Indian Journal of Science and Technology, *Vol 9(S1), DOI: 10.17485/ijst/2016/v9iS1/104115, December 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Number of Scale Points and Data Characteristics: An Experimental Investigation

Anandakuttan B. Unnithan*

IIM Kozhikode, Kozhikode - 673570, Kerala, India; anandunnithan@iimk.ac.in

Abstract

Objectives: This study compares the scale characteristics when the same instrument is used with five and four point scales to test the assumption whether results vary with number of scale points. **Methods/Statistical Analysis:** An experimental study was designed where the same items were organized into two different forms one with five point response format and the second with four points format. 195 respondents were randomly assigned to either one of the formats. Comparison was made between the groups on the scale characteristics and correlations obtained. **Findings:** When the data is transformed to a common scale, it is seen, that data characteristics like mean, standard deviation, skewness and kurtosis are comparable indicating that there is no change in scale characteristics. However when the correlations are estimated, it is seen that five point scales report a significantly higher correlation. **Application/Improvements:** The major implication is that five point scales tend to inflate the reported correlations with a distinct chance of higher type1 error especially when the variables may not be associated.

Keywords: Correlation, Five Point Scale, Likert Scale, Measurement, Scale Points, Type 1 Error

1. Introduction

Multi-item rating scales are the most popular data collection instruments used in business research. They are used extensively in surveys and experiments to measure individual and organizational characteristics. Constructs like customer Satisfaction and quality of service used in marketing and organizational commitment and job satisfaction used in organizational behaviour/HR are typically measured using multi-item instruments. There are examples of similar instruments finding application in the measurement of many personality and attitudinal variables used in marketing[1], strategy, information systems[2], organizational behaviour, human resource management etc. Many of the articles in leading academic journals feature data collected using multi-item instruments[3,4].

Multi item instruments typically contain a battery of statements or questions and the respondents are asked to indicate their agreement or disagreement to these statements or questions by choosing a labelled number. The number chosen by the respondent is taken as a measure of the strength of agreement to the statement. Likert scales are the popular choice of formatting the response options and it is common to see five point and four point response formats in published research. There are examples where the original instrument is adapted to a different scale format without any specific reasoning. It is assumed that the measurement characteristics remain unchanged with change in response format. How tenable is this assumption? Can we assume measurement invariance across scale formats or is it more reasonable to assume that data characteristics also may change with changes in response

*Author for correspondence

formats[5]. This study was designed to answer these questions as there is an ambiguity in existing literature on this issue. Specifically, the research problem addressed in this work is whether there will be changes in data characteristics when four point and five point response formats are used with the same items.

Do the data characteristics vary depending on the scale format used? Several researchers have explored this issue to find a conclusive answer[6]. Could there be an optimal number of scale points? Will it make a difference in data if the number of scale points used is odd or even? Are there any criteria based on which a decision can be taken on the best scale format? There is a stream of research going back by many decades on these topics.

A review of literature reveal that previous studies often sought to find an optimal number of scale points and in this quest explored variation in reliability[7], sensitivity[8], user reported ease or convenience[9] and validity[10,11] as criteria for the best scale format. Studies in[12] based on exhaustive meta-analysis concluded that there exist a positive relationship between the number of scale points and the reliability of the construct measured. Another study[10], based on simulation concluded that the correlation between variables tend to decrease as the number of scale categories decrease. This implies that it is better to use larger number of scale points to obtain a higher correlation between variables. Another study[13] investigated the effect of number of scale points on validity and reported that there is a variation in validity coefficients with the increase of scale points. Studies[14] pointed out the prevalence of five to seven point scales in reported academic studies presumably due to a belief that this range is better. However use of more number of scale points need not be a better practice and some studies suggest there are no difference with use of higher number of scale categories[15].

In general, previous research on the subject does not agree on whether number of scale points has an influence on reliability or validity. There are suggestions that three points are enough[16], while others argue for more categories to increase the sensitivity and discrimination. There are studies that advise five point scales[17] above seven and nine[18]. In practice however, the number of scale points is chosen arbitrarily and the reason could be to make it easier for the respondents.

