The Likert Scale:
A Proposal for Improvement Using
Quasi-Continuous Variables

Carl J. Chimi
cchimi@bloomu.edu
Bloomsburg University 17815 USA

David L. Russell
drussell@wnec.edu
Western New England College
Springfield, MA 01119 USA

Abstract

This paper discusses Likert-type items on a survey instrument, as commonly used is the social sciences and marketing research (often referred to as a "Likert scale"). It briefly discusses antecedents to the Likert item, but focuses more on its limitations, primary among which is the fact that it generates qualitative data. Derivatives of the Likert scale are presented and their limitations are discussed. An alternative that uses information systems to derive quasi-qualitative data is presented, and the analytical advantages are discussed.

Keywords: Likert, Likert-scale, Quantitative analysis, Computer-based surveys

1. INTRODUCTION

The Likert scale is ubiquitous in nearly all fields of scholarly and business research. It is used in a wide variety of circumstances, among them:
- when the value sought is a belief, opinion or affect;
- when the value sought cannot be asked or answered definitively and with precision; and
- when the value sought is considered to be of such a sensitive nature that respondent would not answer except categorically in large ranges.

In this paper, we will focus on the first use of the Likert item stated above. We will use the generic term "response" in place of more specific descriptors like "belief", "affect", opinion" and similar terms. We will use the term "Likert item" as a generic descriptor of the original Likert item and its derivatives.

In its classic form, the Likert item consists of two parts: (1) a positive statement about some feeling, belief, opinion or affect; and (2) a series of responses representing a breadth of potential responses. Most typically, there are five responses designated "Strongly Disagree" though "Strongly Agree" (Aaker & Day, 1998, p. 285). Further, in its classic form, the responses are presented vertically and centered below the statement; commonly "Strongly Disagree" is listed at the top of the list and "Strongly Agree" is listed at the bottom. An example is shown in Figure 1.

The Likert-type scale is also used to capture qualitatively data that is (1) difficult to measure or (2) addresses a sensitive topic, to which a respondent would likely not respond, or would response falsely, if asked directly. An example of the latter would be "What is your salary?". Since most respon-

dents would not answer this item fully or honestly, a series of categories is offered, from which the user chooses the category representing his/her salary. In many instances, the categorical responses include "less than $X" and "greater than $Y" categories. Such open-ended response categories prevent the inference of mean using expectancies based on midpoints. Further the closed-end categorical responses in between tend to be sequential but not linear. As useful as this aspect of the Likert approach is, it is not the subject of this paper. Thus, this paper addresses responses sought in some categorical item grouping, commonly but not always in the "Strongly Disagree" … "Strongly Agree" scale

---

I feel the United States economy is strong.

Strongly Disagree   ☐

Disagree   ☐

Neutral   ☐

Agree   ☐

Strongly Agree   ☐

Figure 1: Classic Likert Item

The Likert item is so ubiquitous in everyday life that the respondents may not be aware of its use. A modern application is the Wong-Baker FACES scale (Hockenberry, 2004, p. 301) used to assess pain in children and others who do not articulate well. The scale uses a semantic-differential horizontal Likert- item (except that the semantic different is expressed is a scalar fashion and is associated with each response), and innovatively uses representational faces as categorical responses.
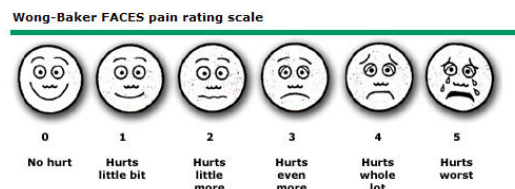
Figure 2: Wong-Baker FACES scale

2. PROBLEMS WITH THE LIKERT ITEM AND ITS DERIVATIVES

Despite its ubiquitousness, we submit that there exist several limitations with the Likert item and in more recent derivatives:

- We submit that the response being elicited through use of the typical Likert item is not static but actually a dynamic, quantitative, and continuous response that is captured poorly by existing Likert items.
- We submit that our ability to analyze, study and draw inferences from such data has been impeded by a limited number of discrete points available for analysis since instruments using Likert-type items generate results of course granularity.
- We submit that the Likert item does not sufficiently address or account for cases of respondents (1) who have sufficient knowledge about the subject of study, but who do not have a response toward it and (2) who are insufficiently knowledgeable about the subject of student to be able to form a response.

