Likert Scales: A History

Diane R. Edmondson, University of South Florida, USA

This article examines the origin of what is now known as the "Likert scale." The two defining articles by Rensis Likert are discussed, as well as several weaknesses not originally mentioned. A detailed literature review examining all articles written in the Journal of Marketing that utilized at least one Likert scale was completed. Finally several future research ideas are discussed.

Rensis Likert changed the way attitudinal research was conducted with his creation of what is now known as the "Likert scale." Although his original intention was not the creation of a new scale, the development of this scale earmarked a shift in how attitudinal research was performed, primarily because of the ease which a Likert scale can supposedly be created and utilized. However with the creation of the Likert scale came several other issues such as researchers' confusion over whether this scale is ordinal or interval in nature. This confusion leads many individuals to utilize statistical methods such as means and standard deviations which are not appropriate for Likert scales.

One of the primary purposes of this particular study is to follow the changes in the usage of the Likert scale from its creation back in 1932 until today (third quarter of 2004) through the analysis of one journal, the *Journal of Marketing*. This journal, created in 1936, yielded 194 articles that utilized at least one Likert scale. A descriptive analysis of the data will then be performed.

The present paper is organized as follows: a brief background of Rensis Likert is given followed by a review of the key concepts of the original Likert scale and a discussion of the scale's major weaknesses. Then the methodology being employed is discussed, results are examined and finally future research ideas are suggested.

ABOUT THE CREATOR

Rensis Likert was born on August 5, 1903 in Cheyenne, Wyoming. He obtained a BA degree in economics from the University of Michigan in 1926 and his Ph.D. from Columbia University in 1932. Likert's philosophical background was primarily functionalistic. While at Columbia, Likert became interested in social psychology and had a vision for survey research. It is this vision that led him to create what is considered by many to

be his best known work: the "Likert Scale." Besides scale development issues and the development of the "Likert Scale," Likert also spent a great deal of his professional career developing theories of business management such as participative management theory and the conceptualization of four types of leader behavior. He was a forward thinker and pushed psychological thought into new frontiers. Likert passed away on September 3, 1981 in Ann Arbor, Michigan. He is survived by his wife and fellow professional colleague, Jane Gibson Likert, and their two daughters, Elizabeth and Patricia (Campbell 1988).
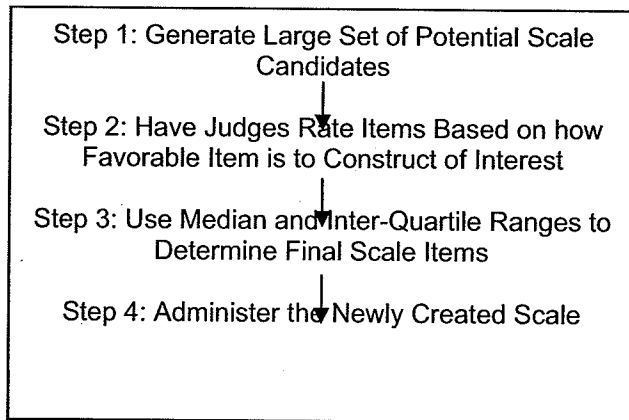
KEY CONCEPTS OF THE ORIGINAL LIKERT SCALE

The original idea for the Likert scale is found in Rensis Likert's 1932 article in *Archive of Psychology* titled, "A Technique for the Measurement of Attitudes." This idea was expanded by Likert's 1934 *Journal of Social Psychology* article titled "A Simple and Reliable Method of Scoring the Thurstone Attitude Scales."

The rationale, according to Likert, for why the 1932 article was written was to reflect on and expand the present knowledge of the procedures developed by Thurstone. The Thurstone scaling technique (see Figure 1 on the next page) involves generating a large set of potential scale items, then having a group of judges or raters rate each statement regarding how favorable the scale item is to the desired concept or construct, and then finally using the median and inter quartile ranges based on the raters responses for each statement to determine which scale items to use in the final scale. Because of the immense amount of work and the laborious assumptions that are necessary to utilize the Thurstone scaling technique, Rensis Likert took a radical departure from the technique when creating what is now known as the "Likert scale." This radical departure from the Thurstone technique was primarily in the usage of judges when developing the scale since Likert's technique did not require judges while Thurstone's method did.

