On the Treatment of Likert Data

Tyler Rinker

University at Buffalo

Department of Learning and Instruction

CEP 510: Psychometric Theory in Education

On the Treatment of Likert Data

Most researchers in the social sciences will come across Likert data derived from Likert scales at some point in their career. For myself this encounter occurred, while I am yet a doctoral student, when I was asked by a professor to analyze reading attitudes data derived from a Likert scale. As an eager student, I readily accepted the challenge, not really understanding the rabbit hole I was about to jump down. Likert scales have been available for psychometric purposes for 80 years, and few measurement tools are as sorely misunderstood and hotly contested than Rensis Likert's attitudinal metric known as Likert scales (Edmondson, 2005; Zand and Borsboom, 2009). This paper will guide the reader through (a) an introduction to Likert scales, including terms associated with Likert scales, (b) the historical evolution of the Likert scale (c) known biases of Likert scales, (d) the debate over the treatment of ordinal data as interval and (e) a discussion in how to approach Likert data. Failure to understand and reflect on issues surrounding Likert data can lead a researcher to make faulty inferences (Zand and Borsboom, 2009). It is important to become aware of Likert scales as a historical method with clear design, intent and underlying theory that has been thoroughly examined and debated by some of psychometrics' greatest minds.

An Introduction to Likert Scales

Likert scales are the creation of Rensis Likert and were first introduced to the field in 1932 in an article entitled, ''A Technique for the Measurement of Attitudes'' in *Archive of Psychology* (Edmondson, 2005; Likert, 1932). The Likert scale was designed to simplify the complexity of the Thurstone scaling technique (Edmondson, 2005, p. 127). Likert constructed his scale as a means of capturing an infinite number of attitudes in an ordinal scale format (Likert, 1932). Likert's scale ''presumes the existence of an underlying (or latent or natural) continuous variable whose value characterizes the respondents' attitudes and opinions...[e]ach Likert-type item provides a discrete approximation of the continuous latent variable.'' (Clason and Dormody, 1994, pp. 31-32). Likert was quite specific in his use and understanding of the scale that bears his name yet the Likert scale is often misunderstood and misused (Jamieson, 2004; Likert, 1932). Before we continue our discussion of Likert scales, it is important to define what is meant by the

term, examining the associated terminology.

Carifio and Perla (2008) state that a *Likert scale* is the summation of a combination of multiple Likert items. A *Likert item* is a single item in the form Likert proposes but in itself does not have the properties of the Likert scale and should not be used for analysis, as this is contrary to Likert's intentions (Carifio and R. Perla, 2007; Carifio and . Perla, 2008). It is common for researchers to confuse the term Likert scale with a Likert item leading researchers to incorrect analysis on individual Likert items (Carifio and R. Perla, 2007).

Uebersax (2006) offers a succinct list of descriptive characteristics original to Likert's (1932) article:

1. The scale contains several items.
2. Response levels are arranged horizontally.
3. Response levels are anchored with consecutive integers.
4. Response levels are also anchored with verbal labels which connote more-or-less evenly-spaced gradations.
5. Verbal labels are bivalent and symmetrical about a neutral middle
   and
6. In Likert's usage, the scale always measures attitude in terms of level of agreement/disagreement to a target statement (para. 11)

Uebersax's (2006) list provides a clear framework for defining the constitution of Likert scales and the items that comprise them. Characteristic five, which requires that the responses are to be symmetrical with a neutral center, indicates an odd number of choices with the non-neutral responses on either side being equal in magnitude yet opposite in polarity (Likert, 1932; Uebersax, 2006). Some researchers, in an effort to avoid *central tendency bias*, or the tendency to choose the neutral response for items the examinee lacks confidence with, stray from the odd number of response, eliminating the neutral choice (James, Demaree, and Wolf, 1984).

The terms Likert and Likert-type are often used interchangeably and may result in a use contrary to Likert's explicit framework for the scale's design (Likert, 1932). Likert's (1932) original summed scale was was derived from a five *point* (or number of discrete points of the responses) response score of multiple items, though he did not specify this quantity, utilizing the

following five point responses: strongly approve, approve, undecided, disapprove, and strongly disapprove (see Figure 1).