In this paper, we will discuss these limitations in detail and offer a solution that addresses many of the limitations posed by the Likert-type item. Note again that we do not address the use of Likert items aimed at sensitive topics like salary and age.

a. Likert Items Provide Coarsely Granular, Qualitative Responses.

When using a Likert item instrument, the response is recorded in one of a small number of discrete categories. Summary results are limited to the count of responses in each category. The categories are intended to contain the full breadth of response on the item being studied, leading to the implicit assumption that the categories "Strongly Disagree" … "Strongly Agree" covers all responses. We submit that this assumption is inadequate. Let us use the "Strongly Agree" categorical response for discussion. We propose that even stronger degrees of agreement are possible (for example, "Absolutely Believe" is stronger statement of agreement than "Strongly Agree"). This leaves the res-

pondent with such an extreme response with a conundrum. Does she or he respond with "Strongly Agree", though that response variable is weaker than the individual's true response? Or, if permitted, does the respondent bypass the question? This problem stems from the fact that the commonplace "Strongly Disagree" … "Strongly Agree" scale is of such coarse granularity that it insufficiently captures the breath and complexity of responses. The issue of "skips" is an important one to which we will turn shortly.

To summarize: the Likert-type item inhibits the capture respondents' beliefs, affect and other responses at a fine level of granularity by (1) not sufficiently covering extremes of response; (2) forcing responses into a limited number of coarsely granulated categories that (3) are sequential but not necessarily linear deriving results as a count of each category representing a qualitative variable that is assumed to be ordinal but like not scalar.

We leave for a later point a discussion about the middle or "Neutral" response.

b. Lack Of Many Statistical Analysis Techniques.

Statistics textbooks and textbooks focusing on many application-specific quantitative methods point out the richer, subtly varied and more powerful analyses available when quantitative data is available (Hildebrand & Ott, 1991, p. 211), (Dillon, Madden, & Firtle, 1993, pp. 134-135). The fundamental problem with Likert items is the response they generate: qualitative data consisting of a count of responses in each category. The resulting analysis is inherently limited, primarily to a frequency table, typically with relative and cumulative relative frequencies computed. Summary statistics can only be inferred using expectancies, and then only if there is some numeric basis for the each response category; however the assumption is often made that the Likert item is interval in nature, and estimated means are commonly computed. It is our argument that this assumption is incorrect. The data generated by an instrument based on Likert items simply cannot be subjected to the more robust, more powerful and more subtle analyses available with quantitative data.

c. Incomplete And Insufficient Accounting For The "Neutral" Response.

Likert items most often have five response categories, although seven is not uncommon. Almost invariably, however, the number of response categories is odd. (The FACES scale for pain assessment, shown above, is an exception.) This presents a problem regarding the center category. Commonly, this category is marked "Neutral" or "Neither Agree or Disagree", but in many cases the category carries no label. To be "neutral" presupposes that the respondent knows about the subject of study, has considered it, and finds that his/her response falls roughly center between the two endpoints.

The ambiguity associated with the middle response category is problematic. Indeed, such a response could mark true neutrality on the item posed, in which case its selection is appropriate. We propose, however, that there are at least two other responses that would lead a rational respondent to pick the center category. They are:

1. The respondent is knowledgeable about the subject matter at hand and has a basis upon which to form a response. However, the respondent is not neutral on the matter. Rather, this respondent simply does not care, one way or the other, about the subject of study. The respondent has considered the matter, but the subject of study is of so little importance to him or her that no response has been developed relative to them. The respondent is indifferent far more than s/he is neutral.
2. The respondent lacks sufficient knowledge to form a response on the subject of study. The respondent cannot be neutral since s/he lacks sufficient knowledge to know what s/he is neutral about.

