Likert Scale: Explored and Explained

Ankur Joshi[1*], Saket Kale[2], Satish Chandel[3], D. K. Pal[4]

1- MD, Assistant Professor, Department of Community Medicine, Gandhi Medical College, Bhopal, 462001

2- MD (Community Medicine), Program Officer-Clinical, Technical Support Unit, Madhya Pradesh State AIDS Control Society, India.

3- MD (Pharmacology), Tutor, Department of Pharmacology, All India Institute of Medical Sciences, Bhopal.

4-MD, Professor & Head, Department of Community Medicine, Gandhi Medical College, Bhopal, 462001.

*- All correspondence to be addressed to Dr. Ankur Joshi, MD, Assistant Professor, Department of Community Medicine, Gandhi Medical College, Bhopal, 462001.

Email: drankurjoshi7@gmail.com.

ABSTRACT:

Likert scale is applied as one of the most fundamental and frequent psychometric tools in educational and social sciences research. Simultaneously it is also subjected to a lot of debates and controversies in regards with the analysis and inclusion of points on the scale. With this context; through reviewing the available literature and then clubbing the received information with coherent scientific thinking; this paper attempts to gradually build a construct around Likert scale. This analytical review begins with the necessity of psychometric tools like Likert scale and its' variants and the focuses on some convoluted issues like validity, reliability and analysis of the scale.

KEY-WORDS: psychometrics, Likert scale, points on scale, analysis, education.

INTRODUCTION:

Nothing is more than a fear you cannot name.

― Cornelia Funke, Inkheart

Since the inception of human race; there is an inclination to capture the ethereal attributes of human behaviour and performance. Simultaneously it has been a challenge from the same time to quantify the thing which cannot be measured through conventional measurement techniques. The perceived need of this quantification lies in the necessity to transform an individual's' subjectivity into an objective reality. Attitude, perceptions and opinions are such qualitative attributes amenable for quantitative transformation due to above mention reason . Qualitative

research techniques do try to compensate by depicting the complexity of human thoughts, feelings and outlooks through several social science techniques; still the quantification of these traits remains a requirement...that's how psychometric techniques come into picture.

PSYCHOMETRICS AND LIKERT SCALE

Psychometrics techniques are being developed, instituted and refined in order to meet the quantification of traits like ability, perceptions, qualities and outlooks- the requirement of social sciences and educational researches [1, 2]. Psychometrics operates through two ways- the first is to formulate of approaches (theoretical construct) for measurements followed by development of measuring instruments and their validation. Stanford Binet test (measures human intelligence) and Minnesota Multiphasic Personality Inventory (measures human personality) are the example for the same. The content in such instruments are rather 'pre-fixed.' [3, 4, 5].The another path is same up to formulation of theoretical construct for the measurement. This conceptualization is followed by operational assembly of abstract ideas /experiences/issues under investigation into some statements (items) largely guided by aim of the study .This permits the contents(items) in such scales/models to be rather flexible and need based. Rasch measurement model (use for estimation of ability), Likert scale (measures human attitude) are the examples of such scales in Psychometrics used widely in the social science & educational research [3, 4, 5].

Likert scale was devised in order to measure 'attitude' in a scientifically accepted and validated manner in 1932 [6, 7]. An attitude can be defined as preferential ways of behaving/reacting in a specific circumstance rooted in relatively enduring organization of belief and ideas(around an

object, a subject or a concept) acquired through social interactions [8]. This is clear from this discourse mentioned above that thinking (cognition), feeling (affective) and action (psychomotor) all together in various combination/permutation constitute delivery of attitude in a specified condition. The issue is how to quantify these subjective preferential thinking, feeling and action in a validated and reliable manner .....a help is offered by Likert scale [9, 10].

The original Likert scale is a set of statements (items) offered for a real or hypothetical situation under study. Participants are asked to show their level of agreement (from strongly disagree to strongly agree) with the given statement (items) on a metric scale. Here all the statements in combination reveal the specific dimension of the attitude towards the issue ;hence necessarily inter-linked with each other. [11].

