

Gildun á mælistikum: Hvaða orðgildi er best að nota á svarmöguleika mælikvarða?

Fanney Þórsdóttir og Friðrik H. Jónsson
Sálfræðideild, Háskóli Íslands

Útdráttur

Gildunarannsókn var gerð þar sem þátttakendur voru beðnir um að meta merkingu íslenskra orðagilda sem hægt er að nota á mælistiku matskvarða. Úrtak 598 Íslendinga var tekið úr 10.000 manna netpanel sem valinn var með tilviljunaraðferð úr þjóðskrá. Alls tóku 398 þeirra þátt og var svarhlutfall því 65,9%. Þátttakendur, sem voru á aldrinum 17 til 73 ára, mátu 96 orð og orðasambönd. Erlendar rannsóknir á merkingu orðagilda voru hafðar til hliðsjónar þegar orðagildi voru valin en einnig var leitað fanga í íslenskum orðabókum. Á grundvelli niðurstaðna er hægt að velja orðagildi sem hafa nákvæma merkingu og spanna þá viðhorfavídd sem verið er að mæla með jöfnu millibili.

Matskvarðar (*rating scales*) eru vinsæl aðferð til að mæla viðhorf (Krosnick og Fabriger, 1997). Þegar slíkir kvarðar eru notaðir taka svarendur afstöðu til fullyrðingar með því að velja svarmöguleika af mælistiku og gefa þeir til kynna stefnu og styrkleika viðhorfs. Rannsóknir á matskvörðum hafa meðal annars beinst að því hvort betra sé að merkja alla svarmöguleika mælistiku með orðagildum eins og *frekar sammála/ósammála* og *mjög sammála/ósammála* eða einungis endamöguleika stikunnar. Niðurstöður benda til að matskvarðar séu áreiðanlegri þegar allir svarmöguleikar eru merktir með orðagildum (Alwin og Krosnick, 1991; Bending, 1953; Boote, 1981; Peter og McCormick, 1966). Almennt er því mælt með notkun orðagilda þegar matskvarðar eru notaðir til að mæla viðhorf (Krosnick og Fabriger, 1997; Schaeffer og Presser, 2003).

Ávinningur af því að merkja alla svarmöguleika með orðagildum hlýtur þó að vera háður því að viðeigandi orðagildi séu valin og er þá einkum litið til þriggja þátta (Jones og Thurstone, 1955; Krosnick og Fabriger, 1997; Schaeffer og Presser, 2003). Í fyrsta lagi þarf merking orðagilda að vera nákvæm (Krosnick og Fabriger, 1997) til að draga úr

tilviljunarkenndri villu í mælingum. Nákvæm merking orðagilda ætti að draga úr tilviljunarkenndri villu í mælingu og auka þannig áreiðanleika mælinga. Í annan stað er talið mikilvægt að jafnlangt bil sé á milli orðagilda á mælistiku til að mælingar hafi eiginleika jafnbilakvarða (Krosnick og Fabriger, 1997; Schaeffer og Presser, 2003). Í þriðja lagi skiptir máli að orðagildin spanni alla viðhorfavíddina og lýsi allt frá mjög neikvæðum viðhorfum til mjög jákvæðra viðhorfa (Jones og Thurstone, 1955).

Erlendis hefur töluleg merking orðagilda verið metin í svokölluðum gildunarannsóknum (*scaling studies*) (sjá til dæmis Cliff, 1959; Jones og Thurstone, 1955; Krosnick og Fabriger, væntanlegt; Spector, 1976). Í slíkum rannsóknum gefa þátttakendur til kynna hvað orðagildi eins og *frekar sammála* og *mjög sammála* merkja með því að velja þeim viðeigandi tölugildi á tilteknu bili. Meðaltalsmat er síðan notað til að ákvarða merkingu eða gildi orðagildis og staðalfrávik hve nákvæm sú merking er. Niðurstöður gefa til kynna hvaða orðagildi er skynsamlegast að nota í viðhorfarannsóknum.

