages the respondent recalls relevant information that can help him answer the question and gather the information into a response (Tourangeau, 1984). According to Krosnick's (1991) satisficing theory, task difficulty can greatly influence whether participants optimize or only satisfice in their answers. In relation to the focus of the present study, the scale format, the number of response options and verbal labels, that has the most task difficulty will result in more participants showing acquiescence in their answers as opposed to scales with easier format. The reason why a participant, strongly satisficing, uses acquiescence, instead of some other response set, could be that social conventions imply that we should be polite and to agree with someone is more polite than disagreeing (Brown and Levinson, 1987; Leech, 1983).

Disacquiescence is, opposite to acquiescence, the tendency to disagree with a statement without regards to the item's content (Baumgartner and Steenkamp, 2001; Couch and Keniston, 1960). Usually disacquiescence is regarded as a less common version of acquiescence (Baumgartner and Steenkamp, 2001; Weijters, Baumgartner and Schillewaert, 2013) but these two response biases are though often considered among the most common ones (Hoffman, Mai and Cristescu, 2013). As with acquiescence, disacquiescence can seriously reduce the data quality. It can be presumed that the reason for disacquiescence, as with acquiescence, is satisficing. But why some employ disacquiescence and other acquiescence or other response sets, has not been studied extensively.

Summary and hypotheses

Evidence suggests that data obtained with fully labelled rating scales are of better quality than data obtained with numerical scales (Alwin and Krosnick, 1991; Andrews, 1984; Dickinson and Zellinger, 1980; Krosnick and Berent, 1993; Wedell et al, 1990; Zaller, 1988). It is therefore recommended to use fully labelled scales in survey research (Krosnick and Fabrigar, 1997). One decision survey researchers face, when using fully labelled scales, is how many response options they should attach to the rating scale. The studies conducted so far on this issue, have focused on five- and seven-point scales (Boote, 1981; McKelvie, 1978; Masters, 1974). These studies do not arrive at the same conclusion. One possible explanation for these mixed results is the interplay between the number of response options and verbal labels

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

attached to the options. In some of the studies the seven-point scales were finer than the five-point scales (Boote, 1981; McKelvie, 1978) whereas the seven-point scales were longer than the five-point scales in others (Masters, 1974). Consequently, the change in task difficulty may not be equivalent when respondents move from five- to seven-point scales in these studies.

In the present study the interplay between the number of response options and the meaning of the verbal labels attached to the options will be explored. It will therefore give new insight into comparison studies on scale lengths. Two pairs of fully labelled five and seven point scales were compared. In one of the pairs, two new options were added on either side of the five-point scale (five-point narrow) to produce the seven-point scale (seven-point long) which then covered the full attitude continuum. Since the verbal labels on the five-point scale do not cover the entire response dimension, the task of selecting a response option should be more difficult for respondents. It is therefore hypothesized that when the scale is lengthened the dis/acquiescence response bias is reduced with increase in response options. The other pair involves making a finer scale by adding two new response options to a five-point scale (five-point wide). It is not as clear in this condition whether the task is easier or more difficult when response options are added to the five-point scales. However on the basis of Boote's (1981) conclusions the hypothesis is that when the scale is made finer the dis/acquiescence response bias is increased.

Method

Participants

The data came from the prerecruited probability-based Capacent panel in Iceland. The data was collected by Capacent via the Internet. The panel was selected from a sampling frame that included all Icelanders 16 to 75 years old who had internet access. In total 2400 respondents were randomly selected from the panel in each of the four following experimental conditions: 1) “five-point narrow scale” group, 2) “5-point wide scale” group, 3) “7-point fine scale” group and 4) “7-point long scale” group. Overall, 1473 respondents took part in the survey (a response rate of 61.4%).

Measurements

Items. All participants answered 26 questions measuring *environmental attitudes*, *attitudes towards the European Union* and *anger*. Five questions measured attitude towards the environment, where higher value on questions one and five corresponded to a more negative attitude towards environmental issues and higher value on questions two, three and four corresponded to a more positive attitude towards environmental issues. An example of a question is *The fight against pollution is not as urgent as often thought*. The reliability of the questionnaire (Cronbach’s alpha) was between 0.715 and 0.780 for the four groups. Nine questions measured attitudes towards the European Union. A higher value on the first four questions corresponded to a more negative attitude towards the European Union and a higher value on the latter five questions corresponded to a more positive attitude towards the European Union. An example of a question is *Iceland’s entry into the European Union would result in higher living standard in Iceland*. The reliability of the questionnaire was between 0.879 and 0.895 for the four groups. *Anger* was measured with 12 statements from the Siegel (1986) anger inventory. All the questions were worded in the same direction and a higher value corresponded to a less degree of anger and an example of a question is *I get angry when forced to work with incompetent people*. The reliability of the questionnaire was between 0.796 and 0.844 for the four groups.