Herein lies the challenge, we propose. The three states of mind discussed above are three entirely different matters, and yet are subsumed under the single response "Neutral". By doing so, the Likert item, as commonly used, presents the opportunity to confound the meaning of "Neutral".

A further problem exists: in some cases, the center "neutral" response is the default response. As the passive default response, no other response will be recorded unless the respondent actively makes another selection. Thus, a fourth state, "No action" can be subsumed under the "Neutral" response. This is especially a problem if item by-pass is permitted. The problem is, we do not know if the respondent bypassed the question because s/he truly is neutral (in effect, utilizing the assumption built into the default setting of the center category), or whether s/he knew about the matter and simply did not care, or whether s/her was not knowledgeable, or simply did not exert the effort to address the item.

In fact, other responses could result in a disproportionately larger "neutral" response. Two possibilities include (1) a degree of illiteracy or lack of fluency such that the respondent could not comprehend the question and (2) a perception by the respondent of being rushed, causing them to select "Neutral" to simply be done with the item. The latter is especially a problem when question by-pass is not permitted. Thus two additional states, illiteracy and insufficient attention due to time pressure, can be seen to be subsumed in the "Neutral" response. In the interest of space, the "No Action", illiteracy and rushed response cases are not addressed in this paper. However, the fact that we can reasonably propose five additional states of mind in addition to true neutrality being subsumed in the "Neutral" category suggests that responses from instruments dependent on Likert-based items might have significant bias to the center.

As a brief example, note that in parallel studies in which a placebo is used, one would expect that in the placebo category "neutral" would be the only response. That is generally not the case, and responses on both sides of "neutral" are often found. This further demonstrates the confused nature of the "Neutral" response.

d. Likert item derivatives

Although not part of Likert's original work, his concept has been applied extensively using semantic differentials. Here, instead of employing an Agree/Disagree scale, the response limits consist of two "anchors" consisting of dichotomous words or phrases, and typically include at least three intermediate points. Two of these intermediate points are phrased representing tendency toward one or the other of the dichotomous anchors. The center category sometimes uses phrasing (often "Neutral") intended to convey some intermediate position between the two dichotomous extremes. Most typically, the responses are arrayed horizontally.

Most commonly, the stem is presented a question; thus, phrases identifying the response categories take the form of a substantive response (as opposed to the extent of agreement with a statement in contrast to the degree of agreement seen in the more classic Likert item). This approach combines the (assumed) equal-interval response of the classic Likert scale with the richer response of the semantic differential. Further, the fact that the stem is interrogative rather than declarative helps avoid bias in favor of the position stated in the instrument.

e. Related issues with the Likert item

The ubiquitous use of instruments based on Likert items has perhaps caused insufficient attention to its limitations. This is particularly so when persons not familiar with research instrument design continue the use of the Likert item because they have never seen anything else. To be fair, this is an advantage as well, since respondents are also so familiar with the Likert item that training or coaching is not required. If, however, the questions raised about the Likert item here have substance, then we should be concerned that significant decisions can be reached based on research findings stemming from instruments based on Likert items. To the extent that the limitations of the Likert item described in this paper dilute the meaning and accuracy of findings, the foundation of subsequent decisions based on them is called into question.

In this paper we propose an improvement to the Likert-type item that is implemented using a GUI interface, but could be applied to paper-based instruments as well. While Likert item instruments have long been used, their use in part has been driven by the ease with which summary results can be

obtained. Again, inference of summary statistics such as mean is based on the assumption that the Liker item is both linear and can capture extremes of response, two assumptions which we question above.

It is ironic that the system we propose has been used on a paper basis in some fields, notably psychometrics and in particular in pain assessment of which the FACES scale shown above is an application specific example (Grant *et al.,* 1999). Known as the Visual Analogue Scale, it is based on a 100mm horizontal line. However, the burdensome nature of a paper-based administration of a scalar response often causes researchers to convert the results to categories, thus defeating the purpose of the scalar approach. We propose that the automated approach we propose here eliminates the need to reconvert the data to qualitative.