There is an implicit assumption that the change in number of scale points does not have any influence on the data characteristics. This assumption has found some empirical support[19]. This means that parameter estimates and the relationships estimated remain invariant even when a different scale configuration is used. Meta analysis reported in leading journals report and compare effect size estimates without taking into consideration changes in number of scale points in the instruments that produced such effect sizes. Is this assumption valid or is it possible that values of estimated parameters may depend on the number of scale points[20]? It is this doubt that motivated the researcher to undertake this study.

Research Question 1: Whether the choice of number of scale points (4 against 5) determine the data characteristics?

Research Question 2: Whether the number scale points any way determine the estimate of relationships between constructs?

2. Methods

An experimental study was designed to compare five point scales against four point scales whereby two different forms with only difference in response formats (five point and four points) were created using the same items and respondents were randomly assigned to the two forms. The resultant data sets can be compared to see if there are differences and whether these differences can be explained by the difference in scale formats. Twenty two items from the revised self consciousness scale for use with general population[21] were formatted into two different forms. The first one was designed with four point agree-disagree scale and the second with a five point one with 3 as neutral point.

The revised Self-consciousness scale is an adaptation of the self consciousness scale[22] and had three sub dimensions namely Private Self Consciousness (hereafter public self), Public Self Consciousness (hereafter public self) and Social Anxiety (hereafter social anxiety). This scale was subsequently revised for use among the general population[21] with the same three dimensions, but with some modifications in items. The revised self consciousness scale[21] had 9 items measuring private self, 7 items measuring public self and 6 items for social anxiety. This instrument had used a four point scale from 0 to 4 and reported correlations of 0.38 (between private self and public self), 0.03 (between private self and social anxiety) and 0.35 (public self and social anxiety)

Using the twenty two items from the revised scale, two different forms were created one with a four point scale response format and the second with a five point scale format. In both forms, the terminal items were

anchored to the verbal labels 'Agree' and 'Disagree' and the in between points were given as numbers. The forms were hosted online and respondents were randomly assigned to one of the two forms with the randomisation algorithm running in the server. The forms also contained other items which are not used in the present study. The items and variables used in the study are listed in Table 1.

Table 1. Variables and items used in the study

Construct	Item
Private Self Consciousness	I am always trying to figure myself out.
	I think about myself a lot.
	I often daydream about myself.
	I never take a hard look at myself.
	I generally pay attention to my inner feelings.
	I am constantly thinking about my reasons for doing things.
	I sometimes step back (in my mind) in order to examine myself from a distance.
	I am quick to notice changes in my mood.
	I know the way my mind works when I work through a problem.
Public Self Consciousness	I am concerned about my style of doing things.
	I care a lot about how I present myself to others.
	I am self-conscious about the way I look.
	I usually worry about making a good impression.
	Before I leave my house, I check how I look.
	I am concerned about what other people think of me.
	I am usually aware of my appearance.
Social Anxiety	It takes me time to get over my shyness in new situations
	It is hard for me to work when someone is watching me.
	I get embarrassed very easily.
	Lt is easy for me to talk to strangers.
	I feel nervous when I speak in front of a group.
	Large groups make me nervous.

An apriori power analysis was performed to determine the samples size for an effect size of 0.35 (correlation between private self and social anxiety reported in the study from which items were drawn). The sample size reported by the software G*power[23] was n = 83 for statistical power $(1-\beta)$ of 0.95. Accordingly a sample size of 85 was targeted for both conditions (five point and four point formats).

Incomplete responses were deleted, and finally 195 responses were available for analysis of which 108 were from the four point condition and 87 from the five point condition. The data was analysed using R –the statistical computing language and environment.