Figure 1: Example Likert Item

It is important to note again that the responses of an individual item do not constitute the scale, rather the summed scores from the responses. The summation of the items is the *Likert scale* where as the term *Likert item* is generally in reference to the format of individual items following in the arrangement of Likert (Clason and Dormody, 1994, p. 31). Items, and thus scales, that deviate mildly from Likert's defined characteristics are termed *Likert-type items*, though a radical departure should not be called Likert-type (Likert, 1932; Uebersax, 2006). Often such scales have more general names (i.e. *visual analogue scale*) that should be utilized (Flynn, van Schaik, and van Wersch, 2004; Uebersax, 2006). Though the distinction between Likert and Likert-type must be respected and reflected in a researcher's analysis, in order to avoid repetitive use of the clause ''Likert and Likert-type'' the remainder of this paper will refer to the term Likert to also subsume the term Likert-type.

A Brief History

Likert (1932) first proposed his scaling technique as a response to the labor intensive method of Thurstone scaling that required multiple judges to assess values for questions (Edmondson, 2005; Edwards and Kenney, 1946; Likert, 1932). The Thurstone scaling method was the first widely used attempt to capture latent variables on a continuous scale, however, the method suffered several drawbacks including the use of multiple judges which was costly and could potentially lead to judging bias(Edmondson, 2005; Edwards and Kenney, 1946; Likert,

Roslow, and Murphy, 1993). Likert's approach was to use a summation of carefully crafted, symmetric scored responses (Likert, 1932). Likert (1932) indicated that reliability scores for his new method were at least as good as Thurstone's and in some cases superior. Edwards (2005) cited that a possible rationale for the increased reliability over Thurstone's method may be attributed to the increased ''steps'' in the scale (p. 77). Likert responded to criticism of the reliability analysis by publishing a similar study in 1934 with similar results, though Edmondson (2005) finds design faults with this analysis as well (Edmondson, 2005, p. 128). Currently, Likert and Likert-type scales are used in many fields related to the social sciences and the scales have evolved to display differing number of points and responses, but the essence of Likert's original scale remains the same (see Vagias (2006) for a sample selection of these formats).

Biases Associated with Likert Scales

Researchers that produce or consume studies derived from Likert scales should be aware of potential biases and weaknesses attributed to the scale when constructing or examining a scale and items. James, Demaree,& Wolf (1984) describe one form of bias, *central tendency bias*, as the bias of examinees to choose the neutral response in an odd point scale, termed *forced choice*, as a way of avoiding items that they are not comfortable or confident in answering. Some attempts to overcome this bias have included the use of an even point scale, however, researchers need to be aware that this could alter the distribution of the data in ways that are more likely to lead to departures from the assumption of normally distributed error terms for linear models (Likert, 1932). A second known potential bias, *acquiescence bias*, is a phenomena in which the examinee tends to give positive responses to the survey questions is sometimes approached by reversing the polarity of the item (Lichtenstein and Bryan, 1965). Semon (2000) notes that acquiescence bias displayed differently depending on the cultural group of the respondent. This differences among cultures in responding to an item is referred to as *cultural bias* (Flaskerud, 2012; Semon, 2000). Chung & and Monroe (2003) describe another form of bias common to Likert scales, *social desirability bias*, in which ''[p]eople have a need to appear more altruistic and society-oriented than they actually are, and social desirability (SD) is the tendency of individuals

to deny socially undesirable actions and behaviors and to admit to socially desirable ones'' (p. 291). Researchers must be conscious of different biases that may affect the inferences that can be made from a study's findings.