The main problem with all attitudinal research, according to Likert (1932, 7), is that "the number of attitudes which any given person possesses is almost infinite." However Likert (1932, 8) also states that there are "certain discernible groups of social responses" and that a response to an attitude, although not inflexible or rigid, can move only within a certain range. Likert stated that there were two prevailing definitions of attitude yet his

preferred definition is such "that attitudes are dispositions toward overt action (Likert 1932, 9)," and that these attitudes can be generally "clustered or linked together" (Likert 1932, 9).

Step 1: Generate Large Set of Potential Scale Candidates

↓

Step 2: Have Judges Rate Items Based on how Favorable Item is to Construct of Interest

↓

Step 3: Use Median and Inter-Quartile Ranges to Determine Final Scale Items

↓

Step 4: Administer the Newly Created Scale

FIGURE 1
THURSTONE SCALING TECHNIQUE

The questions utilized in testing the various survey methodologies employed in this article were taken from preexisting scales whenever possible; however, in some cases the questions needed to be abbreviated, simplified, and/or made less ambiguous. As much as possible, the questions should have permitted the respondent to make a judgment of value between two clearly opposed alternatives rather than just strictly a judgment of fact. There were several different constructs being measured on the questionnaire including internationalism, the Negro, and imperialism.

In the initial questionnaire, there were several different question types presented. The first type was a series of multiple choice questions with five possible answers. The next type was a series of propositions with five response options consisting of strongly approve, approve, undecided, disapprove, and strongly disapprove. The third question type used the same five response options mentioned previously; however, the respondent had to respond to the outcome of different newspaper narratives in which a conflict was presented. The questionnaires were given to, primarily male, undergraduate students across the United States. A total of 650 randomly selected students from seven universities were included in the sample although there were 2000 students from nine universities initially sampled.

The results were scored using a variety of methodologies including the Thurstone method, the Sigma method, and several simpler methods. The first comparison was made between the Sigma method, which assumes that attitudes are distributed normally, and the simpler method. The simpler method of scoring involved the assigning of values from 1 to 5 to each of the different positions on the five-point statements. A value of 1 was always assigned to the negative end while a value of 5 was assigned to the positive end. At first, the mean for each individual was calculated; however since the number of statements per individual was identical, a sum of the numerical scores was used. All in all, the results were similar between the Sigma and simpler methods employed; however, the use of the simpler method yielded additional advantages such as the ability to combine the different question types.

To compare the Thurstone method with that of the simpler method, one of the groups of respondents were "asked to indicate whether he strongly agreed, agreed, was undecided, disagreed, or strongly disagreed with each statement in the Thurstone-Droba War scale, Forms A and B" (Likert 1932, 34). After accounting for questions that could not be easily transformed from the original Thurstone scale to the new agreement scale due to the fact that the initial question was double-barreled or had some other problem, the two scales were compared. Results found that using the simple method yielded similar reliabilities as the Thurstone method with half as many items. In other words, the simpler method, created by Likert, yielded higher reliabilities than the Thurstone method when all items were included.

In 1934, Likert completed a follow-up to his 1932 publication. This study was undertaken in order to show if the results from the 1932 study can be consistently proven. Ten different scales were used including: attitude towards birth control, the Chinese, Communism, Evolution, the Germans, God (Reality of), God (Influence on Conduct), the Negro, War (Droba), and War (Peterson). The directions at the top of the attitude scales stated the following: "If you agree with a statement put a plus; if you strongly agree with a statement, put a plus with a circle around it; if you disagree with a statement put a minus; if you strongly disagree with a statement, put a minus with a circle around it; if you are undecided, put a question mark" (Likert 1934, 229). After each respondent indicated their reaction or agreement to the statements, the scales were then scored by using the Thurstone method. In each of the scales, there were certain statements that could not be utilized because "it was found impossible to determine whether to assign a value of 1 or 5 to the "strongly agree" alternatives" (Likert 1934, 230). A total of 35 questions were deemed unusable. The rest of the results were scored using both the Thurstone and simple method. Like the previous study, all of the respondents were again male students. Scoring using the simple method occurred by summing every response by each individual in order to create an individual score. Results indicated that when correcting for the number of items used, the simpler method was found to possess higher reliabilities than the Thurstone method for all ten attitude scales. Results also found that the correlation between the two methods were consistently high "which indicates that the simpler method of scoring is measuring essentially what is measured by the Thurstone scoring method" (Likert 1934, 234).