The wording and sequence of the 26 questions were the same for all respondents. The questions measuring *environmental attitude*, *attitude towards the European Union* and *anger* were presented in different sections. First, questions measuring *environmental attitudes*, then

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

questions measuring *attitude towards the European Union* were presented in proximity to each other and finally questions measuring *anger*. For all scales respondents were asked to indicate how much they agreed or disagreed with each statement.

Acquiescence response bias was measured using pairs of items where each pair contained two questions worded in opposite direction. One point was given if a responder agreed to both statements within the same pair (Winkler, Kanouse, and Ware, 1982). As recommended by Swain, Weather, and Niedrich (2008) the points were weighted. Agreeing with the most extreme options produced three points, two points were given if the most extreme and the second most extreme were selected, and one point was given for agreeing with the second most extreme option on both statements within the same pair.

It proved to be difficult to make up pairs consisting of two questions which had similar content and were worded in opposite direction but only one pair consisted of items with the same question, worded in opposite directions, question 4 (*The influence of Icelanders in the international forum would increase if Iceland joins the European Union*) and 8 (*The influence of Icelanders in the international forum would decrease if Iceland joins the European Union*) in the European Union questionnaire, see all questions in appendix A. In the end the following six pairs were used: questions 1 and 3 in the environmental questionnaire, questions 4 and 5 in the environmental questionnaire and the following questions in the European Union questionnaire: questions 1 and 5, question 2 and 6, questions 3 and 7 and questions 4 and 8. As mentioned before, the questions were paired based on the fact that they were worded in opposite direction and it was also ensured that each pair consisted of questions from the same questionnaire.

The highest possible score on acquiescence measure was therefore 18 points and lowest zero point. To be able to compare the five-point narrow scale and the seven-point long scale the second most extreme option was defined as response option number two and three (mainly agree and agree a bit) for the seven-point scale but as response option number two (agree a bit) for the five-point scale. Similarly, to compare the five-point wide scale and the seven-point fine scale the second most extreme option was defined as a response option number two and three (somewhat agree and agree a bit) for the seven-point scale but as the response option number two (somewhat agree) for the five-point scale.

Disacquiescence response bias was measured in the same way but using the disagree-options.

Procedure

The survey was conducted from January 13 to January 28, 2009 and was sent to all active panel members. The Capacent panel is based on probability sampling. However, while the panel is expected to better represent the general public than non-probability panels (Dillman, Smyth, and Christian, 2009), a number of factors may cause the panel sample to differ from a simple random sample of the population.

Respondents were randomly assigned to four experimental conditions regarding the verbal labels provided with each question. Results of a scaling study conducted in Iceland were used to select the verbal labels (Þórsdóttir and Jónsson, 2009). It was therefore possible to ensure that the verbal labels on the “5-point narrow scale” did not cover the full response range whereas the labels on the other three scales did. In all groups the verbal labels divided the response continuum into approximately equally large intervals. For the “5-point narrow scale” group the following labels were selected: “mainly agree” (M=4), “agree a bit” (M=2), “neither agree nor disagree” (M=0), “disagree a bit” (M=-2) and “mainly disagree” (M=-4). For the “5-point wide scale” group the following labels were selected: “totally agree” (M=6), “somewhat agree” (M=3), “neither agree nor disagree” (M=0), “somewhat disagree” (M=-4) and “totally disagree” (M=-7). For the “seven-point long scale” the following labels were selected: “totally agree” (M=6), “mainly agree” (M=4), “agree a bit” (M=2), “neither agree nor disagree” (M=0), “disagree a bit” (M=-2), “mainly disagree” (M=-4) and “totally disagree” (M=-7). For the “seven-point fine scale” the following labels were selected: “totally agree” (M=6), “somewhat agree” (M=3), “agree a bit” (M=2), “neither agree nor disagree” (M=0), “disagree a bit” (M=-2), “somewhat disagree” (M=-4) and “totally disagree” (M=-7) (Þórsdóttir and Jónsson, 2009).