The Problem of Likert Data

When I began my initial review of the literature around Likert scales it became abundantly clear that the scale's use was mired in hefty debate among psychometricians since its inception in 1932. The battle in the literature is over the treatment of an ordinal scale as interval. This argument affects the choices a researcher makes in analyzing the data and interpreting the results (de Winter and Dodou, 2010). The *conservatives* consider it a ''sin'' to use parametric tests to analyze what they consider to be strictly ordinal data (Knapp, 1990). The *liberals* contend that the robustness of the F ratio allows for Likert data to be treated as interval (Carifio and R. Perla, 2007; Knapp, 1990). The stance you take will likely affect how you conduct and interpret research and more importantly have to potential to affect type I and II error rates (Anderson, 1961). This sentiment is captured by Jamieson (2004):

> The legitimacy of assuming an interval scale for Likert-type categories is an important issue, because the appropriate descriptive and inferential statistics differ for ordinal and interval variables and if the wrong statistical technique is used, the researcher increases the chance of coming to the wrong conclusion about the significance (or otherwise) of his research. (p. 127)

I do not make a determination as to precisely how a researcher should approach Likert data but instead present the two major viewpoints and the resources to make an informed decision regarding a particular data set and research hypothesis.

S. S. Stevens: The Beginning of a Firestorm

Much of the literature traces the beginning of the ordinal-interval debate back to Stevens' (1946) piece, ''On the Theory of Scales of Measurement'', released 14 years after Likert's article (1932). Stevens (1946) wrote the article in an attempt to provide some guidance and unity in

measurement, particularly that of human perception. Stevens (1946) first defines measurement, ‘‘measurement in the broadest sense, is defined as the assignment of numerals to objects according to rules’’ and then proposes classifying data into one of four scales of measurement: *nominal*, *ordinal*, *interval* and *ratio* (pp. 677-678). He defines the nominal scale as data belonging to different categories with no clear order or zero point (pp. 678-679). It was here that using ‘‘football numbers’’ as an arbitrary assignment to nominal data was first discussed. The ordinal scale shares all the same properties as the nominal scale, but the categories could be ordered according to some rank. This is particulary relevant to Likert data as many, particulary the conservative psychometricians, would typically classify this scale as being ordinal (Edmondson, 2005; Jamieson, 2004). Here Stevens' (1946) warns:

> In the strictest propriety ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the rank-order of data...On the other hand, for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanctification: In numerous instances it leads to fruitful results. (679)

Stevens' (1946) third scale type is interval, which contains order, as the ordinal data, but the spacing between the ranked categories is equidistant. This scale classification is the first type mentioned that he considers ‘‘quantitative’’, though, like the ordinal and nominal scales, lacking a true zero point (p. 679). It is between the ordinal and interval levels that Stevens' (1946) acknowledgs some ‘‘ambiguity of such terms as ‘‘intensive’’ and ‘‘extensive’’. Both ordinal and interval scales have at times been called intensive, and both interval and ratio scales have sometimes been labeled extensive’’ (p. 678). The final scale of measurement, ratio, ‘‘are possible only when there exists operations for determining all four relations: equality, rank-order, quality of intervals and equality of ratios’’ (Stevens, 1946, p. 679).

Though Stevens (1946) article is an attempt to unify the field, it had the opposite effect. Seven years after Steven's piece, Lord (1953) wrote a parable in an attempt to produce a logical counter argument to Stevens' proposed scales and their application; this is considered the second

blow in the ensuing 80 year debate.

Frederic M. Lord: The Debate

Lord (1953), not wholly satisfied with Stevens' (1946) scale classification, used a football numbers story as a logical contradiction, based on Stevens own mention of football numbers as nominal data, to dispute Stevens claim that parametric statistics can not be used with data of the nominal scale. It should be noted that Lord's story does not mention Likert data, nor is it about treating ordinal data as interval, but is very relevant (and often sited) in that it goes further than using parametric tests for ordinal data, suggesting that such statistics can be applied to nominal data (Lord, 1953).

Lord (1953) essentially argues that a test designed for interval data can, in some instances, be applied to nominal data. He develops a story of a professor gone mad because of a love of ''calculating means and standard deviations'' of students' test scores, driven to insanity by his own quantitative hypocrisy (p. 750). In retirement the professor sells football numbers, a supposed arbitrary numeric assignment to nominal data, and is faced with accusations of selling a disproportionate amount of low numbers to the freshman class (p. 751). Without a method of testing the charge, the professor enlists the help of the campus's statistician, who promptly employees parametric measures to the nominal data (p. 751). The professor protests, ''But you can't multiply 'football numbers,' '' the professor wailed. ''Why, they aren't even ordinal numbers, like test scores.'''', to which the statistician retorts, ''The numbers don't know that'' (p. 751). The statistician defies the professor to disprove his applying parametrics to nominal data and the professor promptly sets out to do so through random sampling of his football numbers. After repeated samplings the professor is convinced that you can apply means and standard deviations to nominal data.