Gildunarannsóknir hafa verið gerðar erlendis síðastliðna áratugi (sjá yfirlit í Krosnick og Fabriger, væntanlegt) og mælt er með því að niðurstöður slíkra rannsókna séu notaðar til að velja orðagildi (Krosnick og Fabriger, 1997; Schaeffer og Presser, 2003). Erlendir rannsakendur hafa því tók á að bæta aðferðafræðileg gæði sinna rannsókna með því að velja orðagildi sem hafa nákvæma merkingu og spanna alla viðhorfavíddina með jöfnu millibili. Hér á landi liggja hins vegar ekki fyrir upplýsingar um gildi orða og orðasambanda sem unnt er að nota á mælistikum. Á þessu er mikilvægt að ráða bót og í þessari rannsókn var upplýsingum um gildi íslenskra orða og orðasambanda sem unnt er að nota á viðhorfamælistikur safnað.

Aðferð

Úrtak

Þátttakendur í þessari rannsókn voru valdir úr 10.000 manna netpanel Capacent og voru þeir alls 598. Tilviljunarúrtak Íslendinga er valið í panelinn og fólki boðið að taka þátt gegn umbun sem var tveir happadrættisvinningar. Annars vegar vinningur sem rann til þátttakenda og hins vegar vinningur sem rann til góðgerðarsamtaka. Alls tóku 394 þátt í rannsókninni og er það 65,9% svarhlutfall. Viðmælendur voru á aldrinum 17 til 73 ára ($M=41$ ár, $sf=13$ ár) og voru 43,7% karlar og 56,3% konur.

Orðagildi

Erlendar rannsóknir á merkingu orðagilda voru hafðar til hliðsjónar þegar orðagildi voru valin (Cliff, 1959; Jones og Thurstone, 1955; Krosnick og Fabrigar, væntanlegt; Spector, 1976). Einnig var leitað fanga í íslenskum orðabókum eins og Íslenskri samheitaorðabók (1985) og Íslenskri orðabók (2007). Í upphafi voru 74 orð eða orðasambönd valin og þau forprófuð á nemum í Háskóla Íslands (Faney Þórsdóttir og Friðrik H. Jónsson 2007). Á grundvelli þeirra niðurstaðna var ákveðið að taka út 27 orðagildi. Tveimur orðagildum sem lýsa hlutlausri afstöðu var bætt við. Alls voru því 49 orð eða orðasambönd notuð í þessari rannsókn. Hvert orð og orðasamband var gildað tvisvar, annars vegar með orðinu sammála og hins vegar með orðinu ósammála. Þetta var gert í ljósi þess að tiltekið orð getur teygst mismikið á lýsingarorði eftir því hver stefna þess lýsingarorðs er. Þannig þarf orðagildið mjög sammála ekki nauðsynlega að vera 4 þó svo að orðagildið mjög ósammála fái gildið -4. Hverju orðagildi fylgdi 11-svarmöguleika mælistika. Hver svarmöguleiki stikunnar var merktur með tölum frá -5 til +5 og endagildi hennar var merkt með orðagildunum *eins ósammála og hægt er að vera* og *eins sammála og hægt er að vera*.

Framkvæmd

Könnunin var lögð fyrir á netinu. Þátttakendur fengu í upphafi dæmi þar sem verkefnið var útskýrt. Í framhaldi af því fengu þeir lista af orðum og orðasamböndum sem lýsa því hve

sammála eða ósammála fólk getur verið og voru beðnir um að gefa til kynna hve sterka afstöðu hvert orðagildi hefur.