Results

Descriptive statistics

Table 1 shows response rate, which varies between 60.5% for the 7-point wide scale to 62.5% for the 5-point wide scale. Table 1 also shows participants age division for each scale. The largest age group varies between scales but it is always above 35 years of age. Participant's

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

education division for each scale, can also be seen in table 1. The largest education group is nearly always those who have a university degree, but most of participants answering the 5-point wide scale have only finished compulsory education and some additional education.

[Table 1]

Table 2 shows descriptive statistics for the response styles: acquiescence and disacquiescence, for each scale. The range for acquiescence between scales is comparable but the means vary with the mean for 5-p narrow being the highest. The range and means for disacquiescence vary more between the scales, but the highest mean is again for the 5-p narrow scale. The values for skewness show that all the distributions have some deviation from a normal distribution in that they are all skewed to the right. The values for kurtosis show, similarly, some deviation from a normal distribution, some are lower than a normal distribution and some are higher.

[Table 2]

Hypotheses testing

As expected, comparison of acquiescence between the five and seven point scale revealed a decrease in acquiescence when the seven point scale ($M=2.46$, $Sd=2.15$) was longer than the five point scale ($M=3.38$, $Sd=2.29$), $t(670)=5.327$; $p<0.001$. No difference was found between the five point wide scale ($M=2.51$, $Sd=2.12$) and seven point fine scale ($M=2.42$, $Sd=2.04$), $t(656)=0.556$; $p=0.578$. Figure 1 shows mean-acquiescence by number of points and verbal labelling.

[Figure 1]

Comparison on disacquiescence, between five and seven point scales, also revealed a significant decrease in disacquiescence when the five-point narrow scale ($M=2.78$, $Sd=2.11$) scale was lengthened to a seven-point long scale ($M=1.93$, $Sd=1.92$), $t(670)=5.482$; $p<0.001$. No difference was found between the five-point wide scale ($M=2.04$, $Sd=1.88$) and the seven-point fine scale ($M=2.14$, $Sd=1.70$), $t(656)=-0.707$; $p=0.480$.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Figure 2 demonstrates the decrease in disacquiescence when the scale is lengthened but there is little difference in the mean disacquiescence when the scale is refined.

[Figure 2]

Discussion

The main objective of the present study was to compare the quality of data obtained with fully labelled five- and seven-point scales when the seven-point scale is either finer or longer in terms of the attitude response continuum.

It was predicted that when a five point rating scale is lengthened to a seven-point rating scale response bias would be reduced. The results supported this hypothesis in that both acquiescence and disacquiescence were significantly lower for the seven-point long scale than the five-point narrow scale. These results are in line with Master's (1974) findings that the data, produced with a long seven-point rating scale, is of better quality than the data produced with a narrow five-point rating scale. A likely explanation for these findings is that the task of selecting a response option is easier for respondents when they use a seven-point scale that covers the whole response continuum than a five-point scale that covers a narrower range of the continuum.

For the refinement pair, a five-point wide scale and a seven-point fine scale, it was predicted that the response bias would increase with an increase in number of points. The results failed to reveal a difference in either acquiescence or disacquiescence for the refinement pair. A possible explanation for the failure to reveal a difference could be the response rate. The lowest response rate, for all scales, was for the seven-point fine scale and it is possible that unmotivated participants, likely to employ a response style, were more likely to not answer the seven-point fine scale than other scales. If those participants had answered the questionnaire a difference in response styles between the five-point wide scale and seven-point fine scale could have been found.

The results are not in agreement with Boote's (1981) results where a five-point wide scale showed higher reliability than a seven-point scale. In addition to the reason outlined above, a possible reason for these inconsistent results on comparison between equally long five- and seven-point rating scales is that the rating scales used in Boote's (1981) study were unipolar whereas the scales used in this study were bipolar. The difference in the meaning of verbal labels chosen for a unipolar scale is likely to be less than the difference between meanings of verbal labels chosen for a bipolar scale as the range is narrower for the unipolar scale. When the difference in verbal label's meaning is unclear the task is more difficult (Krosnick and Fabrigar, 1997) and should therefore be more difficult when the response options are seven than when they are five. Another reason for the conflicting results could be that Boote (1981) used reliability as a measure of data quality which, as mentioned above, can

increase even though there is a decrease in data quality (Gove and Geerken, 1977; Peer and Gamliel, 2011).