With this context, this exploratory article attempts to describe two confusing issues related with Likert scale- (would be) preferable numbers of points on a scale and analysis of the scale. During one of the contributing author's participation in a web based conversational learning forum on medical education, these two issues emerged as thrust area amenable for further exploration and lucid explanation for the educational researchers . An initial literature search by authors led to aggregation of mutual conflicting evidences which compelled us to re-explore and further construction of arguments based upon accumulated knowledge.

LIKERT SCALE AND ITS VARIATION

Before proceeding further; let's have a brief look on several constructional diversities of a Likert scale as the analytical treatment and interpretation with Likert scale largely depends upon these diversities.-

Symmetric versus asymmetric Likert scale- If the position of neutrality (neutral/don't know) lies exactly in between two extremes of strongly disagree (SD) to strongly agree (SA) ; it provides an independence to a participant to choose any response in a balanced and symmetric way in either directions .This construction is known as symmetric scale . On the other hand , asymmetric Likert scale offers less choices on one side of neutrality (average) as compared to other side. Asymmetric scale in some cases also indicate ipsative (forced) choices where there is no perceived value of indifference/neutrality of researcher [12, 13,14].

Seven /ten point scale - They are the variation of 5 point scale in which adjacent options are less radically different(or more gradually different) from each other as compare to a 5 point scale. This larger (step by step) spectrum of choices offers more independence to a participant to pick the 'exact' one (which he prefers most) rather than to pick some 'nearby' or 'close' option. [15]. These variations are discussed in more details (in reference with validity and reliability) further in this paper.

Likert and Likert type scale- The construction of Likert (or Likert type) scale is rooted into the aim of the research. When the purpose of inquiry is to understand the opinion/perception around a single phenomenon or concept ; all the statements in the scale in a complementary manner address the specific portion of the phenomenon /concept. Here during analysis, the scores of the all items of the questionnaire are combined (sum) to generate a composite score, which logically in totality measures a uni-dimensional trait . this instrument is known as Likert scale.

Sometimes the primary interest of the researcher is not to synthesize the stance of the participants per se but to capture feelings, actions and pragmatic opinion of the participants about mutually exclusive issues around phenomenon/s under study. This fact demands the

individual analysis of item to ascertain the participants' collective degree of agreement around that issue. The scale used so can be labeled as Likert type and not Likert scale [16]. A word of caution ; this 'direction of enquiry' must be decided during the planning phase and at least during the designing of questionnaire and not at the time of analysis.

IS 7 POINT LIKERT SCALE BETTER THAN 5 POINT LIKERT SCALE? - A PERSPECTIVE CONTROVERSY OR ESTABLISHED WITH A CONSENSUS?

Since the advent of Likert scale in 1932, there have been debates among the users about its best possible usability in term of reliability and validity of number of points on the scale [17-20].

Likert (1932,7) in his original paper discussed about the infinite number of definable attitudes existing in a given person with possibility of grouping them into "clusters" of responses. He further conversed about the assumption of his "survey of opinions" on which he provided his results and psychological interpretations[21] The key assumptions of his survey being firstly, the presentation of item on scale are such that, so as to allow the participants to choose clearly opposed alternatives. Secondly, the conflicting issues chosen were empirically important issue thus result themselves constituting an empirical check on the degree of success.

Thus, it is argued in particular context of clustering of attitudes Considering reliability of the responses from participants in a survey, chances are that the 7 point scale may perform better compared to 5 point scale owing to the choice of items on scale defined by the construct of survey . The 7 point scale provides more varieties of options which in turn increases the probability of meeting the objective reality of people. As a 7-point scale reveals more description about the motif and thus appeals practically to the "faculty of reason" of the participants [19, 20]. A respondents' absolute agreement with the motif of topic may lie between the two descriptive options provided on a 5 point scale. On repeated administration he/she may differ in choosing

one of the options e.g. 3 instead of 4 when the person thinks in between the two of the response option on 5 point scale. A 7 point scale may eliminate this problem up to an extent, by eliciting retrieval beyond the utmost level of agreement provided by a 5 point scale. Hence, the dilemma of choosing between the two undesirable points on 5 point Likert scale may convert into the slight subjective predicament towards a response option or even an absolute agreement as a result of provision of degree of measurement beyond the extent of 5 point scale [22-24]. The provision of number of scale points, 5 point or 7 point, would be more engaging to the minds of respondents when the items on the scale carry the statement of ideas near the truth of the universe for both the participants and the surveyor. It may create the 'curves of reliability' around the 'zenith of validity'. The dilemma of choice and explicit greater extent of measurement by 7 point scale is very much in the territory of the reason of response without which consideration of reliability is of no weight [19].

