

When “Strongly Disagree” Doesn’t Mean Strongly Disagree

Donald Shearman

(University of Western Sydney, Sydney)

Peter Petocz

(Macquarie University, Sydney)

Abstract

Student evaluations of faculty and courses have been used as a method of quality control in higher education for almost 100 years. Analysis of the data generated by these surveys has been the focus of considerable research for at least the past 20 years, with the bulk of this analysis using techniques which assume that the survey data, usually Likert item responses, are quantitative variables. By utilising correspondence analysis, a statistical tool which makes no such assumptions on the variables, we have produced results from one set of data comprising 2749 surveys which suggest that the assumption that Likert item responses can be treated as quantitative variables could be challenged. Through this we provide new insight about student evaluations and question these assumptions at least for one data set.

Background

Student evaluations of instruction and of perceived quality of units have been a part of the higher education process for at least 90 years (Algozzine et al., 2004) in post-secondary institutions around the world. These evaluations usually take the form of a series of items which students are asked to rate using some form of Likert type response (Strongly Agree – Agree – Neutral – Disagree – Strongly Disagree) often with an open response area for general comments. Data from such surveys of instructors has been used as both a formative assessment of teaching – to measure the effectiveness of faculty teaching ability and to assist with faculty development (Nowell, Gale, & Handley, 2010) – and as a summative assessment for promotion, course assignment and tenure (Agbetsiafa, 2010), whilst data from evaluation of course quality may be used for evaluation and modification of curricula. Basic reporting of survey data usually involves assigning each response category a numeric value (often 1 – 5) and calculating averages (means) and standard deviations for each survey item using these values. This process has shown little change throughout the history of such evaluations.

In analysing evaluation data to identify underlying concepts much use has been made of statistical methods such the use of mean (average) scores and analysis of variance (ANOVA) (Simione, Cadden, & Mattie, 2008), factor analysis (Simon & Soliman, 2003; Sohail & Shaikh, 2004), linear regression (Denson, Loveday, & Dalton, 2010) and structural equation modelling (Toral, Barrero, Martinez-Torres, Gallardo, & Duran, 2009). These methods all assume that the data being analysed are quantitative, an assumption which cannot always be guaranteed.

Data, from a statistical perspective, can be classified broadly into two main categories; quantitative and qualitative. Quantitative data according to Keller are “real numbers such as heights, weights, incomes, and distances” (Keller, 2008) while qualitative data are categories such as the response to questions about marital status. The responses to student evaluation items, while clearly not

quantitative, have more structure than that of a qualitative variable since they have a natural order (Strongly Agree > Agree > Neutral > Disagree > Strongly Disagree). Such variables are usually classified as ordinal. Data types are important in determining appropriate calculations which can reasonably be performed on the data and hence the appropriate analysis. In particular calculations which involve the addition of data values should be restricted to quantitative variables where the difference between two values has a consistent meaning such as height, weight, exam marks, etc.

Much discussion in elementary statistics focuses on the use of appropriate statistics and statistical methods to describe data of different types (Keller, 2008; Moore, McCabe, & Craig, 2009) and cautions against the use of quantitative techniques for ordinal variables. In addition, several writers have particularly identified problems with treating individual Likert items as quantitative variables (Allen & Seaman, 2007; Carifio & Perla, 2007) although the combining of several Likert items to construct a Likert scale may allow the combined scale to be treated as quantitative. The issues surround the fact that although ordinal variables are often represented by numbers, the numbers convey only a rank or order and differences between these numbers generally do not suggest an absolute difference in quantities. For example in a Likert item a difference between Strongly Agree (5) and Agree (4) is 1, but this value does not measure some absolute difference in opinion.

Statistical analysis of ordinal variables tends to be overlooked in part because such variables do not fall into the two main groups but have some characteristics of each. Such techniques as do exist are relatively recent developments and are not well understood by many (Agresti, 2010). Consequently qualitative methods are frequently applied.

Methods

Our data set consists of 2749 responses to a student evaluation survey comprising 20 items (10 relating to evaluation of instructors and 10 relating to evaluation of unit) completed by students from an institute of higher education in NSW, Australia in November 2011. All items in the survey were answered using a five point Likert scale response varying from Strongly Agree to Strongly Disagree. The responses covered 60 individual units of study with the maximum number of responses for an individual unit being 147, and the minimum number of responses for a unit being four. Items in the survey related to a number of issues identified by the institution for which they required feedback as displayed in Table 1. Item identifiers (U1 – U10 and T1 – T10) are used in the correspondence maps which follow.