3. Data Analysis

The data represents ratings provided by respondents assigned to two different scale formats on the same set of items. It is important to convert the two set of ratings to a common scale to make meaningful comparison of scale characteristics. There are a number of methods available for this conversion to a common scale of which the method suggested by[5] and reported by[15] was used in this study. They used a formula as given

$$TransformedScaleValue = \frac{Rating - 1}{(Numberof Responsecategories - 1)} \times 100$$

This formula re-scales the original values to a common score out of 100. In this study, we adapted the formula to get scores out of 10 rather than 100 by multiplying with 10. This application of the formula transforms the raw scores to a new scale where the scores will be distributed from 0 to 10.

After the conversion, the total scores of the three constructs 'Private Self', 'Public Self' and 'Social Anxiety' were computed.

In order to check whether there are any differences in data characteristics between the scale formats, means and standard deviations were computed and are given in Table 2.

Table 2. Means and standard deviations of constructs after re-scaling

Construct	Mean-4 Point	Mean-5 point	SD-4 Point	SD-5 point
Private Self	55.72727	55.47753	12.61143	11.77365
Public Self	45.42424	43.23034	12.31476	14.44623
Social Anxiety	30.12121	31.54494	12.67429	12.23878

It can be seen from Table 2 that, the mean values of the constructs as well as the standard deviations are comparable for all the three variables. Eyeballing does not reveal any large differences. It seems that irrespective of the scale format used, the means and standard deviations obtained are close are comparable.

The skewness and kurtosis was also computed from the data for the two groups which is provided in Table 3.

Table 3. Skewness and Kurtosis of variables

Construct	Skewness-4 Point	Skewness-5 point	Kurtosis-4 Point	Kurtosis-5 point
Private Self	-0.10506792	-0.006350211	2.637771	2.651398
Public Self	0.1857418	0.4894282	2.404652	3.066709
Social Anxiety	-0.03720593	-0.03564683	2.536532	2.536532

In the case of skewness and kurtosis also, data from both forms show comparable characteristics. However it is important that the data must be subjected to a more rigorous comparison to see whether the distribution can be considered to be similar across the scale formats. This was accomplished with the help of Levene's test for equality of variances between the groups, and subsequent t test for equality of means.

Levene's test of equality of variances was performed on the data to see of the two sets of data show significant differences in variance. Levene's test is performed on a null hypothesis that the two groups have the same variance against the alternate hypothesis that the groups have unequal variances. The results of the test are provided in Table 4.

Table 4. Levene's test for equality of variances

Construct	DF	F Ratio	P value
Private Self	1	0.3482	0.5558
Public Self	1	2.1254	0.1465
Social Anxiety	1	0.3209	0.5717

The p values obtained indicate that the assumption of equal variances cannot be rejected at alpha = 0.05. Considering that the variances are equal between the groups, three separate t tests were conducted to test the hypothesis whether the group means are significantly different. The results of the t tests are reported in Table 5, 6 and 7.

Table 5. t test for equality of means private self

Private Self	Mean	t-value	df	P value
4 point	55.72727	0.14306	197	0.8864
5 point	55.47753			

Table 6. t test for equality of means public self

Public Self	Mean	t-value	df	P value
4 point	45.42424	1.1562	197	0.249
5 point	43.23034			

Table 7. t test for equality of means social anxiety

Social Anxiety	Mean	t-value	df	P value
4 point	30.12121	-0.80006	197	0.4246
5 point	31.54494			

Here, the obtained p values indicate that the assumption of equal means across the groups cannot be rejected at a significance level of 0.05. Therefore the null hypothesis finds acceptance and the group means can be considered to be equal for all the three variables between four point and five point data sets. Both the Levene's test of equality of variances and t test for equality of means provide support to the argument that the data distribution is same in both the groups. Irrespective of the use of four-point or five-point scale, the means and the variances seems to be similar. Thus with regard to the research objective 1, we may conclude that the distribution of data is not dependent on the scale format used. Both five point and four point scales yield comparable distributions and the means, standard deviations, skewness and kurtosis are similar.