Validity of Likert scale is driven by the applicability of the topic concerned, in context of respondents' understanding and judged by creator of the response item. We can appreciate it by an example: "How efficacious is a therapeutic modality in treating a particular disease?" This question when asked to a group of individuals, indifferent with the disease or the modality, the response pattern may remain similar, independent of the number of point on the scale. The responses may cluster around center or to the extreme ends. On the contrary, when the topic concerned is relevant to the respondents' context provision of more option may add to the content & construct validity of the scale. Providing options more close to the original view of the respondent reduce the role of ambiguity in the responses. [23, 12] Furthermore, comprehension of all items and points on a scale needs a judgment time and a memory span different for different means and also depends on communication mode. While listening to the responses of a

long scale may discern the various options on the scale with lesser time to judge compared to a written scale. Written scale thus will add to validity even with more points on the Likert scale. Also research concerning span of immediate memory support this notion of accuracy of response categories around seven as human mind has span of absolute judgment that can distinguish 7 categories at a time [25].

ANALYSIS OF THE RESPONSE ITEM

Before we proceed to the method of analysis available to Likert scale, a very fundamental but equally controversial question should be addressed- which type of scale Likert is?

There are two schools of thoughts - One school considers Likert scale as ordinal and other treats it as Interval scale. This conflict is primarily rooted into the question- whether points on a items are equivalent and equidistant? Points on scale are not close enough to consider them equal (in other words strongly agree is definitely away from agree and agree is away from neutral); they should be considered as non-equivalent entity. There is an agreement in both school for the above fact. The conflict arises on asking another question- if the points on scale are non - equivalent; are they equi-distant (in other words is 'neutral' of same distance from 'agree' as 'agree' from 'strongly 'agree')? This question is important as by answering of this question only, one can decide whether Likert scale can be treated as Interval scale?

The first school of researchers and statisticians consider Likert scale as ordinal scale. They argue that choices or responses are arranged in some ranking order,however as this scale doesn't show the relative magnitude and distance between two responses quantitatively ;it can not be treated as interval scale. The other school interprets this dilemma from a different perspective

stating that when the aim of the researcher is to ‘combine’ all the items in order to generate a ‘composite’ score for an individual rather than separate analysis of single item responded by all individuals; then this individualistic summative score (for all the items) of a participant shows a sensible realistic distance from the individual summative score of another individual ; hence can be labeled as ‘interval estimates’ [26, 16].

To understand this concept; let’s assume a scenario in which the aim of the researcher is to measure the attitude towards classroom lectures and to make out relative preferences (library reading and small group teaching) compared with lecture. He designs the following survey instrument on a 5 point Likert scale for the stated aim-

Fig.1. Survey Instrument for Measuring Attitude towards Classroom Lectures

Items	SD	D	N	A	SA
1. Lectures are usually dominated with monologue without revealing the application of what is being taught.	1	2	3	4	5
2. The information provided in lecture is very much textual and can be retrieve easily though text book.	1	2	3	4	5
3. Lectures usually fail to generate curiosity and further motivation to know more about the topic	1	2	3	4	5
4. I would like to spend time in self directed reading owing to the opportunity of control on pace and choices of learning	1	2	3	4	5
5. I would like to attend Small Group Teaching (SGT) session owing to the opportunity of more interaction and liveliness of the environment.	1	2	3	4	5

The first question of importance is-'Can these items be clubbed (see together) in order to generate a composite index for measuring the attitude?' In order to evaluate their appropriateness for transformation into a single composite index; following points can be considered-

a) Whether the items are arranged in logical sequence,

b) Whether the items are closely interrelated but provide some independent information as well

c) Whether there is some element of 'coherence /expectedness' between responses (whether next response can be predicted up to some extent based upon previous one)

d) Whether each item measures a distinct element of the issue

If answer to all the above questions are affirmative for all the items of a set ,they may be combined to construct a composite index which measures the collective stance of the participant

towards phenomenon under study .In the above example as item-1,2 and 3 fulfill all four criteria for each other ,they may be combined and can be treated further in unison.