MAJOR WEAKNESSES/ISSUES WITH ORIGINAL LIKERT SCALE

Originally the Likert scale was created in order to alleviate the need for the usage of judges when developing scales. However judges may still be needed when creating scales using the simpler or Likert method. All of the scales used when testing this new methodology were scales that had already been tested previously by experts or judges. Therefore it is impossible to state that judges are not needed when developing Likert scales since the scales used to create the Likert scoring technique had previously used the judge rating technique. In fact, Trochim (2001) states that the procedure for developing a Likert scale consists of the following steps: generating the items, rating the items, selecting the items, and finally administering the scale. The rating of the items step, according to Trochim (2001), consists of using judges to rate the items on a one to five scale regarding how favorable the item is to the concept or construct. This rating procedure is similar to the rating procedure utilized in the creation of a Thurstone scale. Although the original scale was created in order to remove judges from the process of scale creation, judges are still needed in a similar fashion as with Thurstone scales.

Another weakness of the original Likert scale is the statement that either a mean score or summated score could be utilized. An inherent assumption with the usage of any Likert scale is that although the scale is truly ordinal in nature, it is assumed to be on an interval scale with which statistical properties such as the mean can be justifiably used. However this assumption is never mentioned in the original Likert study. In fact, this assumption is down right incorrect. Just because a 1 to 5 scale is used to show the level of agreement with a particular scale item and construct does not mean that the average score should be used when conducting statistics on the results. The 1 to 5 scale is ordinal in nature rather than a true interval scale; therefore, employing means and standard deviations is incorrect. Usage of these interval scale statistics when the scale is truly ordinal is considered a subtle sin in marketing research, according to Martilla and Carvey (1975). The use of 1, 2, 3, 4, and 5 to represent strongly disagree, disagree, undecided, agree, and strongly agree is strictly a coding structure. It is impossible to state that the difference between strongly agree and agree is the same as the difference between agree and undecided. Because of this, assuming an equal distribution or equal intervals between each data point so that a mean can be used is an improper assumption although it is an assumption made quite frequently in empirical studies. Instead the median or mode should be used to describe the central tendency of the data and the range, percentiles or inter quartile range should be used to describe variability in the data.

Besides the judge and statistical issues, one other flaw or cause of concern with the original Likert studies is in the usage of all student samples. It is not practical to assume that students would have the same attitudes on the selected topic areas as non-students; therefore, the amount of variability that exists between student samples versus the amount of variability that exists in mixed samples is probably different. It appears that the sample was based primarily on a convenience sample as there is little rationale given as to why students were selected over other possible samples for these articles. It is possible that using non-student samples would have led to different results regarding the comparability of the various survey methodologies employed; yet this issue was not even remotely mentioned in the original study. The fact that a large percentage of the students were male is another issue which should have been mentioned. Again it is unlikely that males and females have the same attitudes regarding the topics addressed in the original Likert studies; however, the possibilities of these differences were not alluded to in the article. Therefore it is impossible to state for certain that the same conclusions regarding the similarity of the various survey methodologies (Thurstone, sigma, and Likert) would have occurred if a mixed (student/non-student and male/female) sample of respondents would have been surveyed instead of strictly male, student samples. Although the large percentage of male students is a factor of the times since few females were enrolled in college in the 1930's; in 2004, the usage of almost entirely male samples would be deemed problematic and any results obtained may be questioned.

METHODOLOGY

In order to examine the changes in the Likert scale over time, a review of the usage of the Likert scales in the *Journal of Marketing* was conducted. How the relevant articles were identified and coded, including the applicable inclusion criteria is briefly discussed below.

Literature Review

There were several strategies utilized in the attempt to identify relevant studies. First, a computer search of ABI/Inform was completed. The search was limited to articles in the *Journal of Marketing*. The rationale behind why only this journal was considered was due to the intent of this study being to examine the changes in the usage of Likert scales in the marketing field since its 1932 creation. Since the only marketing journal that goes back almost 70 years is the *Journal of Marketing* (1936), it was selected as the journal to search. Any studies that contained the term *Likert or scale* in their title, abstract, and/or full text were considered. The reason that the search term *scale* was used was because some studies may have used a Likert scale yet not explicitly stated the word *Likert* in the article; therefore, a search strictly for the word *Likert* would not have resulted in these relevant articles being found. Because ABI/Inform does not allow full text viewing for the last twelve months, the second strategy undertaken was to manually search all 2003 issues and the first three 2004 issues of the *Journal of*

Marketing. There were a total of 84 articles found using the word *Likert* as well as 1086 articles using the word *scale*; however, not all of these articles met the inclusion criteria shown below. In fact, only 129 articles out of these 1170 possible articles were applicable for this study. There were also a total of 21 relevant articles found in the current 2003 and 2004 *Journal of Marketing* issues. Because of some inconsistencies in the ABI/Inform search engine results, it was determined that several years were not being adequately searched; therefore, a manual search of years 1980 through 1991 was also conducted, yielding 44 additional studies. In total, there were 194 articles that met the following inclusion criteria.