Þess var sérstaklega gætt að tilviljun réði því í hvaða röð orðagildi birtust svarendum (*random order*). Það er hugsanlegt að þátttakendur ígrundi betur svör sín í upphafi fyrirlagnar en í lok fyrirlagnar sökum þreytu. Í því tilfalli er hætta á að tilviljun ráði vali á svarmöguleika sem leiðir til vanmats á áreiðanleika þeirra orðagilda sem koma í lok fyrirlagnar. Röð orðagilda getur því haft villandi áhrif á niðurstöður og því nauðsynlegt orðagildi birtist svarendum ekki í kerfisbundinni röð.

Í verkefni sem þessu er hætta á að svarendur misskilji fyrirmæli eða leggi sig ekki fram. Til að útiloka slík svör var stuðst við miðmið sem sett eru fram í Cliff (1959) og Jones og Thurstone (1955). Í fyrsta lagi voru svarendur sem mátu innan við helming atriða ekki teknir með í úrvinnslu. Í öðru lagi var athugað hvort tiltekið mynstur kæmi fram í svörum. Ef þátttakendur völdu til dæmis einungis einn eða tvo svarmöguleika þá voru þeir útilokaðir frá þátttöku. Í þriðja lagi voru óvenjuleg svör skoðuð nánar. Ef til dæmis *lítilllega sammála* var talið lýsa sterkari afstöðu en *fullkomlega sammála* voru svör viðkomandi skoðuð nánar. Ef slíkt ósamræmi kom fram þegar sex eða fleiri önnur orðagildi voru skoðuð var litið svo á að viðkomandi hefði annað hvort misskilið verkefnið eða ekki sinnt því sem skildi og ekki var unnið úr því svári. Að lokum var skoðað hvort orðagildum var gefið andstætt gildi við það sem við stefnu orðagildis. Ef þátttakandi valdi til dæmis svarmöguleikann 5 fyrir orðagildið *mjög ósammála* í stað þess að velja neikvætt tölugildi voru svör hans skoðuð nánar. Ef ósamræmi af þessum toga kom fram oftast en átta sinnum var viðkomandi útilokaður frá þátttöku. Ef slíkt ósamræmi kom hins vegar fram átta sinnum eða sjaldnar voru svör endurkóðuð þannig að orðagildið fékk gildi andstætt stefnu sem fólst í merkingu orðagildis. Alls þurfti að fjarlægja 99 svarendur úr gagnasafninu vegna þess að þeir mátu innan við helming atriða. Fimm völdu einungis einn eða tvo svarmöguleika og voru því útilokaðir frá þátttöku. Í átján tilfellum var ekki unnið úr svörum vegna þess að viðkomandi annað hvort misskildi verkefnið

eða sinnti því ekki sem skildi. Að lokum þurfti að endurkóða 191 svör. Heildarfjöldi svarenda var því 272.

Útreikningar á gildum

Röðunaraðferðin (*method of successive intervals*) (Guildford, 1954) var notuð til að reikna út gildi og staðalfrávik þeirra orða og orðasambanda sem voru metin. Sú aðferð er talin heppilegri en aðferð paraðs samanburðar (*method of paired comparisons*) þegar áreiti eru mörg (Moiser, 1940), líkt og þau voru í þessari rannsókn. Röðunaraðferðin einnig talin heppilegri en jafnbilaaðferðin (*method of equal appearing intervals*) því síðarnefnda aðferðin hvílir á þeirri forsendu að jafnt bil sé á milli svarmöguleika á mælistikunni sem fylgir orðagildum. Röðunaraðferðin gerir hins vegar einungis ráð fyrir að svarmöguleikar séu á raðkvarða Edwards, 1952; Guildford, 1954).