On the other hand ; Item-4 and item-5 offer separate and sovereign (mutually exclusive) preferences regarding two different teaching-learning methods- self directed reading and small group teaching. Hence they cannot be combined and further they should be analyzed independently from item 1,2 and 3 and even from each other.

After this assertion of eligibility for combination, the next question arises- On what scale can item1, 2 and 3 be treated and what is the appropriate measurement scale for item4 and item-5?

The answer of the above question lies in another question asked by Stevens in his famous paper- 'what are the rules (if any) under which numerals are assigned?' Here we see (a) the minimum score one can secure for first three item is 3 (and not a absolute zero). The reason for this apparently dislike for zero lies in the fact that in psychometrics, attitude is preferably measured in positive degree and being the 'strongly disagree 'cannot be equated with 'absolute disagreement'....there is always something below than strongly disagree. Zero also gives the notion of neutrality rather disagreement (the attitude is zero...means one is apathetic to issue) (b) each numeral conveys the same meaning in all three items (i.e. 3 denotes the neutral in all three items) (c) as mentioned above, all three items can be clubbed while satisfying the content and criterion validity. This sentence needs a little more explanation. The idea or concept behind framing item1,2 and 3 is to capture the opinion of participants about the lecture. This theoretical construct how well can be transformed into operating reality, can be ascertained by looking at relevant content domains (content validity/reflection of construct),ability to distinguish opinion on lecture from other teaching modality (concurrent validity) .and similarities among items 1 to 3 and dissimilarities from item 4 and 5 (convergent and discriminant validity).Concurrent

,convergent and discriminant validities are the domains of criterion validity. Before deciding any statistical treatment to items, all the items must be scrutinized for validity issues.

If we look into point (a),(b) and (c) in cohesion for the set of item 1,2 and 3 ; that composite score for the item-1,2 and 3 can be compared with another composite score for another individual on an interval scale . A 'rank-order' among the composite scores can be presumed as well as equality of interval among related composite scores can also be postulated . The specific point on a particular item is conveying the same meaning for all individuals (for item -2 point 3 on Likert scale denotes 'neutral' among all individuals.) Moreover a specific point (say 2 for disagree) is conveying the same meaning (same extent of disagreement) in all the items and there is no absolute zero in scale (minimum achievable score is 3). From the discourse, this can be safely assumed (after going through all these mathematical characteristics with due consideration of validity related issues) that the obtained composite data for item1,2 and 3 for all the participants can be treated on an interval scale.

The truth has different dimension in case of item 4 and item 5. Item -4 and 5 being an mutually exclusive observation from each other (opinion on self directed reading/ small group teaching) and from item1, 2 and 3; should be treated differently. They may not be combined (validity restriction) for an individual as they are nowhere providing complementary observation. Still as a specific point on a particular item is conveying the same meaning for all individuals same as of previous data set. Moreover response variables obtained for a single item from all the individuals can be arranged in any order preserving transformation (like square, multiplication, square root etc.) to the response variable(the rank order remains unaffected) .... so an ordinal scale's assumptions and treatment is applicable on this subset of items (4 and 5).

Once it is clear that under which rules the items are categorized and what the direction of inquiry is; it becomes obvious the further statistical treatment as per their assignment into ordinal or interval scale.

AUTHORS'NOTE: CONCUSION-

The crux that can be extracted from the above inductive arguments and logical interpretation that the methods adopted for Likert scale analysis largely depends on the item response variable assignment into ordinal or interval scale which in turn depends on the construct of the research instrument. This construct of research instrument can be derived from objectives of study and objectives are the operational form of theoretical construct of phenomenon under inquiry . In other words design of instruments based upon objectives and frameworks of study decides further statistical treatment. .

Hence ,If one wishes to combine the items in order to generate a composite score (Likert scale) of a set of items for different participants then the assigned scale will be an interval scale (fig.2). The measures for central tendency and dispersion for an interval scale are mean and standard deviation. Further this data set can be statistically treated with Pearsons' correlation coefficient (r), Analysis of Variance (ANOVA)- and regression analysis.