Table 1:
Student Survey Items

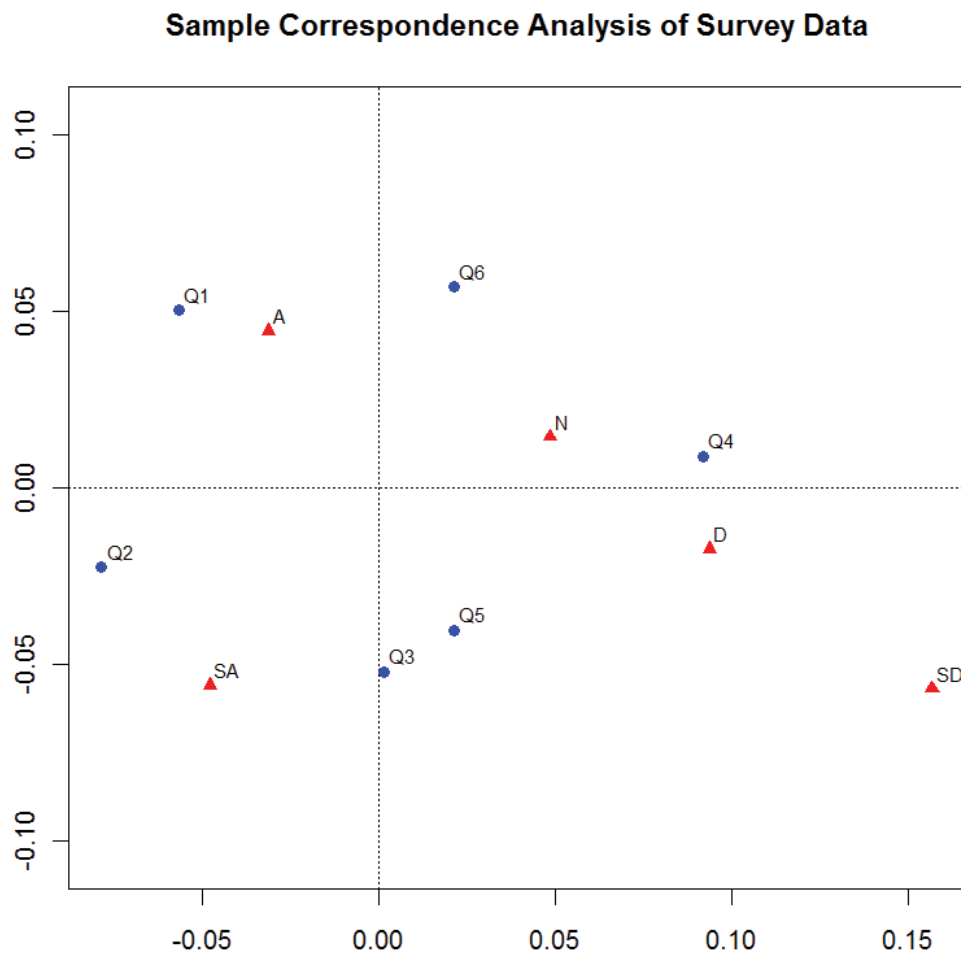
Unit Items		Instructor Items	
U1	Clarity of aims and objectives as stated in documentation for the unit	T1	Instructor's competence in relating subject material to real life
U2	Clarity of statement of assessment criteria	T2	Encouragement of class discussion
U3	Perceived appropriateness of assessment tasks	T3	Clear communication of unit requirements
U4	Depth of cover of material compared to previous learning	T4	Motivation to make students work independently
U5	Appropriateness of topic order;	T5	Organisation and punctuality
U6	Usefulness of learning activities	T6	Constructiveness of feedback
U7	Usefulness of teaching materials	T7	Structure of classroom activities
U8	Encouragement of critical thinking	T8	Respect of different opinions

U9	Timeliness of feedback	T9	Availability and helpfulness
U10	Pace of teaching	T10	Enthusiasm for teaching

In order to avoid the previously discussed problems associated with the analysis of ordinal data we opted to use a statistical method known as correspondence analysis which can be used to identify and display relationships within qualitative data. As Greenacre notes in the preface to his book (Greenacre, 2007):

“Correspondence analysis is a statistical technique which is useful to all [...] who collect categorical data, for example data collected in social surveys. The method is particularly helpful in analysing cross tabular data in the form of numerical frequencies, and results in an elegant but simple graphical display which permits more rapid interpretation and understanding of the data.”