The work proposed here has been called for in other fields, often by arguing that Likert-item scales produce interval data to which hypotheses can be tested using an F-test

In using straightforward present-day information technology; this paper questions these two assumptions. Besides the inferred mean, computation of resulting data is automated, and there is no longer any need to rely on Likert instruments solely for their ease of computation. At the same time, the scale itself can be improved and the resulting data can be of a form amenable to richer data analysis. This will be discussed in more detail in Section 3d.

3. A PROPOSAL FOR AN AUTOMATED IMPROVEMENT TO LIKERT ITEMS

a. Rationale behind the proposed improvement

The difficulties with the Likert item described above can be addressed. Using existing technology, we can provide an improved research instrument which will employ Likert's basic methodology while permitting a high degree of granularity, leading to quantification. In turn, this will facilitate a much richer analysis of the resulting data.

It is central to our proposal that the degree of response varies over such a finely-granulated range that can be closely approximated by a quantitative variable. We propose that the response being elicited is not static but actually dynamic. The typical respondent possesses a degree of response that lies on a continuum somewhere between these two extreme statements, thus capturing a far more dynamic and disperse phenomena than a Likert item can accurately capture due to its coarse granularity. This opens a whole range of statistical tools and analyses not possible with the qualitative responses provided by a Likert item.

Researchers in various fields have sensed this limitation and taken steps to address it. For example, some argue to assign the lowest category as a zero point. Employing the assumption of scalarity, the argument results that the data interval discrete quantitative data. They then argue that standard statistical tools such as the *F*-test.

While the discrete quantitative data generated by this approach is an improvement over the qualitative data typically generated by Likert-type items, we submit that the proposal generated here is better because just as quantitative data has a richer array of analysis tools than qualitative data, so does continuous data allow for richer analysis than discrete data. (Carfio and Perla, 2007)

Thus, we propose to introduce actual linearity in place of the assumption of linear response in Likert items. We do this by introducing so many response values, and structuring the response, in such a way as to ensure a strong assumption of linearity.

As a result of what we see as weak assumptions, we propose a refinement of the Likert scale to harness the positive aspects of the Likert items with a finely-granulated *quantitative* approximation of responses. As a result of advances in computer technology, combined with the rapid advances in the availability of microcomputers, the proposed refinements of the Likert item are both practical and not so different from conventional use that it would be off-putting to respondents. The refinement generates a format that: (1) allows data to be collected as

a quasi-continuous quantitative variable; (2) is practical; (3) enhances the precision of research; and (4) takes advantage of ubiquitous computing.

Responses of rich diversity can be captured in a more finely-granulated item, and if we strive for granularity to level of precision below which the researcher has no interest in pursuing, then it is possible to treat the response as a quasi-quantitative variable. This, we propose, is correct: while no metric may be available to measure a given response, the subject's *relative* response can be measured as a scalar variable that can be expressed in such a degree of precision that it will be accepted as if it were quantitative. This is important since we often want to capture very fine and subtle movement in the population on a given point of study. Coarsely-granulated instruments would require a major shift to take place before they captured the movement, whereas more finely-granulated instrument could catch subtle movement early and quickly.

We go one step further: in addition to proposing to capture responses finely-granulated manner, we propose that instruments can be constructed with items intended to assess potentially correlative effect on related but distinct matters. The fact that the response variables can be treated quasi-quantitatively permits such a correlation analysis.

To demonstrate, we will describe the interface in some detail, both in terms of its construction and its use. We will then present an exhibit of the interface, noting an important side benefit in handling the "don't care/don't know" response. We will argue for some advantages to the proposed method.

b. A description of the survey interface

The interface begins by presenting the user with a full-screen GUI interface. The computer screen shown below represents on item. It is envisioned that the instrument will contain a series of items addressing the subject of study. An illustration of the screen is can been seen in the Appendix

The item is presented to the user as a horizontal semantic differential. The key difference is the response scale, which here is a continuous line between the semantic differentials. There are no categorical responses.