The second objective of research was to see whether the differences in scale formats yield differences in relationships between the constructs. Accordingly, correlations were estimated between each pair of variables for both scale formats. Karl Pearson coefficient of correlation r between the variables pairs are provided in Table 8, 9 and 10.

Table 8. Correlation between private self and public self

Scale Format	Correlation	Sample Size	P Value
4-point	0.3265***	108	0.0005
5-point	0.5366***	87	0.0000

Table 8 reports the correlation between the variables private self and public self for four point and five point scenarios. It is seen that both correlations are positive and significant. Further the correlation for the five point scale is higher than that obtained for the four point scale. The size of correlations indicates a large sizeable difference. A similar pattern is seen in the case of correlations between private self and social anxiety in Table 9.

Table 9. Correlation between private self and social anxiety

Scale Format	Correlation	Sample Size	P Value
4-point	0.0301	108	0.755
5-point	0.2837**	87	0.007

From Table 9, it can be seen that the correlations are higher for the five point scale version compared to the 4 point format. The same pattern is seen in the case of correlation between public self and social anxiety also.

Table 10. Correlation between public self and social anxiety

Scale Format	Correlation	Sample Size	P Value
4-point	0.1440	108	0.1333
5-point	0.3627***	87	0.0005

Table 10 shows that the correlation between public self and social anxiety is 0.3627 for five point scales which is significant at 0.01 while the corresponding correlation is only 0.1440 for four point scales.

It can be seen from Tables 8, 9 and 10 that the estimated correlations are always higher for the five point scale version compared to the four point one. To test whether the sample correlations indicate significant differences in the population parameter estimates, a test of equality of correlations was conducted between each pair of correlation coefficients the results of which are reported in Table 11.

Table 11. Test of equality of correlations

		Correlations		
	Sample Size	Private Self and Public Self	Private Self and Social Anxiety	Public Self and Social Anxiety
4-point	108	0.3265***	0.0301	0.1440
5-point	87	0.5366***	0.2837**	0.3627***
Test of equality of correlations		Z=1.78 p value=0.04	Z=1.79 p value=0.04	Z=1.61 p value=0.05

It is seen that the null hypothesis of equality of correlations cannot be accepted at 5% significance level (one tail test) for at least two pairs of variables. There is an indication that the correlation between variables for five point scales is higher than those for four point scales. This is quite a counter intuitive result as the correlation between constructs is not expected to vary with the scale points. The randomised assignments of respondents rule out other causes behind the difference in correlations. The pattern of higher correlations in five point scales can be only attributed to the scale format.

4. Discussion

The results clearly indicate the following. 1. The univariate data characteristics do not show any differences when both data sets were rescaled to a common scale. 2. However the estimated correlations are significantly higher in the case of five point scales compared to four point options. Five point scales tends to obtain higher correlations and the differences are significant for all the three pairs of variables. The implications of this result are very important when the constructs are unrelated. The sub-dimensions of self consciousness private self and social anxiety are expected to be uncorrelated as reported in prior studies[21,22]. In this study while correlations are very low (0.1440) and not significant (p value = 0.133) between public self and social anxiety in line with the theoretical expectations, the correlations obtained are much higher (0.3627) for the five pint scale. Further the correlation is significant also. The difference between correlations was found to be significant at 0.05. This implies that the use of five point scales for both variables may result in a high and significant correlation even when the variables may not be related. It is actually type 1 error when we wrongly reject the null hypothesis and conclude the presence of a relationship when there exist none. In most studies, the endeavour of the researcher is to reject the null hypothesis and report a significant relationship. In this quest, it is possible that false positives may result from the use of certain scale point configuration (in this case five point)

5. Conclusions

The major implication from the study is that probability of type one error (rejecting a true null hypothesis) could be inflated in the case of five point scales. Though there

are no difference in data characteristics (mean, standard deviation, skewness and kurtosis) between data collected using five point and four point scales, the correlations between constructs show significant differences in one direction.