Implications

The main implication of the findings is that survey methodologists must take in to account the interplay between verbal labels and the number of response options when studying the optimal number of response options on fully labelled scales, as the findings showed that the question of whether it was better to use a five- or a seven-point scale could not be answered without considering the verbal labels presented to respondents. Survey researchers must keep this in mind when using findings from methodological studies to decide on the number of response options on fully labelled rating scales. Moreover, conflicting results of previous studies focusing on this issue could at least partly be explained by the fact that in some studies the seven-point rating scale was longer, in terms of the attitude continuum, than the five-point rating scale and in some the seven-point rating scale was finer, in terms of the attitude continuum, than the five-point rating scale. Thus, the findings shed new light on previous findings and raise significant questions for further research on the optimal number of response options on fully labelled scales. They must, furthermore, focus on fully labelled scales but at the same time be aware of the location of it's verbal labels on the attitude continuum.

An additional implication from this study is the undermining of reliability as a measure of data quality when comparing rating scales of different lengths, as the results from this study suggests that as task difficulty increases so does systematic response bias. An increase in systematic response bias can produce an increase in reliability and researchers should therefore refrain from using reliability as a measure of data quality unless ensuring that the systematic response bias is the same between scales.

Turning to the theoretical implications of the findings, the most important implication is for satisficing theory (Krosnick, 1991). Satisficing theory states that as task difficulty increases so does the likelihood of satisficing. One form of satisficing is dis/acquiescence response bias and since the task of selecting a response option from the five-point narrow scale is more difficult than selecting an option from the seven-point scale, satisficing theory predicts that respondents are more likely to demonstrate dis/acquiescence responding when presented with the five-point scale. The findings supported this prediction; in that both disacquiescence and acquiescence responses were more likely when respondents used the five-point scale.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Finally, even though it is impossible from this study to determine which scale length produces data with the most quality the results indicate that it is safer to use a seven-point rating scale than the five-point rating scale. However, since the research design did not include a seven-point narrow scale, it is impossible to know how that scale type would compare to the five-point wide scale. In any case it is recommended that the verbal labels selected, cover the entire response continuum. Further comparison studies are needed on the interplay between number of response options and verbal labels in order to better advise survey researchers on the optimal response number. They could for instance include comparisons between a five-point wide scale and a seven-point narrow scale and also a comparison between a nine- and a seven-point scale.