As opposed to, if researcher wishes to analyze separate item (no composite score; Likert type scale) the assigned scale for such data set will be ordinal (fig.2). Needless to say; the recommended measure of central tendency and dispersion for the ordinal data are the median (or the mode) & frequency (or range) . An ordinal data set can further be statistically tested by non-parametric techniques such as Chi-square test, Kendall Tau B or C test.

Before wrapping up ; it is imperative to transform an abstract issue into figurative shape in order to measure it up to best possible extent . Simultaneously this is an integrate process reason

being influenced by perspective and subjectivity of researcher. Still all attempts should be directed for quantification of such qualitative attributes as -'what get measured ,get managed.' (Peter Druker).

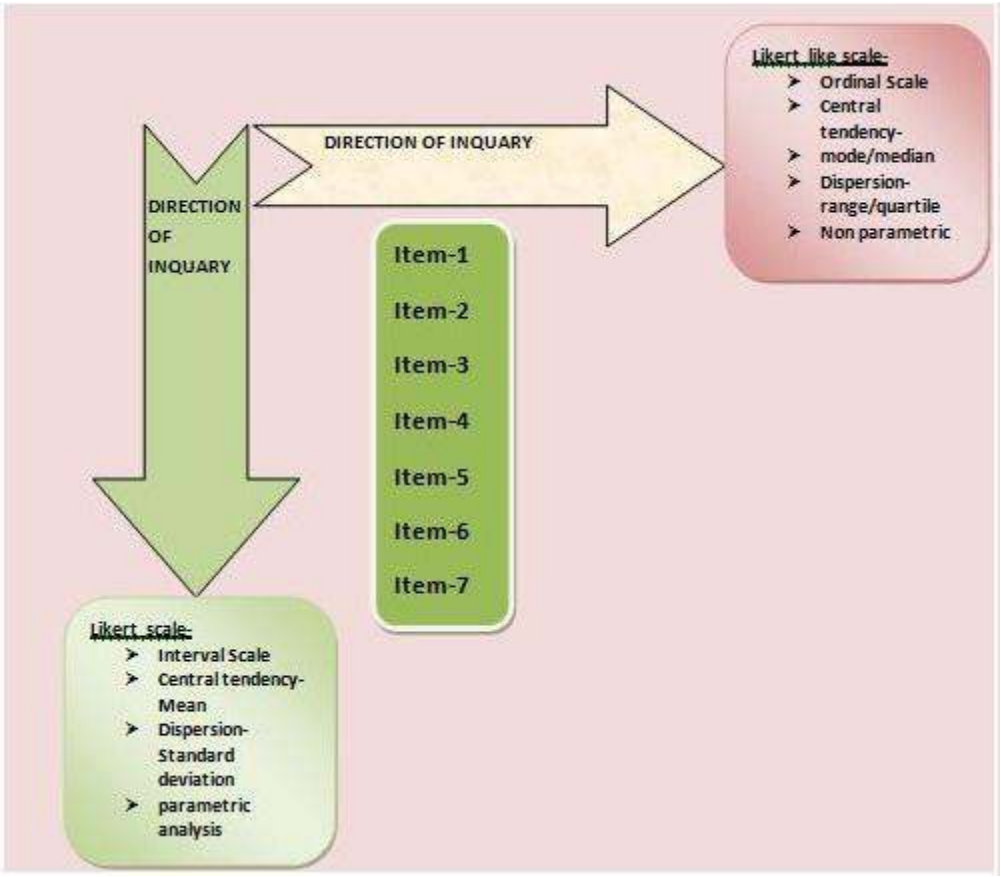

Fig.2. Choice of Analysis of Likert Items: Aim and Construct of Research

COMPETING INTERESTS

Authors have declared that no competing interests exist.

AUTHORS' CONTRIBUTION

Aj- Initiated the idea, wrote the first draft and contributed in further refinement with critical inputs to literature review. Sk-Conceptualized the variation of Likert scale and provided critical input to the several drafts of manuscript and literature review. Sc- Contributed for common understanding of psychometrics with critical inputs. Dp- reviewed and facilitated the final shape of paper.