Following Lord's (1953) football numbers story a debate has raged over the appropriateness of treating nominal or even ordinal data with tests designed for interval and ratio data. Behan & Behan (1954) quickly retort, ''But, let us note that when we are all finished, we know something about the number signs, not something about the football players'' (p. 262)''. Bennett

(1954) humorously replied to Lord:

> So it is with Lord's parable. The freshman-sophomore argument settled by the statistician was one of cardinal highness or lowness in a set of numbers used in an entirely different context to identify football players. Our Professor X had best re-retire; his helpful statistical friend had best return to his TV set. I at least shall continue to lock my door when computing the means and standard deviations of test scores.' (p. 263)

Lord (1953) addressed Bennett's (1954) concerns with a more relaxed stance:

> It would be unfortunate if what has been written here were to lead anyone to ignore the very serious pitfalls actually present. Let me hasten to agree with Dr. Bennett that incorrect or meaningless conclusions can easily be reached...The conclusion to be drawn is that the utmost care must be exercised in interpreting the results of arithmetic operations upon nominal and ordinal numbers; nevertheless, in certain cases such results are capable of being rigorously and usefully interpreted, at least for the purpose of testing a null hypothesis. (p. 265)

This indicates his true intention, to warn against arbitrarily applying a statistical test without considering the data and measurement, to caution against apply rules rather than reason. Unfortunately, this article is cited much less frequently and the message is lost in the debate between the conservatives and liberals (Zand and Borsboom, 2009). Though Likert scales weren't explicitly mentioned in either Stevens (1946) or Lord (1953) the Likert battle of ordinal-interval uses both Lord and Stevens as spring boards for arguments of their position.

Current Views

The debate between ordinal and interval treatment of Likert data is still raging (Zand and Borsboom, 2009). Often the debate is in the logical realm, in theory, rather than practice (Lord, 1954). Lord (1954) gives permission to stop the debate by testing the charge, ''It seems very fortunate that any fundamental disagreement here between critic and statistician need not long remain a matter of opinion since the question is so readily submitted to wholly objective,

practical verification'' (p. 265). Knapp (1990) provides a great deal of insight into testing Lord's challenge, summarizing many of the arguments held by both sides. It becomes clear that the F test is quite robust, that ''one can usually tease normality and homogeneity- of variance quite a bit without doing serious injustice to t or F, particularly with equal sample'' (Knapp, 1990, p. 122). Knapp (1990) also discusses the possibility of the quantity of break points for the scale effecting the distribution of data, with more points tending to ''continuize'' (p. 123). Perhaps the most striking claim Knapp (1990) makes is regarding power, one of the underlying rationale for preferring parametric tests:

> But both camps are mistaken regarding... the, alleged power superiority of parametric tests over non parametric tests. The wilcoxon tests for independent samples and for paired samples are never much less powerful than t, and when the population distribution is not normal (for ordinal or interval measurement) they can be much more powerful (Blair & Higgins, 1980; 1985)...If you claim that-you have an interval scale, you are more likely to prefer parametric techniques, but should you have qualms about normality and/or homogeneity of variance and elect some nonparametric counterpart, don't be apprehensive about losing power; it maybe even-higher. (pp. 122-123)

Denny Borsboom, a respected leader in the field of psychometrics, co-authored a piece that takes a radical and sensible approach to the Likert scale debate. Zand & Borsboom (2009) attack Lord's (1954) argument in a different way than previous critiques. Rather than confront the theory and logic they discredit the logical contradiction of Lord's football number selling professor. Lord's (1954) argument rests on the fact that nominal data with numeric representation can be treated as parametric. Zand & Borsboom (2009) show that the numbers in Lord's story are not serving a nominal role but are a representation of the bias of the vending machines they're being distributed from (p. 72). At this point Lord's contradiction is debunked and no longer serves as an argument for treating Likert data as interval. Zand & Borsboom (2009) are not taking the position that Likert data should be consistently treated as non parametric; instead, they contend,

as Lord did, that ‘‘Stevens’ rules should not be applied mindlessly’’ (p. 74). Zand & Borsboom (2009) further the point, ‘‘The numbers don’t have to know where they came from; researchers have to know where they came from, since they assigned them in the first place.’’ (p. 74). The argument is transformed from one of ‘‘ordinal versus interval’’ to that of sound measurement methods, reflective research practices and consideration of the inferences made from statistical tests.