References

- Alwin, D. F. and Krosnick, J. A. (1991). The Reliability of Survey Attitude Measurement The Influence of Question and Respondent Attributes. *Sociological Methods & Research*, 20(1), 139-181.
- Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48, 409-442.
- Baumgartner, H. and Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143-156.
- Bendig, A. W. (1954). Reliability and the Number of Rating Scale Categories. *Journal of Applied Psychology*, 38, 38-40.
- Bentler, P. M., Jackson, D. N., and Messick, S. (1971). Identification of content and style: a two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76(3), 186.
- Birkett, N. J. (1986). Selecting the Number of Response Categories for a Likert-type Scale. In *Proceedings of the American Statistical Association* (pp. 488-492).
- Boote, A. S. (1981). Reliability testing of psychographic scales: five-point or seven-point? Anchored or labeled? *Journal of Advertising Research*, 21, 53-60.
- Brown, P. and Levinson, S. C. (1987). *Politeness: Some Universals in Language Use*. New York: Cambridge University Press.
- Champney, H. and Marshall, H. (1939). Optimal refinement of the rating scale. *Journal of Applied Psychology*, 23, 323-331.
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66, 27-44.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Couch, A. and Keniston, K. (1960). Yeasayers and naysayers: agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 60, 151-174.
- Cox, E. P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research*, 17, 407-422.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational & Psychological Measurement*, 6, 475-494.
- Dickinson, T. L. and Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of applied Psychology*, 65, 147-154.
- Dillman, D. A., Smyth, J. D. and Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. New Jersey: Hoboken.
- Eriksen, C. W. (1954). The case for perceptual defense. *Psychological review*, 61, 175.
- Eriksen, C. W. and Hake, H. W. (1955). Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *Journal of Experimental Psychology*, 49, 323-332.
- Fanney Þórsdóttir and Friðrik H. Jónsson. (2009). Gildun á mælistikum: Hvaða orðagildi er best að nota á svarmöguleika mælikvarða? In Ragnar P. Ólafsson, Árni Kristjánsson and Einar Guðmundsson (Eds.), *Sálfræðiping 2009: Ráðstefnurit*. Reykjavík: Sálfræðingafélag Íslands.
- Garner, W.R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, 46, 373-380.
- Garner, W. R. and Hake, H. W. (1951). The amount of information in absolute judgements. *Psychological Review*, 58, 446-459.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Givon, M. M. and Shapira, Z. (1984). Response to rating scales: a theoretical model and its application to the number of categories problem. *Journal of Marketing Research*, 21, 410-419.
- Gove, W. R. and Geerken, M. R. (1977). Response bias in surveys of mental health: an empirical investigation. *American Journal of Sociology*, 82, 1289-1317.
- Green, P. E. and Rao, V. R. (1970). Rating scales and information recovery – How many scales and response categories to use? *Journal of Marketing*, 34, 33-39.
- Greenleaf, E. A. (1992). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research*, 29, 176-88.
- Hoffmann, S., Mai, R., and Cristescu, A. (2013). Do culture-dependent response styles distort substantial relationships? *International Business Review*, 22, 814-827.
- Hui, C. H. and Triandis, H. C. (1985). The instability of a response sets. *Public Opinion Quarterly*, 49, 253-260.
- Husek, T. R. (1961). Acquiescence as a response set and as a personality characteristic. *Educational and Psychological Measurement*, 21, 295-307.
- Jackson, D. N. and Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243-252.
- Jacoby, J. and Matell, M.S. (1971). Three-point Likert scales are good enough. *Journal of Marketing Research*, 7, 495-500.
- Jenkins, G. D. and Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, 62, 392-398.
- Jones, L. V. and Thurstone, L. L. (1955). The psychophysics of semantics: An experimental investigation. *The Journal of Applied Psychology*, 39, 31-36.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Kieruj, N. D. and Moors, G. (2010). Variations in response style behaviour by response scale format in attitude research. *International Journal of Public Opinion Research*, 22, 320-342.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Krosnick, J. A. and Berent, M. K. (1993). Comparisons of party identification and policy preferences: the impact of survey question format. *American Journal of Political Science*, 37, 941-964.
- Krosnick, J. A. and Fabrigar, L. R. (1997). Designing ratings scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141-156). New York: Wiley.
- Leech, G. N. (1983). *Principles of Pragmatics*. London/New York: Longman.
- Lehmann, D. R. and Hulbert, J. (1972). Are three-point scales always good enough? *Journal of Marketing Research*, 9, 444-446.
- Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin*, 35, 659.
- Lissitz, R. W. and Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60, 10-13.
- Nunnally, J. and Husek, T. R. (1958). The phony language examination: An approach to the measurement of response bias. *Educational and Psychological Measurement*, 18, 275-282.
- Martin, W. S. (1973). The effects of scaling on the correlation coefficient: A test of validity. *Journal of Marketing Research*, 10, 316-318.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Martin, W. S. (1978). Effects of scaling on the correlation coefficient: Additional considerations. *Journal of Marketing Research*, 15, 304-308.
- Masters, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement*, 11, 49-53.
- McCrae, R. R., Herbst, J. H. and Costa Jr, P. T. (2001). Effects of acquiescence on personality factor structures. *Personality and temperament: Genetics, evolution, and structure*, 217-231.
- McKelvie, S. J. (1978). Graphic rating scales-How many categories. *British Journal of Psychology*, 69, 185-202.
- Messick, S. (1991). Psychology and methodology of response styles. In R. E. Snow and D. E. Wiley (Eds.), *Improving inquiry in social science* (pp. 161-200). Hillsdale, NJ: Erlbaum.
- Miller, G. A. (1956). The magical number of seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81-97.
- Muniz, J., García-Cueto, E., and Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences*, 38, 61-69.
- Paulhus, D. L. (1991) Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, and L. S. Wrightman (Eds.), *Measures of Social Psychological attitudes Series* (pp. 17-59). San Diego, CA: Academic Press.
- Peer, E. and Gamliel, E. (2011). Too Reliable To Be True? Response Bias as a Potential Source of Inflation in Paper-And-Pencil Questionnaire Reliability. *Practical Assessment, Research and Evaluation*, 16 (8), 1-8.
- Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America*, 24, 745-749.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Preston, C. C. and Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1–15.
- Rammstedt, B., Goldberg, L. R. and Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44, 53-61.
- Rammstedt, B. and Kemper, C. J. (2011). Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias. *Journal of Research in Personality*, 45, 121-125.
- Rammstedt, B., Kemper, C. J. and Borg, I. (2013). Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study. *European Journal of Personality*, 27, 71-81.
- Ray, J. J. (1983). Reviving the problem of acquiescence response bias. *Journal of Social Psychology*, 121, 81-96.
- Revilla, M. A., Saris, W. E., and Krosnick, J. A. (2014). Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*, 43, 73-97.
- Rosenstone, S. J., Hansen, J. M., and Kinder, D. R. (1986). Measuring change in personal economic well-being. *Public Opinion Quarterly*, 50, 176-192.
- Scherpenzeel, A. and Saris, W. (1993). The evaluation of measurement instruments by meta-analysis of multitrait-multimethod studies. *Bulletin de Methodologie Sociologique*, 39(1), 20-44.
- Schuman, H. and Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Information*. Chicago: University of Illinois Press.
- Shulman, A. (1973). A comparison of two scales on extremity response bias. *Public Opinion Quarterly*, 37, 407-412.
- Siegel, J. M. (1986). The multidimensional anger inventory. *Journal of personality and social psychology*, 51(1), 191.
- Smith, T. W. (1994). A comparison of two confidence scales. *GSS Methodological Report No. 80*. Chicago, IL: National Opinion Research Center.
- Soto, C. J., John, O. P., Gosling, S. D. and Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94, 718.
- Spector, P. E. (1976). Choosing response categories for summated rating scales. *Journal of Applied Psychology*, 61, 374-375.
- Srinivasan, V. and Basu, A. K. (1989). The metric quality of ordered categorical data. *Marketing Science*, 8, 205-230.
- Swain, S. D., Weathers, D. and Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, 45, 116-131.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.
- Ware, J. W. (1978). Effects of acquiescence response sets on patient satisfaction ratings. *Medical Care*, 16, 327-336.
- Warr, P., Barter, J., and Brownridge, G. (1983). On the interdependence of positive and negative affect. *Journal of Personality and Social Psychology*, 44, 644-651.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

- Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology*, 55, 128-141.
- Wedell, D. H. and Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology*, 55, 341-356.
- Wedell, D. H., Parducci, A., and Lane, M. (1990). Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and types of anchors. *Journal of Personality and Social Psychology*, 58, 319-329.
- Weijters, B., Baumgartner, H., and Schillewaert, N. (2013). Reversed Item Bias: An Integrative Model. *Psychological Methods*, 18, 320-334.
- Weijters, B., Cabooter, E., and Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236-247.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading MA: Addison-Wesley.
- Winkler, J. D., Kanouse, D. E., and Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67, 555.
- Zaller, J. (1988). "Vague Minds vs. Vague Questions: An Experimental Attempt to Reduce Measurement Error." Presented at the annual meeting of the American Political Science Association, Washington, DC.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Caption for table 1

Table 1. Participant's information for each scale.

	5-p narrow	7-p long	5-p wide	7-p fine
Response rate	60.8%	61.7%	62.5%	60.5%
Age division				
16-24 years	19.1%	19.2%	19.1%	19.1%
25-34 years	18.1%	18.0%	20.9%	18.6%
35-44 years	24.1%	20.1%	18.9%	21.5%
45-54 years	18.7%	24.4%	19.2%	17.7%
55-75 years	20.0%	18.3%	22.0%	23.0%
Education division				
Comp	6.5%	4.6%	4.0%	3.6%
Comp+	26.5%	32.9%	29.8%	32.3%
Secondary education	21.1%	20.7%	25.4%	16.5%
Secondary+	11.0%	9.6%	10.1%	12.0%
University degree	33.4%	31.3%	29.3%	34.4%
Unsure	1.5%	0.8%	1.3%	1.1%