All authors read and approved the final manuscript

REFERENCES:

1. Cook DA, Beckman TJ. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *The American Journal of Medicine.* 2006; 119(2): doi:10.1016/j.amjmed.2005.10.036
2. Shea JA, Fortna GS. Psychometric Methods. In: Norman GR, Van der Vleuten CPM, Newble DI, editors. *International Handbook of Research in Medical Education.* Dordrecht: Kluwer Academic Publisher; 2002. p. 97-126.

3. Youngstrom EA, Glutiing JJ, Watkins MW. Stanford-Binet Intelligence Scale: Fourth Edition (SB4): Evaluating the Empirical Bases for Interpretations. In: Reynolds CR, Kamphaus R W, editors. *Handbook of psychological and educational assessment: Intelligence, aptitude, and achievement.* New York: Guilford press; 2003.p. 217-242.
4. Wu M, Adams R. *Applying the Rasch model to psycho-social measurement: A practical approach.* Melbourne: Educational Measurement Solutions; 2007.
5. Croasmun JT, Ostrom L. Using Likert-Type Scales in the Social Sciences. *Journal of Adult Education*.2011; 40: 19-22.
6. Edmondson DR. Likert Scales: A History. Proceedings of *the 12th Conference on Historical Analysis and Research in Marketing (CHARM).*April 28-May1, California, USA; 2005.
7. McLeod S. Likert Scale. *Simplypsychology.org*. Retrieved January 27, 2014, from http://www.simplypsychology.org/Likert-scale.html/pdf.
8. Park K. Social Sciences and Health. In Park K, editor. Park Text Book of Preventive and Social Medicine. Jabalpur: Bhanot Publishers; 2013.p. 54x-54x.
9. Culbertson JM. What is an attitude? Journal of cooperative extension. 1963: 79-84.
10. Schwarz N, Bohner G. The Construction of Attitudes. In: Tesser A & Schwarz N, Editors. *Blackwell Handbook of Social Psychology*. Oxford: Blackwell; 2001. p. 436-457.
11. Singh YK. *Fundamental of Research Methodology and Statistics*. New Delhi: Newage International (P) Ltd. Publisher; 2006.

12. Lobsy J, Wetmore A. CDC Coffee Break: Using Likert Scales in Evaluation Survey Work [Internet].Cdc.gov. Retrieved January 12 , 2014, from www.cdc.gov/dhdsp/pubs/docs/CB_February_14_2012.pdf.
13. Tsang KK. The use of midpoint on Likert Scale: The implications for educational research. *Hong Kong Teachers Centre Journal*. 2012; 11:121-130.
14. Malhotra NK. Questionnaire Design and Scale Development. In: Grover R, Vriens M, editors. *The Handbook of Marketing Research*. California: Sage Publications, Inc; 2006.
15. Dawes J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and10-point scales. *International Journal of Market Research.* 2008; 50(1):61-77.
16. Boone HN, Boone DA. Analyzing Likert Data. *Journal of Extension*. 2012; 50(2). Available from http://www.joe.org/joe/2012april/tt2.php.
17. Colman AM, Norris CE, Preston CC. Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports.* 1997; 80: 355-362.
18. Preston CC, Colman AM. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent Preferences. *Acta Psychologica*. 2000. 104; 1-15.
19. Chang L. A psychometric evaluation of 4-point and 6-point Likert-type scale in relation to reliability and validity. *Applied Psychological Measurement*. 1994; 18:205-215.
20. Cox EP. The optimal number of response alternate for a scale: a review. *Journal of Marketing Research.* 1980; 17(4):407-422.

21. *Likert R. A Technique for the Measurements of Attitudes. Archives of psychology. 1932; 140(22) : 5-55.*
22. Pearse N. Deciding on the scale granularity of response categories of Likert type scales: The Case of a 21-Point Scale. *Electronic Journal of Business Research Methods.* 2011; 9(2):2011.
23. Finstad K. Response interpolation and scale sensitivity: evidence against 5-point scales. *Journal of Usability Studies*. 2010; 5(3):104-110.
24. Komorita SS and Graham WK. Number of Scale Points and the Reliability of Scales. *Educational and Psychological Measurement*. 1965; 25(4):987-995.
25. Miller GA. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review. 1956. 63:81-97.*
26. Carifio J and Perla RJ. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences.* 2007; 3 (3):106-116.