Upon encountering the item, the cursor is situated at the center point, 50.0. The numerical equivalent just below the cursor is intended for demonstration, and its use by an actual researcher is optional but not recommended.

The rules of the system require that the respondent move the cursor (an error trap will capture those who click "Submit" without moving the cursor). In this way, the respondent is mandated to express his or her response, even if his or her action is to consciously return the cursor to its middle position. This will confirm that the respondent's choice is truly the middle position, or neutrality between the two semantic differentials offered. This will be explored further below through the use of the two "opt out" options offered.

The user moves the cursor to the left or right to express the degree of his or her response of the topic being assessed. It is the *proportion* of the line that is being measured, as a surrogate of a continuous response. By definition, complete agreement with the left side semantic differential term is assigned the value of 0.0 while complete agreement with the right side semantic differential term is assigned the value of 100.0. Actual use of these values is expected to be rare. Thus, 1,001 potential responses are possible if precision is set at one decimal place. There is no reason why the scale could be more finely-granulated if necessary, nor is there any reason that the left-hand could not be defined at 100.0 and the right-hand end defined as 0.0. Any other linear values, including reversing the values of the extreme points, would suffice.

After the respondent moves the cursor to the point corresponding to his or her response and clicks "Submit", the numerical value corresponding to the proportion the line to the left is captured as the quasi-quantitative response.

c. Execution of a specific item using the interface

Having responded to the item as described above and having clicked "Submit", the score is recorded for that item and the respondent moves on to the next item. It is a matter for the researcher to decide if a response if required or whether a "skip" is permitted. However, the presence of specific "opt out" options, as discussed below, gives such valuable additional information that our recommendation would be to mandate a response.

d. Opt-out responses

In response to a singular limitation of the Likert item, the proposed method contains specific opt-out responses, as seen above. The first opt-out response, "I don't know enough to answer this question", indicates lack of sufficient knowledge to respond meaningfully. Presumably, the sample has been designed *a priori* to address that segment of the population for whom the subject of study is assumed to be appropriate. If, *a postieri¸* it is determined that a meaningful proportion of respondents lack sufficient knowledge to form a response, a valuable dimension of knowledge is discovered. Reasonable conclusions from this finding might be a need for greater education on the subject of study or more extensive advertising, promotion or other efforts to diminish the portion of the population lacking sufficient knowledge.

The second opt-out response, "I am indifferent on this question", addresses the individual who is knowledgeable about the subject of study, but who simply does not care about it. As noted above, we hold that this is a distinctly different response than a "neutral" response, as the "neutral" response (in this example, a score of 50.0) is a measure of a truly neutral response on the point covered by the item.

For these reasons, the proportion of "opt-outs" due to indifference and "opt-outs" due to lack of knowledge provide a very valuable part of the study. If the proportion of opt-outs for either reason is more than incidental, a very important finding, heretofore unavailable, has been obtained in the past.

Again, presumably, the sample *a priori* has been designed to address that segment of the population for whom the subject of study is assumed to be interested. If, *a postieri¸* it is determined that a meaningful proportion possess sufficient knowledge to form a response, but simply do not care one way or the other, yet another dimension of knowledge is discovered. This is based on the proposition that neutrality and indifference are distinct and different responses. Reasonable conclusions from this finding might be a need for more refinement in sample selection, an education effort on the importance of the subject at hand or other efforts to motivate informed but indifferent subjects to make the effort to form a response.

Note that here we do not address here issues related to no action (that is, a "skip", if that action is permitted), illiteracy, lack of fluency, or a rushed responses.

e. Benefits of use of the proposed interface

The primary benefit of the interface has been suggested earlier. The sample or survey, over a number of items, generates comprehensive, quasi-quantitative measure response on each item representing various aspects of topics of study in the population. The degree of precision is up to the researcher and in any case would be of greater granularity than can be achieved from discrete qualitative responses.