The Treatment of Likert data

Zand & Borsboom (2009) discredited Lord’s logical contradiction and called for attentive research practices, therefore, it is necessary to understand the direction a researcher must approach after making decisions regarding measurement, scale, analysis and inferences. It is important to realize, as stated by Carifio & Perla(2007), that the F test is actually quite robust to use of Likert data, even skewed data:

> The non-parametric statistical analyses only myth about ‘‘Likert scales’’ is particularly disturbing because many (if not all) ‘‘item fixated’’ experts seem to be completely unaware of Gene Glass famous Monte Carlo study of ANOVA in which Glass showed that the F-test was incredibly robust to violations of the interval data assumption (as well as moderate skewing) and could be used to do statistical tests at the scale and subscale (4 to 8 items but preferably closer to 8) level of the data that was collected using a 5 to 7 point Likert response format with no resulting bias. (p. 110)

Carifio & Perla(2007, 2008) also make it clear that this robustness only holds true when Likert data is analyzed as a scale, that is a summed composite score, not individual items. Anderson (1961) and Knapp (1990) warn that interactions may be affected more harshly than the main effects and need to be considered and analyzed carefully in using Likert data.

The consideration of power is the major concern for researchers in choosing parametric vs. nonparametric (Knapp, 1990). Knapp (1990) and Anderson (1961, 2004) both indicate that

under equinormality, power for both the parametric and nonparametric are close, and may actually be greater for the nonparametric under violations of the normality assumption. Zand & Borsboom (2009) also discuss the tendency to use parametric tests because of their availability and ease of use. As statistical computer programs, such as R (2012), become reflective of psychometric theory they have come to incorporate procedures for handling non parametric data with ease. It is absurd to allow ease of use to affect scientific inquiry that may lead to policy changes. Osterlind (2010) makes mention of more sophisticated graded response IRT models suited for Likert data that should become another tool in the researchers tool box (p. 298). It is incumbent upon the researcher to become familiar with appropriate techniques or to hire a professional statistician so that ‘‘‘measurement levels...guide the choice of statistical test’’ (Zand and Borsboom, 2009, p. 69).

Conclusion

It was the intention of this paper, not to give a ready to follow road map for analyzing Likert data, but to provide the insight and knowledge necessary for researchers to properly approach measurement, analysis and interpretation with a more cautious and informed perspective. When I first jumped down the rabbit hole of Likert data I was amazed at the misconceptions and ignorance I had. This paper is by no means comprehensive, but does provide the major considerations the reader should be aware of and can serve as a guide in exploration of this deeply seeded and inherently important topic that has been a source of debate and contention for generations of physchometricians. Zand & Borsboom (2009)offer great insight for the probing researcher attempting to grasp measurement with Likert scales:

> Research findings and conclusions depend on arbitrary, and usually implicit, scaling decisions on part of the researcher. This hinders scientific progress because it obscures a factor, namely the choice of scaling, that is influential in determining conclusions based on empirical research. It is important, therefore, to have a clear understanding of how level of measurement can affect our conclusions. (p. 69)

It is our duty to scrutinize our own data and the research of others as we attempt to build our collective understanding of various issues in social sciences. The Likert scale is a tool that may be useful but must be used with sensible understanding of the scale, its intended use, potential weaknesses, analysis approaches, interpretations of the results and of inferences gathered.

References

Anderson, N. H. (1961). Scales and statistics: Parametric and nonparametric. *Psychological Bulletin*, *58*(4), 305–316. doi:10.1037/h0042576

Behan, F. L., & Behan, R. A. (1954). Football numbers (continued). *American Psychologist*, *9*(6), 262 –263. doi:10.1037/h0053500

Bennett, E. M. (1954). On the statistical mistreatment of index numbers. *American Psychologist*, *9*(6), 264. doi:10.1037/h0059284

Bertram, D. (n.d.). *Likert scales ...are the meaning of life*.