Niðurstöður og umræða

Niðurstöður eru í meginráttum að orð og orðasambönd raðast í nokkuð góðu samræmi við almenna máltilfinningu. Í 1. töflu er að finna gildi og staðalfrávik hvers orðs og orðasambands. Gildi segir til um merkingu orðagilda og staðalfrávik segir til um hve nákvæm sú merking er. Þegar gildi í 1. töflu eru skoðuð sést að styrkur hvers orðagildis er yfirleitt háð stefnu þess og tekur neikvæða stefnan yfirleitt sterkara gildi. Þannig hefur orðagildið algjörlega sammála veikari merkingu (gildi=6) en orðagildið algjörlega ósammála (gildi=-7). Þetta er áhugvert í ljósi hugmynda um að neikvæð viðhorf hafa yfirleitt meiri áhrif á hegðun en jákvæð viðhorf (Cacioppo og Berntson, 1994). Rétt er þó að benda á að munur þarna á milli var yfirleitt ekki mjög mikill. Niðurstöður sýna einnig að almennt séð er merking orðagilda sem lýsa mjög sterkri afstöðu annars vegar og veikri afstöðu hins vegar nákvæmari en merking orðagilda sem lýsa afstöðu sem fellur nær miðju mælistikunnar.

Niðurstöður í 1. töflu geta nýst við gerð matskvarða. Á grundvelli þeirra er hægt að velja orðagildi sem hafa nákvæma merkingu og spanna viðhorfavídd með jöfnu millibili. Ef lítið er fyrst til mælistika með fimm svarmöguleika benda niðurstöður til þess að hin hefðbundnu orðagildi frekar

sammála/ósammála og mjög sammála/ósammála spanni viðhorfavíddina

Tafla 1. Gildi og staðalfrávik orðasambanda

	Sammála		Ósammála	
	Gildi	Sf.	Gildi	Sf.
Að öllu leyti	6	0,4	-7	0,3
Algjörlega	6	0,3	-7	0,2
Skilyrðislaust	6	0,5	-6	0,4
Fullkomlega	6	0,2	-6	0,4
Alveg	6	0,5	-6	0,6
Alfarið	6	0,5	-6	0,4
Tvímælalaust	6	0,7	-6	0,6
Eindregið	6	0,7	-6	0,6
Fyllilega	6	0,8	-6	0,6
Ofur	5	1,0	-6	1,2
Sannarlega	5	1,0	-6	1,0
Virkilega	5	0,9	-6	1,0
Mjög	5	1,0	-6	1,0
Ákaflega	5	0,9	-6	1,0
Óskaplega	5	1,2	-6	1,2
Feikilega	5	1,1	-6	1,0
Afskaplega	5	1,0	-6	1,0
Afar	5	1,0	-6	1,0
Einkar	5	1,1	-5	1,2
Verulega	4	1,0	-5	1,0
Að mestu	4	0,6	-4	0,6
Að mestu leyti	4	0,6	-4	0,5
Ansi	4	0,9	-4	0,9
Í meginatriðum	4	0,8	-4	0,9
Töluvert	4	0,7	-4	0,9
Að miklu leyti	4	0,9	-4	0,6
Býsna	3	1,0	-4	0,9
Að talsverðu leyti	3	0,7	-4	0,7
Talsvert	3	0,9	-4	0,8
Þó nokkuð	3	0,7	-4	0,7
Fremur	3	0,9	-4	0,8
Tiltölulega	3	0,9	-3	1,0
Frekar	3	0,8	-3	0,8
Nokkuð	3	0,9	-3	0,7
Að vissu marki	2	0,8	-3	0,8
Að nokkru leyti	2	0,7	-3	0,7
Sumpart	2	0,7	-3	0,8
Hálfpartinn	2	0,9	-3	0,7
Hallast að því að vera	2	1,1	-2	1,1
Hneigist til að vera	2	0,9	-2	1,0
Að einhverju marki	2	0,7	-2	0,7
Dálítið	2	0,7	-2	0,7
Að dálitlu leyti	2	0,6	-2	0,8
Svolítið	2	0,7	-2	0,7
Smávegis	1	0,5	-2	0,5
Lítilllega	1	0,6	-2	0,5
Ögn	1	0,4	-2	0,5
Hlutlaus	0,0	0,2		
Hvorki né	0,0	0,1		