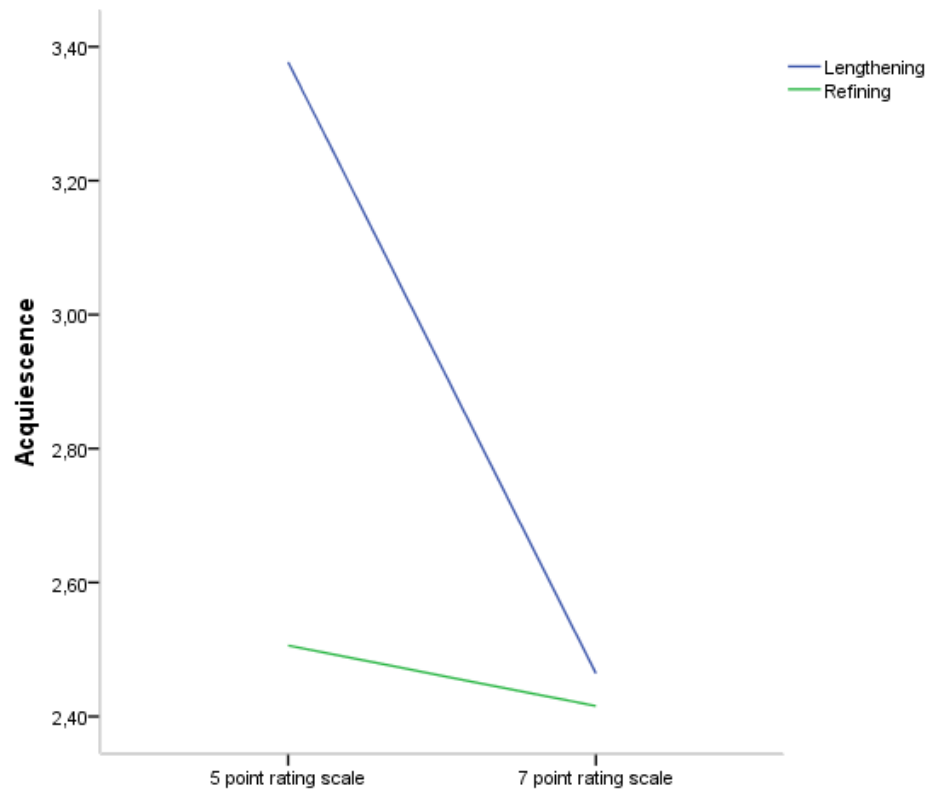
FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Caption for table 2

Table 2. Descriptive statistics for acquiescence and disacquiescence, for each scale.

	N	Min	Max	M	Sd	Skewness	S.E	Kurtosis	S.E
Acquiescence									
5-p narrow	334	0	11	3.38	2.29	.372	.133	-.291	.266
7-p long	338	0	10	2.46	2.15	.712	.133	-.076	.265
5-p wide	338	0	10	2.51	2.12	.790	.133	.495	.265
7-p fine	320	0	10	2.42	2.04	.785	.136	.489	.272
Disacquiescence									
5-p narrow	334	0	9	2.78	2.11	.533	.133	-.396	.266
7-p long	338	0	9	1.93	1.92	1.189	.133	1.192	.265
5-p wide	338	0	12	2.04	1.88	1.387	.133	3.166	.265
7-p fine	320	0	10	2.14	1.70	.876	.136	1.033	.272