Inclusion Criteria

The first major criterion for inclusion was that the article had to be in the *Journal of Marketing*. Because the *Journal of Marketing* was developed in 1936 and because it was the earliest marketing journal written, it was determined that a review of this journal would enable the author to examine the changes in the usage of the Likert scale over the time frame 1936 to 2004. The second criterion for inclusion was that only the original Likert format, which examined agreement, was permitted. Any article that utilized a Likert-type scale such as one for frequency (always, often, about half the time, seldom, never) or satisfaction (very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied, very dissatisfied) were not included in this analysis.

Study Coding

There were several variables coded for each study deemed relevant for this study. The variables used in this analysis include sample size, demographic variables, number of scale points, whether any Likert study was cited (i.e. 1932, 1934, 1967), what statistics were reported (mean, median, etc), and job and industry classification. Demographic variables included variables such as the percentage of male respondents and average age of the respondent. Although no formal meta-analysis was conducted due to the apples-and-oranges issue that exists when measuring the reliability of different constructs, a descriptive analysis was conducted. Although many studies may possess more than one Likert scale in the questionnaire design, only one Likert scale from each independent study was considered in the descriptive analysis. The reason why only one Likert scale per independent study was used in the descriptive analysis was to make sure that a study with a large amount of different Likert scales did not receive an unequal weighting. Fortunately, for the most part, the information coded specifically for a scale such as the number of scale points was consistent throughout a study, regardless of the number of scales a particular study possessed. If, however, a study did utilize several different scale point numbers (i.e. some 5-point and some 7-point scales), then the number of scale points that possessed the majority was recorded.

ANALYSES

Out of over 1200 possible articles, 194 articles were deemed relevant for this study. Although there were 194 articles, there were 219 independent studies included in the analysis. There were a total of 94,669 respondents included in this analysis. A detailed listing of the relevant variables coded for this study can be found in the Appendix[i].

An initial analysis shows that the usage of Likert scales has increased since originally developed. This is shown by the increase in the frequency of studies using a Likert scale over time (see Figure 2 below). However there are two other interesting findings which should be noted. First of all, there were no articles found that utilized a Likert scale prior to 1970. Therefore it can be inferred that even though the Likert scale had been invented back in 1932, its usage in marketing did not begin until almost forty years later. One possible reason for why the Likert scale was not used before 1970 is because of the differences in what was being published in the *Journal of Marketing*. The early editions of *Journal of Marketing* tended to have a great deal of theoretical and secondary research studies. There were little empirical studies performed; hence, no surveys using Likert scales were completed. Another possible reason for non-usage in the *Journal of Marketing* could be that marketing researchers did not read psychology and economics-type journals; therefore, some marketing researchers may have been unaware of the creation of the Likert scale. A second interesting finding visible in this figure is that there appears to be lulls in the usage of Likert scales throughout time (for example, 1982 and 1991). It should be noted that some variability is expected considering the same number of articles are not published every year. However the fact that only a few studies utilizing a Likert scale were published in these years is very interesting. One possible reason for this limited usage is that articles using Likert scales may have appeared in other journals besides the *Journal of Marketing*. For example, studies using the Likert scale may have appeared in other marketing journals, such as the *Journal of Marketing Research* and *Journal of Consumer Research*.

Table 1 displays some overall descriptive statistics for the 219 independent studies included in the analysis. This table looks at the countries included in the analysis, the type of respondent being sought (consumer, employee, etc), the type of job held, the central tendency and variability measure reported, and whether a Likert study was referenced. The weighted average number of scale points and the weighted average percentage of male respondents can also be found in this table. These two figures were calculated after weighting by sample size. When looking at this table, several interesting findings are evident.
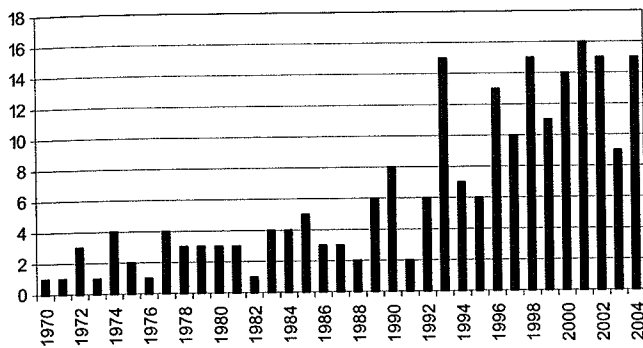
FIGURE 2
FREQUENCY OF LIKERT STUDIES BY YEAR