sæmilega vel. Þær benda einnig til að þessi orðagildi hafi í samanburði við önnur með svipuð gildi tiltölulega nákvæma merkingu. Að vísu fengu orðagildin *mjög sammála* og *mjög ósammála* gildin 5 og -6. Það er því ekki hægt að líta svo á að matskvarði með ofangreind orðagildi sé jafnbilakvarði (Krosnick og Fabrigar, 1997; Schaeffer og Presser, 2003) (2. mynd). Í því sambandi er þó rétt að benda á að niðurstöður rannsókna sýna að jafnlangt bil á milli orðagilda tryggir ekki að mælingar hafi eiginleika jafnbilakvarða (Faney Þórsdóttir, 2005). Þrátt fyrir það er full ástæða til að velja orðagildi sem hafa nokkuð jafnt bil sín á milli um leið og þess er gætt að orðagildin spanni alla viðhorfavíddina. Þegar bil á milli orðagilda er nokkuð jafnt ætti að vera auðveldara fyrir svarendur að greina á milli merkinga ólíkra orðagilda. Þar af leiðandi ætti áreiðanleiki mælinga að aukast þegar orðagildi dreifast tiltölulega jafnt yfir viðhorfavíddina.

Þakkir

Þakkir fá Rannsóknasjóður Rannís, Rannsóknasjóður Háskóla Íslands og Capacent sem styrktu þessa rannsókn.

Heimildir

- Alwin, D. F. og Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondents attributes. *Sociological Methods and Research*, 20(1), 139-181.
- Bending, A. W. (1953). The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *The Journal of Applied Psychology*, 37(1), 38-41.
- Boote, A. S. (1981). Reliability testing of psychographic scales. *Journal of Advertising Research*, 21(5), 53-60.
- Cacioppo, J. T. Og Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115(3), 401-423.
- Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, 66(1), 27-44.
- Edwards, A. L. (1952). Scaling the stimuli by the method of successive intervals. *Journal of Applied Psychology*, 36, 118-122.
- Faney Þórsdóttir. (2005). On the validity of attitude measurements in survey research: A comparison of psychometric models. Óbirt doktorsritgerð: London School of Economics and Political Science.
- Faney Þórsdóttir og Friðrik H. Jónsson. (2007). Gildun á mælistikum. Í Gunnar Þór Jóhannesson (Ritstj.), *Rannsóknir í félagsvísindum VIII* (bls. 527-536). Reykjavík: Félagsvísindastofnun Háskóla Íslands.
- Guildford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Íslensk orðabók. (2007). Mörður Árnason (ritstjóri). Reykjavík: Edda.
- Íslensk samheitaorðabók. Svavar Sigmundsson (ritstjóri). Reykjavík: Háskóli Íslands.
- Jones, L. V. og Thurstone, L. L. (1955). The psychophysics of semantics: An experimental investigation. *The Journal of Applied Psychology*, 39(1), 31-36.
- Krosnick, J. A. og Fabrigar, L. R. (1997). Designing ratings scales for effective measurement in surveys. Í L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz og D. Trewin (ritstj.), *Survey measurement and process quality* (bls. 141-156). New York: Wiley.
- Krosnick, J. A. og Fabrigar, L. R. (væntanlegt). *Designing great questionnaires: Insights from psychology*. New York: Oxford University Press.
- Mosier, C. I. (1940). A modification of the method of successive intervals. *Psychometrika*, 5(2), 101-107.
- Peters, D. L. og McCormick, E. J. (1966). Comparative reliability of numerically anchored versus job-task anchored rating scales. *Journal of Applied Psychology*, 30(1), 92-96.
- Schaeffer, N. C. og Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65-88.
- Spector, P. E. (1976). Choosing response categories for summated rating scales. *Journal of Applied Psychology*, 61, 374-375.