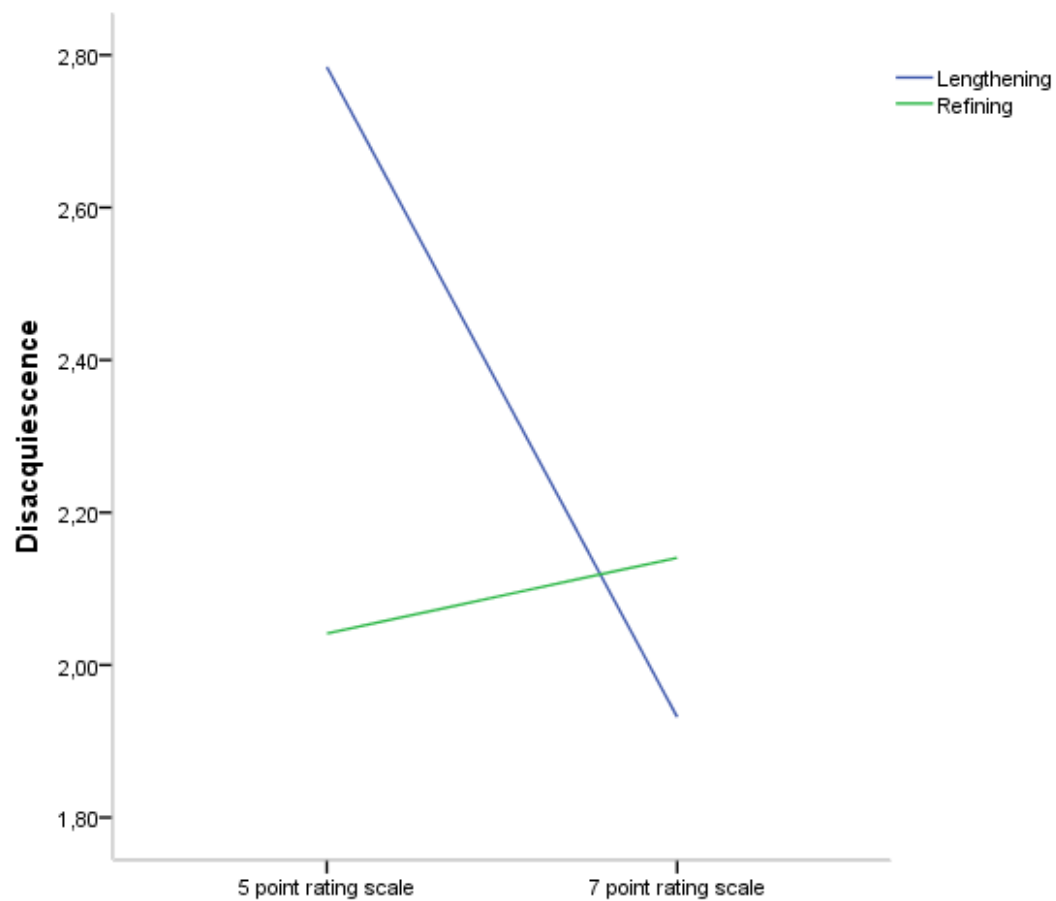
FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS



Caption for figure 1

Figure 1. Acquiescence by number of response options and verbal labelling

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS



Caption for figure 2

Figure 2. Mean disacquiescence by number of response options and verbal labelling.

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

Appendix A

Environmental issues questionnaire

- 1 .I would agree to tax increases should the money be used to prevent environmental pollution?
- 2 .If we want to increase prosperity in Iceland we must accept environmental problems?
3. Fight against pollution is not as urgent as often thought?
4. The government should put emphasis on power consuming heavy industry?
5. In the coming years improvements in environmental issues should have priority over attempts to increase economical growth?

European Union questionnaire

1. If Iceland joins the European Union it would result in lower interest rate in Iceland?
2. If Iceland joins the European Union it would result in lower food prices in Iceland?
3. If Iceland joins the European Union it would result in higher living standard in Iceland?
4. The influence of Icelanders in the international forum would increase if Iceland joins the European Union?
5. Demands for fishing right within Icelandic fisheries jurisdiction would increase if Iceland joins the European Union?
6. Are Icelanders forfeiting custody over their natural recourses if Iceland joins the European Union?
7. Icelanders forfeit their independence if Iceland joins the European Union?
8. The influence of Icelanders in the international forum would decrease if Iceland joins the European Union?
9. Membership in the European Union would be expensive for Icelanders?

Anger questionnaire

1. I can get irritated by people who are simply near me?
2. I get angry when forced to work with incompetent people?

FIVE- OR SEVEN-POINT RATING SCALES: THE INTERPLAY BETWEEN
VERBAL LABELS AND NUMBER OF RESPONSE OPTIONS

3. Habits some of my friends have can both irritate and annoy me greatly?
4. When I get angry my anger will take a long time to subside?
5. I get angry when I get orders from those more incompetent than I?
6. Sometimes I feel that I get angry for no reason?
7. I am wary when people show me unexpected kindness?
8. So called specialists I have met are no better than I?
9. I get angry when people fail me?
10. I get angry when people show injustice?
11. I get angry if I do not get thanks for accomplished work?
12. I can get angry just thinking about something in the past?