First of all, there were 0 studies found that referenced one of Likert's seminal pieces. Although it is impossible to know exactly why these authors did not refer back to Likert's original work, it is easy to create some possible reasons for why Likert was not cited. One possible reason is because of the fact that there were no articles written in close proximity to one of Likert's seminal pieces (1932 or 1934). Considering there were no Likert scales used in the *Journal of Marketing* articles from 1932 until 1970, there is a possibility that authors did not examine articles over forty years old. Another possible reason for why Likert was not cited was because of the uncommon journal that the seminal article was published. *Archives of Psychology* is not a readily available journal unless a researcher was actively looking for it; therefore, it is quite possible that searches for the seminal piece yielded no results. A third rationale is that the search capabilities before online searches such as ABI/Inform or PychINFO made researching and finding older articles extremely tedious and complicated. A fourth and final possible reason for why Likert was not cited could be the fact that Likert's original work was founded in Psychology. Until individuals like Wroe Alderson (1957), there were very few concepts taken from other disciplines outside of marketing and economics. Because of this, it is possible that until researchers actively pursued areas outside of the business realm, Likert's work was not considered.

As previously stated, only ordinal statistical techniques such as modes, medians, percentiles, inter quartile ranges, and ranges should be utilized when examining Likert scales (Martilla and Carvey 1975); however, many studies continued to employ interval statistical techniques such as means and standard deviations. In fact, when examining whether means, modes or medians were utilized, there were only two studies that actually listed the median while there was 114 studies that listed the mean. Approximately 47% of the studies did not list any measures of central tendency (which obviously is acceptable); however, the remaining 52% specified the mean. Figure 3 below displays the frequency of the central tendency type by year. Looking at this figure, it is evident that the usage of means did not

become prevalent until about 1993. In fact, 94 out of 114 studies that used the mean as the reported central tendency measure did so on or after 1993.

TABLE 1
DESCRIPTIVE STATISTICS

Variable	Item	N	%
Country	Belgium	2	1.29%
	Canada	3	1.94%
	China	2	1.29%
	Germany	1	0.65%
	Holland	1	0.65%
	India	1	0.65%
	Japan	1	0.65%
	Multinational	15	9.68%
	Netherlands/ Scandinavia	3	1.94%
	Norwegian	1	0.65%
	Unknown	17	10.97%
	United States	108	69.68%
Respondents	Both Employees and Consumers	1	0.46%
	Consumer	82	37.44%
	Dyad	4	1.83%
	Employees	128	58.45%
	Other	4	1.83%
Job Type	Managers/Executives	50	22.83%
	Other	22	10.05%
	Purchasing agents	10	4.57%
	Sales	29	13.24%
	Students	20	9.13%
	Unknown	66	30.14%
	Various	22	10.05%
Weighted Avg. % of Males*		48.95%	
Weighted Avg. # of Scale Pts*		6.56	
Central Tendency Measure Reported	Mean	114	52.05%
	Median	2	0.91%
	None	103	47.03%
Variability Measure Reported	Inter quartile Range	1	0.46%
	None	120	54.79%
	Range & St. Dev.	1	0.46%
	St. Dev.	97	44.29%
Likert Referenced	Yes	0	0.00%
	No	219	100.00%
* Weighted by Sample Size			
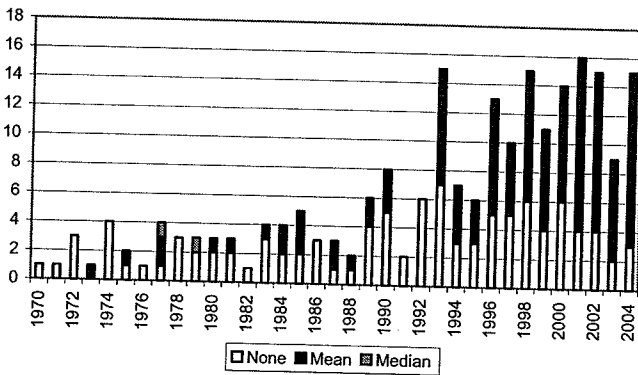
FIGURE 3
FREQUENCY OF CENTRAL TENDENCY BY YEAR AND TYPE