The method demonstrated above permits the far richer body of quantitative analysis to be brought to bear, including, but not limited to, the following:
1. The appropriate use of more comprehensive statistics. The inferred mean would be substituted by the *computed* sample mean, while measures of dispersion would now be available.
2. More refined measures, such a kurtosis and skewness, would also be available, as would additional measures of dispersion such as quartile or quintile divisions.
3. If a particular parameter is hypothesized *a priori* to be of a certain value, the corresponding statistic could be subjected to a t-test of difference.

4. If a particular variable was hypothesized *a priori* to have a particular distribution, χ^2 of goodness of fit would be available to test this.

5. Variances of those instrument items using the method described above would be available and would form a matrix leading potentially to common factor analysis.

6. If two or more parameters were projected *a priori* to have significant commonality or significant difference, pairwise *t*-tests to test these hypotheses would be possible.

7. If the sample or survey retains items for which the response must be qualitative (for example, gender), then one-way ANOVA analysis becomes available.

8. If the sample or survey retains two items for which the response must be qualitative, and particularly if one or both represent a specific "treatment", a two-way ANOVA analysis of a quasi-quantitative response becomes available.

The list above represents only a fraction of the analyses that become available if the belief or affect can be operationalized as a quasi-qualitative variable, and the list above does not exclude even more specific analyses.

5. RELEVANCE TO IS EDUCATION

The question may be fairly raised about the relevance to the work described here to education in Information Systems and related fields. We argue that is does, but limit ourselves to two arguments here:

1. Many information systems include the means to capture user reaction to use of the system. This is particularly prevalent among websites. Today, these instruments rely on Likert-items or some derivate. Just as we argue that research other fields are limited by the Likert-items, so too are attempts at user response in Information Systems.

2. Many areas of business are facing increased competition both in competitors who formerly could not compete and in the scope of products available. A good example is the field of financial services after the passage of the Gramm-Leach-Bliley Act of 1999, which effective removed barriers between types of finan-

cial services. In such a market, constantly tracking customer sentiment is crucial. Survey such as that described here is used a great deal, often as pop-ups at the end of a session. Thus, it is reasonable to assume that IS graduates will be working on the very kinds of systems described here.

6. CONCLUSION

The use of the Likert items and its derivatives generate result data that is qualitative and ordinal, and can only be presumed to be linear. Moreover the categorical responses provided by the Likert-tem are presumed to cover the full range of possible responses. We contend that these presumptions are questionable.

In response, a GUI-based item presentation has been devised, presenting several advantages while retaining a Likert item derivative that: (1) has a question instead of a declarative positive statement as its stem; (2) uses a horizontal presentation; (3) uses semantic differentials as anchors yet (4) preserves sufficient flavor of conventional uses of Likert items to provide users with a comfortable and familiar item format. This conforms with many common uses of the Likert item today.

Our major innovation is the continuous scale presented to the user in order to measure their response. Its major advantage is data from each item that is quasi-qualitative and can be generated to any reasonable degree of precision. Other advantages include: (1) the presentation of specific opt-out responses which are not subsumed under the middle or neutral response; (2) that user is mandated to move the cursor to a point on the continuous scale reflecting the degree of their response (even if their response causes them to return the cursor to the exact center point); (3) by measuring the proportion of the cursor's ending point on the continuous scale, the resulting data can be viewed as quasi-quantitative data; (4) explicitly establishes linearity of response; (5) allows users to express the most extreme responses**;** and (6) generates results of fine granularity.