The process can be seen as being akin to the familiar technique of the scatterplot as used to display quantitative data, but centres on creating a graph of the profiles (frequencies of responses for each category of answer) of each survey item. Because these profiles require more than two dimensions to graph, correspondence analysis uses a process similar to factor analysis to reduce the graph to the dimensions which contain the most information. The process works on a quantity known as the chi-squared measure of the data which compares each profile with “expected” or overall profile constructed by adding all profiles together. A typical example of such a map is shown in Figure 1. In this map the location of the responses, Strongly Agree – Strongly Disagree, are represented by triangles which can be seen to follow an arch from Strongly Agree at the left to Strongly Disagree at the right. The arch structure, known as the Horseshoe Effect, is typical of an ordinal response variable such as that found in survey responses (Weller & Romney, 1990). The location of the dots representing item profiles indicates the relative position of each compared to the average profile for all items, the closer a profile is to a response triangle, the more important that response is in the item’s profile.



Fig

ure 1:

Typical correspondence map for Likert item survey data

Although the primary use of correspondence analysis is as a graphical tool, it has been shown (Greenacre, 2007) that where a variable can be interpreted as being of an ordinal type the coordinates for the categories on the principal axis can be used to construct a scale for the categories which may be used to assign quantitative values to them. These values, known as an optimal scaling, then give a measure of the relative distances between the ordered categories and may be used to compute means and standard deviations for the various survey items which better reflect the difference between categories than the arbitrary assignment of the numbers one to five to the categories. In addition the total variability for the data can be measured in terms of the chi-squared measure and percentages of this variability can be associated with each dimension of the correspondence map.

Results

The surveys used by the institution we were studying included both unit and instructor items on the same survey instrument since at this institution each unit is generally taught by a single instructor and

it was felt that a single instrument would be easier to administer. Surveys were distributed in class in paper form and were returned to an administrative assistant for transcription of results to a spreadsheet.

As a first point of analysis, all survey questions were subjected to correspondence analysis together. The resultant correspondence map can be seen in Figure 2. The unexpected shape for the response categories in the map together with the aggregation of the instructor-focused issues around the Strongly Agree category led us to suspect that responses for the two classes of items may be fundamentally different and hence suggested that analysis of the unit and instructor items should be carried out separately. In terms of optimal scaling values the correspondence analysis suggests a value of 1.37 for Strongly Agree, -0.39 for Agree, -1.21 for Neutral, -1.55 for Disagree and -0.40 for Strongly Disagree. The two dimensions of the map account for about 94% of the total variability of the data.