When looking at the variability measures utilized, this trend of using an interval statistical method instead of ordinal statistical method since the mid-1990s is also present (Figure 4). Again it is clearly evident that the usage of standard deviations was not common until 1993 as, for the most part, no measures of variability were given until that time. Because of this, there is evidence that more studies utilized statistical tools geared towards interval scales rather than using the statistical tools designed for ordinal (Likert) scales.
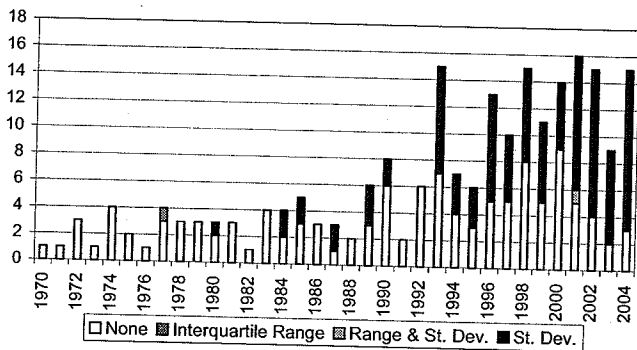
FIGURE 4
FREQUENCY OF VARIABILITY BY YEAR AND TYPE

FUTURE RESEARCH

The research completed in this paper focused primarily on one key marketing journal, the *Journal of Marketing*, because it is the only journal that has existed since around the time the original Likert scale was invented. However through this analysis, it has been determined that, in the marketing field, Likert scale usage did not really start climbing until the mid-1970s. Because of this, further research that reviews the usage of Likert scales in other major marketing journals, especially the *Journal of*

Marketing Research, should be conducted to determine if similar results are found.

It has also been shown that one major assumption, albeit an unstated assumption, made by a majority of researchers is the consideration of the Likert scale as an interval scale rather than an ordinal scale. The usage of Likert scales as an interval scale implies that means and standard deviations can be utilized when making interpretations of the results; however, this has been shown to be incorrect. However this leads to a future research study that needs to be completed, which is a determination of how the usage of means and standard deviations over medians/modes and inter quartile ranges changes the results of the study. For example, it would be useful for a series of studies to be completed which fully examines this impact.

Another area where future research regarding Likert scales needs to be completed is in the completion of a series of bona fide meta analyses which examines several different constructs that utilize a Likert scale in order to determine the impact that this scale has on results. Since it would be virtually impossible to actually code every study ever completed that used a Likert scale, coding a subset of studies on specific constructs can be used to determine what impacts this scale has. Completing a series of meta analyses also eliminates one of the major problems with meta-analysis, the problem of comparing apples to oranges. By measuring the same construct, this problem is alleviated.

Throughout this literature review, it was determined that several different data collection methods, such as surveys and experiments, were conducted. However no study found examined the impact that the data collection method had on the results. For example, a researcher may find that the results for the same Likert scale using a mail survey may differ from the results when the data is collected using a mall intercept. Because of this, a series of studies need to be completed that examines how and if Likert results change based on the data collection method employed.

Furthermore, the original Likert study has not, to my knowledge, ever been re-tested to see if similar results would exist. Because of this, a similar study to Likert's original studies should also be conducted. Prior to conducting the study; however, a few changes need to be made. First of all, the sample should not consist of strictly student samples. The sample should also be based on the population and; therefore, both females and males should be included in the sample. Finally the scales being employed will need to change as these scales, such as the Negro scale, are no longer relevant in today's society.

NOTES

[1]   The Appendix is available upon request

REFERENCES

Alderson, Wroe. 1957. *Marketing Behavior and Executive Action: A Functional Approach to Marketing Theory.* Homewood, IL: Richard D Irwin, Inc.

Campbell, A. 1988. *International Encyclopedia of the Social Sciences, Biographical Supplement.* New York: The Free Press.

Likert, R et al. 1934. A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*: 228-238.

Likert, R. 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology* 140: 5-55.

"Market and Survey Researchers." U.S. Department of Labor, Bureau of Labor and Statistics. Accessed on April 20, 2004 at http://stats.bls.gov/oco/ocos013.htm

Martilla, John A. and Davis W. Carvey. 1975. Four Subtle Sins in Marketing Research. *Journal of Marketing* 39 (January): 8.

Trochim, William M.K. (2001). *The Research Methods Knowledge Base.* Cincinnati, OH: Atomic Dog Publishing.

References for the *Journal of Marketing* articles are available upon request.