It needs more research to replicate these findings as well as understand the process by which such a phenomena is manifested. The current research also has limitations of a one off study, using online data collection methods and the constructs drawn from personality variables. More replications with different respondent profiles, and other constructs are required to validate these findings. At the same time, it is important to remember that perhaps five point scales may result in inflated estimates of relationships between constructs.

6. References

1. Ko SH. Effect of consumer value on the brand identification and loyalty in airline service. Indian Journal of Science and Technology. 2016 Jul 21; 9(26):1–6.
2. Yeo HJ. Information system success disparity between developer and users. Indian Journal of Science and Technology. 2016 May 30; 9(20):1–6.
3. Deepa S. Computer-related health problems among white-collar employees: Communicating a blueprint. Indian Journal of Science and Technology. 2016 Aug 30; 9(32):1–10.
4. Lee CS, Park SU, Hwang YK. The structural relationship between mother's parenting stress and child's well-being: the mediating effects of mother's growth mindset and hope. Indian Journal of Science and Technology. 2016 Sep 29; 9(36):1–6.
5. Cabooter E, Weijters B, Geuens M, Vermeir I. Scale format effects on response option interpretation and use. Journal of Business Research. 2016 Jul 31; 69(7):2574–84.
6. Jones WP, Loe SA. Optimal number of questionnaire response categories. Sage Open. 2013 Apr 1; 3(2).
7. Weng LJ. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. Educational and Psychological Measurement. 2004 Dec 1; 64(6):956–72.
8. Contractor SH, Fox RJ. An investigation of the relationship between the number of response categories and scale sensitivity. Journal of Targeting, Measurement and Analysis for Marketing. 2011 Mar 1; 19(1):23–35.
9. Preston CC, Colman AM. Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. Acta Psychologica. 2000 Mar 31; 104(1):1–15.
10. Martin WS. The effects of scaling on the correlation coefficient: A test of validity. Journal of Marketing Research. 1973 Aug 1; 10(3):316–8.
11. Maydeu-Olivares A, Kramp U, García-Forero C, Gallardo-Pujol D, Coffman D. The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. Behavior Research Methods. 2009 May 1; 41(2):295–308.
12. Churchill Jr GA, Peter JP. Research design effects on the reliability of rating scales: A meta-analysis. Journal of Marketing Research. 1984 Nov 1; 360–75.
13. Rodgers WL, Andrews FM, Regula Herzog A. Quality of survey measures: A structural modeling approach. Journal of Official Statistics. 1992 Sep 1; 8:251–75.
14. Leitz P. Research into questionnaire design: A summary of the literature. International Journal of Market Research. 2010; 52(2):249–72.
15. Lee J, Paek I. In search of the optimal number of response categories in a rating scale. Journal of Psychoeducational Assessment. 2014 Oct 1; 32(7):663–73.
16. Jacoby J, Matell MS. Three-point Likert scales are good enough. Journal of Marketing Research. 1971 Nov 1; 8(4):495–500.
17. McKelvie SJ. Graphic rating scales-How many categories? British Journal of Psychology. 1978 May 1; 69(2):185–202.
18. Revilla MA, Saris WE, Krosnick JA. Choosing the number of categories in agree-disagree scales. Sociological Methods and Research. 2013 Dec 10; 43(1):73–97.
19. Dawes JG. Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. International Journal of Market Research. 2008 Feb 29; 51(1).
20. Eutsler J, Lang B. Rating scales in accounting research: The impact of scale points and labels. Behavioral Research in Accounting. 2015 Jul; 27(2):35–51.
21. Scheier MF, Carver CS. The self-consciousness scale: A revised version for use with general populations. Journal of Applied Social Psychology. 1985 Dec 1; 15(8):687–99.
22. Fenigstein A, Scheier MF, Buss AH. Public and private self-consciousness: Assessment and theory. Journal of Consulting and Clinical Psychology. 1975 Aug; 43(4):522–37.
23. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods. 2009 Nov 1; 41(4):1149–60.