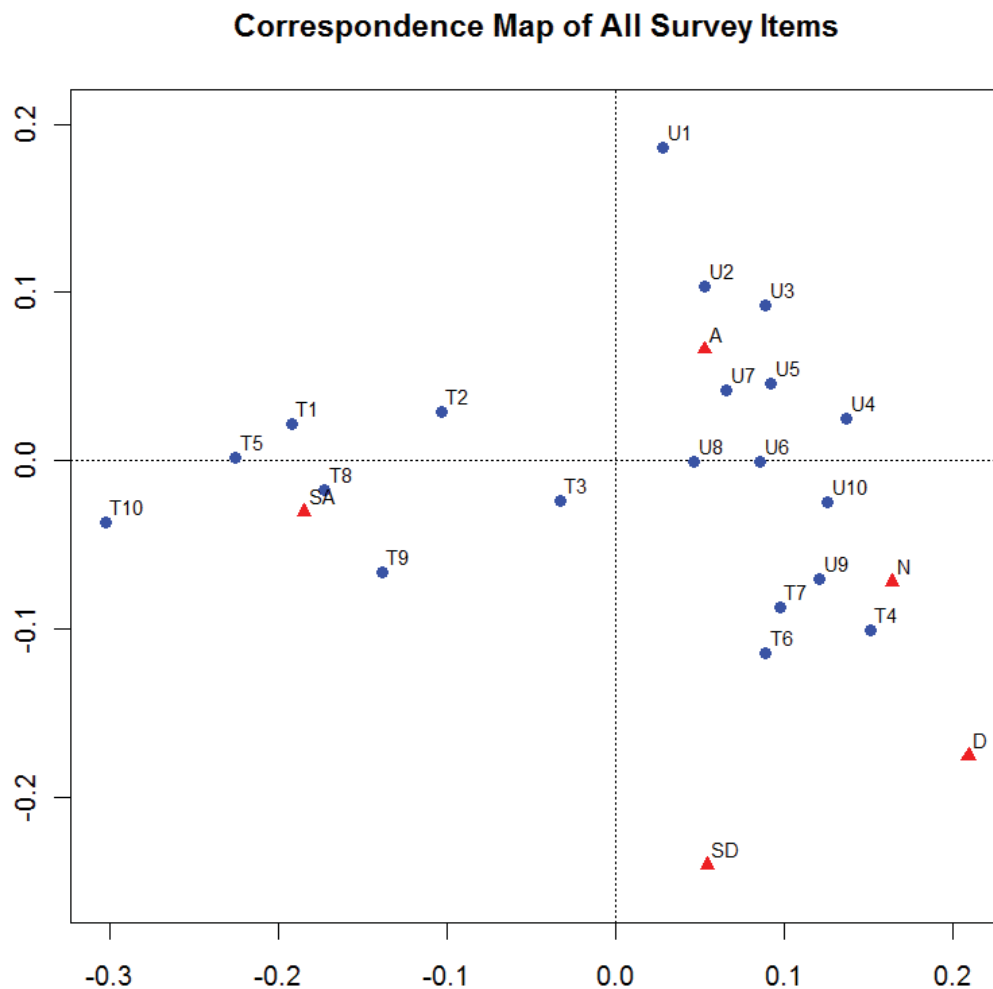


Figure 2:
Correspondence map for all survey items

A correspondence analysis of the instructor items resulted in the correspondence map in Figure 3. While this has resulted in a separation between the Agree and Strongly Disagree categories, the placement of the Strongly Disagree category on the map is still not in a location which would be expected for an ordinal variable, nor is there evidence of the Horseshoe Effect which would be expected if the responses were ordinal. The optimal scaling value obtained from this analysis are Strongly Agree 1.09, Agree -0.17, Neutral -1.55, Disagree -2.20 and Strongly Disagree -1.00 with 96% of the variability of the data shown in the map.