Carifio, J., & Perla, . (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150fffd1152. doi:10.1111/j.1365-2923.2008.03172.x

Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116. doi:10.3844/jssp.2007.106.116

Chung, J., & Monroe, G. S. (2003). Exploring social desirability bias. *Journal of Business Ethics*, *44*(4), pp. 291–302. Retrieved from http://www.jstor.org/stable/25075038

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual likert-type items. *Journal of Agricultural Education*, *35*(4), 31–35. doi:10.5032/jae.1994.04031

de Winter, J. C. F., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation*, *15*(11), 1–12. Retrieved from http://pareonline.net/pdf/v15n11.pdf

Edmondson, D. R. (2005). Likert scales: A history. *CHARM*, *12*, 127–133. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCkQFjAA&url=http%3A%2F%2Ffaculty.quinnipiac.edu%2Fcharm%2FCHARM%2520proceedings%2FCHARM%2520article%2520archive%2520pdf%2520format%2FVolume%252012%25202005%2F127%2520edmondson.pdf&ei=8AqKT4WaD4W40QGN1K3mCQ&usg=AFQjCNHITtvhd9XFkFBUWVWseN7yuOmZPA&sig2=ERTm7gL9B8yUit44t84kAA

Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology*, *30*(1), 72 –83. doi:10.1037/h0062418

Flaskerud, J. H. (2012). Cultural bias and Likert-type scales revisited. *Issues in Mental Health Nursing*, *33*(2), 130–132. doi:10.3109/01612840.2011.600510

Flynn, D., van Schaik, P., & van Wersch, A. (2004). A comparison of multi-item likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, *20*(1), 49 –58. doi:10.1027/1015-5759.20.1.49

James, L. R., Demaree, R., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*(1), 85–98. Retrieved from http://people.sabanciuniv.edu/~gokaygursoy/ISTATISTIK_OLD/BOLUM_CALISMALARI/ILL/2007/OnlineSaglananDokumanlar/12192.pdf

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*, 1212–1218. doi:10.1111/j.1365-2929.2004.02012.x

Knapp, T. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research*, *39*(2), 121–123. Retrieved from http://www.mat.ufrgs.br/~viali/estatistica/mat2282/material/textos/treating_ordinal_scales[1].pdf

Lichtenstein, E., & Bryan, J. H. (1965). Acquiescence and the mmpi: an item reversal approach. *Journal of Abnormal Psychology*, *70*(4), 290 –293. doi:10.1037/h0022412

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55.

Likert, R., Roslow, S., & Murphy, G. (1993). A simple and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology*, *46*(3), 689 –690. doi:10.1111/j.1744-6570.1993.tb00893.x

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, *8*(12), 750–751. doi:10.1037/h0063675

Lord, F. M. (1954). Further comment on ”football numbers”. *American Psychologist*, *9*(6), 264–265. doi:10.1037/h0059284

Osterlind, S. J. (2010). *Modern measurement: Theory, principles, and applications of mental appraisal* (2nd ed.). Boston: Pearson Education.

R Development Core Team. (2012). *R: A language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Semon, T. T. (2000). No easy answers to acquiescence bias. *Marketing News*, *34*(3), 7. Retrieved from http://search.ebscohost.com.gate.lib.buffalo.edu/login.aspx?direct=true&db=bth&AN=3568208&site=ehost-live&scope=site

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. doi:10.1126/science.103.2684.677

Uebersax, J. S. (2006). Likert scales: Dispelling the confusion. Retrieved from http://john-uebersax.com/stat/likert.htm

Vagias, W. M. (2006). Likert-type scale response anchors. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University. Retrieved from http://www.clemson.edu/centers-institutes/tourism/documents/sample-scales.pdf

Zand, A. S., & Borsboom, D. (2009). A reanalysis of Lord’s statistical treatment of football numbers. *Journal of Mathematical Psychology*, *53*(2), 69–75. doi:10.1016/j.jmp.2009.01.002