We submit that use of this form of item will generate data that can be seen as sufficiently quantitative to permit exploitation of

the far richer repertoire of analysis tools possible only with quantitative data.
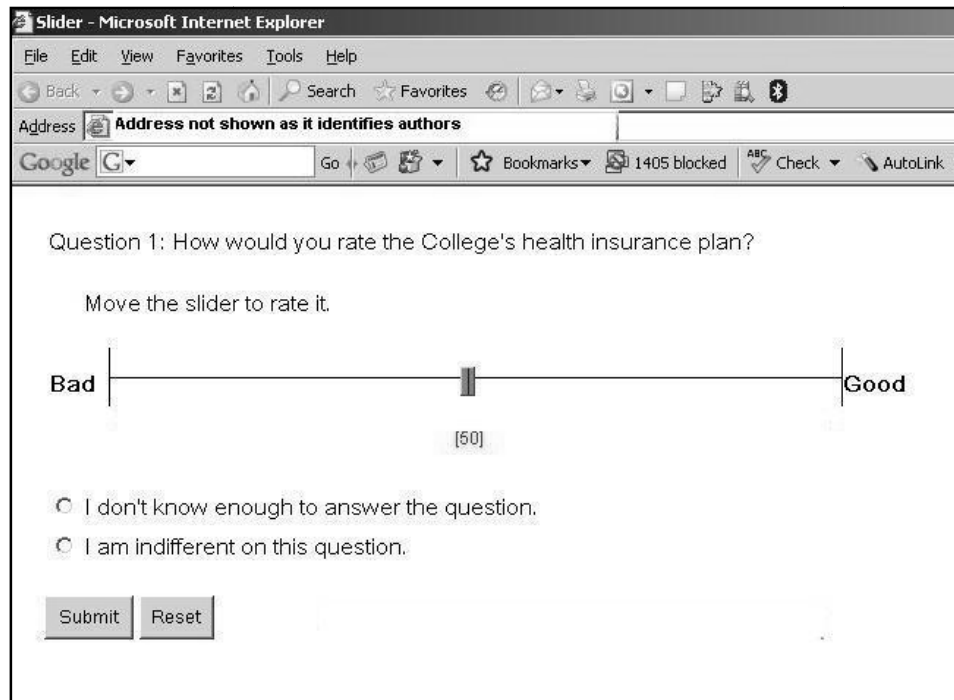
7. ACKNOWLEDGEMENTS

The authors would like to thank Prof. Tom Madden, now of the University of South Carolina, in whose class many years ago the germ of the ideas expressed here arose in class discussion.

8. REFERENCES

Aaker, K and G. Day (2003) , Marketing Research (6th ed.). Wiley, New York.

Babbie, E (2003), Practice of Social Research (10th ed.). Thomas Wadsworth.

Bogardus, E. (1926), " Social Distance in the City", Proceedings and Publications of the American Sociological Society , 20, 40-46

Dillon, W. R., Madden, T. J., & Firtle, N. H. (1993). Essentials of Marketing Research. Irwin.

Grant, S. E. (1999). A comparison of the reproducibility and the sensitive to change of visual analogue scales, Borg scales and Likert scales. Chest 116:5 (November 1999), 1209-1217.

Guttman, L. (1950). "The basis for scalogram analysis" In E. A. Stouffer, Measurement and Prediction. Wiley, New York

Hildebrand, D. K., & Ott, L. (1991). Statistical Thinking for Managers (3rd ed.). Duxbury Press.

Hockenberry, M. (2004). Wong's Essentials of Pediatric Nursing (7th ed.). Mosby, St. Louis.

Likert, R. (1932). "A Technique for the Measurement of Attitudes". Archives of Psychology , 140, 5-55.

Likert, R. (1948). "Public Opinion Polls: Why Do They Fail?" Scientific American , (117:6) 7-11.

Marshall, G. (1998). "Thurstone scale. In A Dictionary of Sociology" in Osgood, C. E., G. Suci, and P Tannenbaum The Measurement of Meaning. University of Illinois Press, Urbana

Snider, J. G., & C.E. Osgood (1969). Semantic Differential Technique: A Sourcebook. Aldine. Chicago

Thurstone, L. L. (1927). "A Method of Paired Comparison for Social Values" Journal of Abnormal and Social Psychology, 21, 384-400.

van Schuur, W. H. (2003). "Mokken Scale Analysis: Between the Guttman Scale and the Parametric Item Response Theory". Political Analysis , 11, pp. 139-163.

9. APPENDIX FOR ILLUSTRATIONS

Figure 3: Interface for proposed improvement