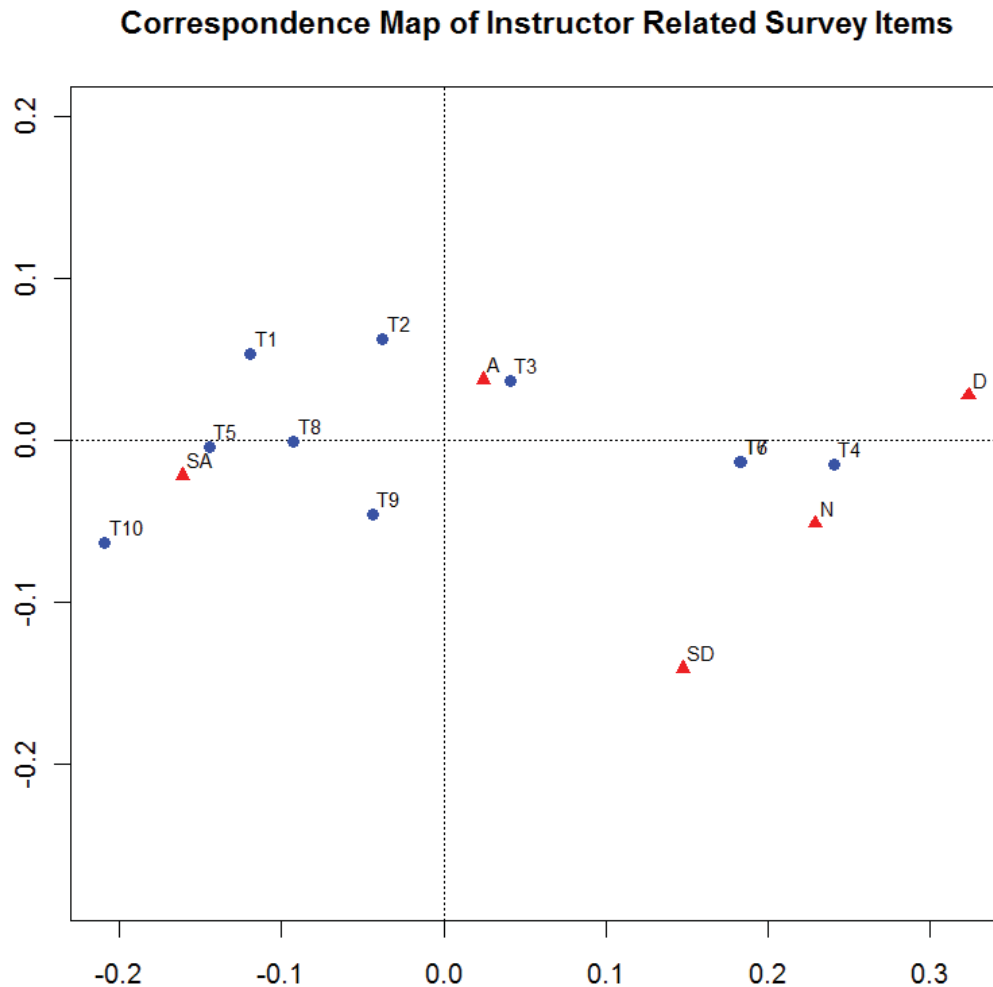


Figure 3:

Correspondence map for instructor related items

A similar analysis of the unit items resulted in the correspondence map displayed in Figure 4. In this instance the lower three categories appear in their expected order, but the Agree and Strongly Agree categories have reversed positions from what would be expected. Optimal scaling values for the categories are Strongly Agree 0.13, Agree 0.65, Neutral -1.15, Disagree -3.13 and Strongly Disagree -3.76. A total of 92% of the variation in the data is contained in the map.

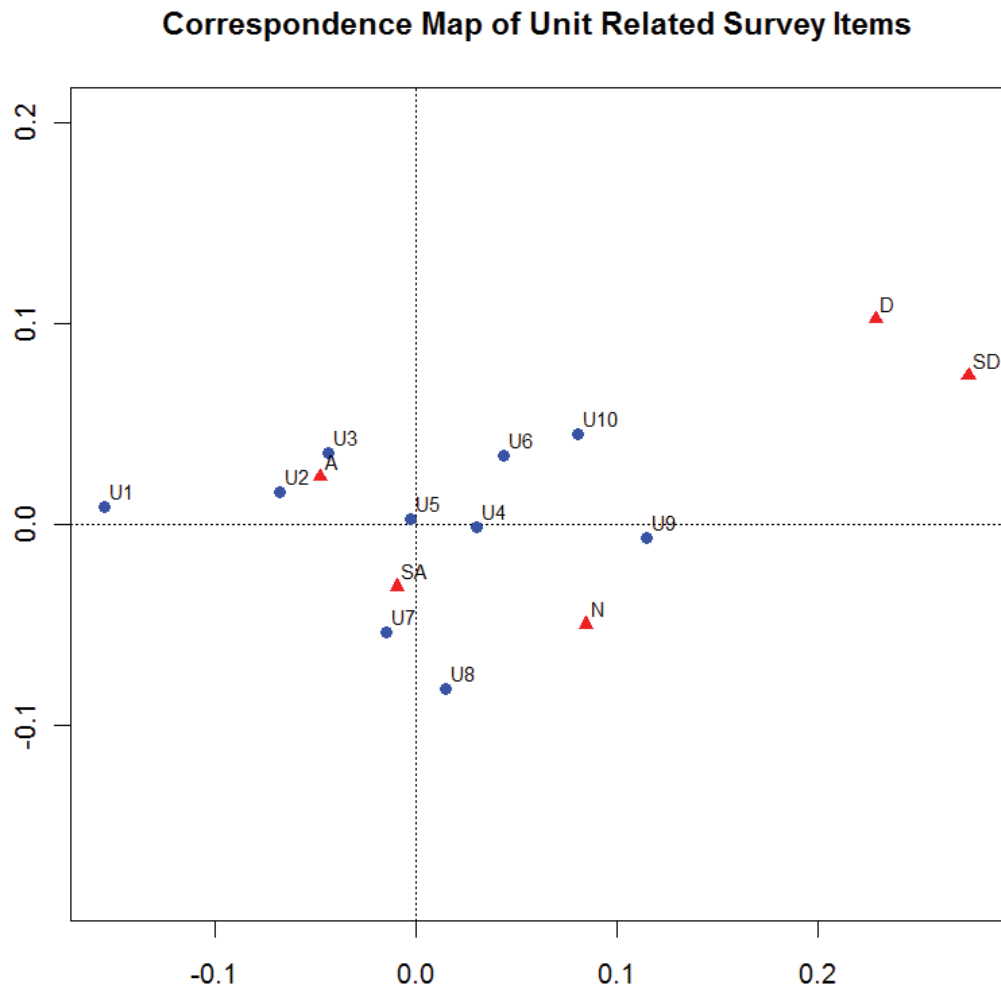


Figure 4:

Correspondence map for unit related items

Discussion

The initial analysis of all survey items suggests that students respond to items about their instructors more favourably than those for the unit as evidenced by the aggregation of the instructor items around the Strongly Agree category marker, indicating that these items have a higher proportion of Strongly Agree responses than other items. Because correspondence analysis, like factor analysis, identifies dimensions or factors in the data in terms of decreasing strength, the correspondence map for the combined questions suggests that the difference in response to the unit and instructor items may be a stronger factor than the differences created by the response categories. Thus the first (horizontal) dimension of the correspondence map shows the presence or absence of Strongly Agree, the difference between Instructor and Unit item responses, while the second (vertical) dimension of the

map separates the other response categories. This is also supported by the fact that the other response categories appear in their usual order from top to bottom of the map.

While this explanation may describe the unexpected shape of the correspondence map for the questions overall, it does not help when investigating the classes of items separately. The inclusion of several class groups with low numbers of responses, almost all being Strongly Agree and Agree, was also considered as a possible reason for the unusual shape of the correspondence map; however, excluding these groups from the data and reanalysing made no significant difference to the results. Another possibility is that the results reflect a randomness of responses due to disinterest from the students completing the surveys or to survey fatigue caused by students being required to complete surveys for several units within a short space of time.

While the effect described in this paper has been observed in a previous survey from the same institution, preliminary studies with similar surveys from another institute of higher education in NSW, Australia have not shown the effects described here – these results were used to produce the correspondence map shown in Figure 1.

Conclusion

The results of our analysis of data from student evaluations of instructor and unit of study from one institute of higher education in NSW suggest that the assumptions that responses to such surveys can be treated as quantitative variables and subjected to analysis as such requires careful consideration. In our data set, responses which were expected to follow a simple ordered pattern from Strongly Agree to Strongly Disagree failed to do so when subjected to correspondence analysis. While no compelling explanation for this failure to behave in the expected manner can be found from the data, survey fatigue, student apathy and the inclusion of spurious results from a number of units with very small response rates may be contributing factors. In any event further investigation of the effect is highly recommended.

References

- Agbetsiafa, D. (2010). Evaluating Effective Teaching In College Level Economics Using Student Ratings Of Instruction: A Factor Analytic Approach. *Journal of College Teaching & Learning*, 7(5), 57-65.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Hoboken: John Wiley & Sons.
- Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., . . . Spooner, F. (2004). Student Evaluation of College Teaching: A Practice in Search of Principles. *College Teaching*, 52(4), 134-141.
- Allen, E. I., & Seaman, C. A. (2007). Likert Scales and Data Analyses. *Quality Progress*, 40(7), 64-65.
- Carifio, J., & Perla, R. J. (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.
- Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: what predicts satisfaction? *Higher Education Research & Development*, 29(4), 339-356.
- Greenacre, M. (2007). *Correspondence Analysis in Practice*. London: Chapman & Hall/CRC.
- Keller, G. (2008). *Statistics for Management and Economics*. Mason: South-Western Cengage Learning.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2009). *Introduction to the Practice of Statistics*. New York: W. H. freeman and Company.

- Nowell, C., Gale, L. R., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, 35(4), 463-475.
- Simione, K., Cadden, D., & Mattie, A. (2008). Standard Of Measurement For Student Evaluation Instruments. *Journal of College Teaching & Learning*, 5(12), 45-58.
- Simon, J. C., & Soliman, K. S. (2003). An Alternative method to measure MIS faculty teaching performance. *International Journal of Educational Management*, 17(5), 195-199.
- Sohail, M. S., & Shaikh, N. M. (2004). Quest for excellence in business education: a study of student impressions of service quality. *International Journal of Educational Management*, 18(1), 58-65.
- Toral, S. L., Barrero, F., Martinez-Torres, M. R., Gallardo, S., & Duran, M. J. (2009). Modeling Learner Satisfaction in an Electronic Instrumentation and Measurement Course Using Structural Equation Models. *IEEE Transactions on Education*, 52(1), 190-199.
- Weller, S. C., & Romney, A. K. (1990). *Metric Scaling Correspondence Analysis*. Newbury Park: Sage University.