Spring 6-10-2017

A THRESHOLD FOR A Q-SORTING METHODOLOGY FOR COMPUTER-ADAPTIVE SURVEYS

Sahar Sabbaghan
University of Auckland, s.sabbaghan@auck-land.ac.nz

Lesley Gardner
University of Auckland, l.gardner@auck-land.ac.nz

Cecil Eng Huang Chua
University of Auckland, aeh.chua@auckland.ac.nz

A THRESHOLD FOR A Q-SORTING METHODOLOGY FOR COMPUTER-ADAPTIVE SURVEYS

Research in Progress

Sahar Sabbaghan, Department of ISOM, University of Auckland, Auckland, NZ, s.sabbaghan@auckland.ac.nz

Lesley Gardner, Department of ISOM, University of Auckland, Auckland, NZ, l.gardner@auckland.ac.nz

Cecil Eng Huang Chua, Department of ISOM, University of Auckland, Auckland, NZ, aeh.chua@auckland.ac.nz

Abstract

Computer-Adaptive Surveys (CAS) are multi-dimensional instruments where questions asked of respondents depend on the previous questions asked. Due to the complexity of CAS, little work has been done on developing methods for validating their content and construct validity. We have created a new q-sorting technique where the hierarchies that independent raters develop are transformed into a quantitative form, and that quantitative form is tested to determine the inter-rater reliability of the individual branches in the hierarchy. The hierarchies are then successively transformed to test if they branch in the same way. The objective of this paper is to identify suitable measures and a "good enough" threshold for demonstrating the similarity of two CAS trees. To find suitable measures, we perform a set of bootstrap simulations to measure how various statistics change as a hypothetical CAS deviates from a "true" version. We find that the 3 measures of association, Goodman and Kruskal's Lambda, Cohen's Kappa, and Goodman and Kruskal's Gamma together provide information useful for assessing construct validity in CAS. In future work we are interested in both finding a "good enough" threshold(s) for assessing the overall similarity between tree hierarchies and diagnosing causes of disagreements between the tree hierarchies.

Keywords: Computer-adaptive, survey, construct validity, q-sorting, threshold

1 Introduction

A new kind of survey in business research is Computer-Adaptive Surveys (CAS). Unlike in a traditional survey, where every question is asked (Hayes, 1992), in a CAS, the previous questions determine the next questions asked. CAS differs from traditional surveys in several ways. First, the items in CAS are arranged in a hierarchy, whereas traditional methods assume a "flat" set of items. Second, respondents legitimately only fill in some questionnaire items, hence unfilled questions cannot be treated as non-responses.

A typical CAS item bank can contain hundreds of items. In addition constructs are mapped together in a hierarchy, with constructs concerning higher level concepts linking to constructs with greater precision. To illustrate the process, consider the CAS we have developed, which is designed to elicit the problems customers have with cafés.

CAS uses an adaptive version of branching for respondents to move from one set of items to another set according to pre-defined criteria. To illustrate, see Figure 1. If food is the area that the customer is least satisfied with, CAS then retrieves questions about the quality of the food (i.e., preparation, portion, menu choice). If the customer is least satisfied with preparation, CAS then retrieves questions about how the food was cooked, taste, special needs, options, and availability. CAS does not retrieve further questions on constructs the respondent rated satisfactorily. As the customer continues to answer questions, CAS navigates deeper down the tree until it finally identifies that the food item is too salty. As the goal of CAS is to measure individuals' perceptions, CAS typically identifies one or a few narrowly defined constructs that respondents as a whole have the greatest or least affiliation to. In the CAS we developed, each customer goes through the survey and selects the constructs which they are least happy with. Hence, CAS provides paths which start from top level constructs to bottom constructs. If customers agree that particular constructs are the cause of their dissatisfaction, then certain path(s) would be chosen more.

These differences in structure mean traditional methods for validating surveys do not work for CAS.

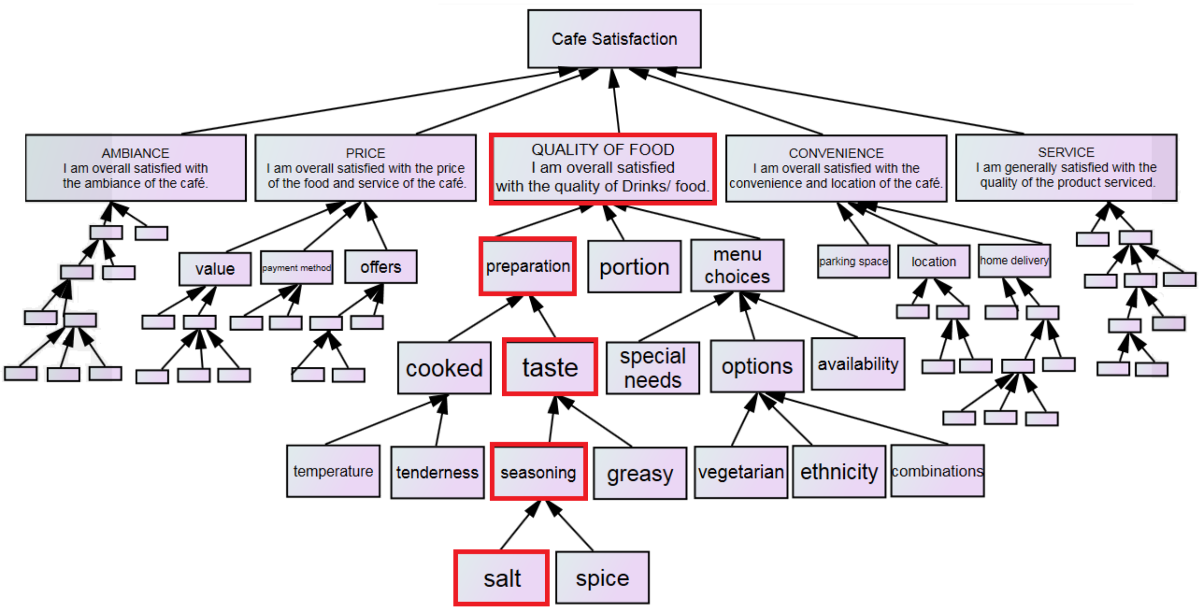


Figure 1. How CAS works for café satisfaction

2 Review of Literature

Traditional IS survey research has been useful for developing causal relationships between constructs, but has been poor at "unpacking" constructs to develop an in depth understanding for the questions of why (Pinsonneault and Kraemer, 1993). Bagozzi (2007) highlights that research on the technology acceptance model has demonstrated the relationship between perceived usefulness, ease of use, and intention to use, but cannot articulate why this relationship holds. Similarly, if one does a customer satisfaction survey, one often wants to know not only that something is wrong, but precisely why it is wrong (Fundin and Elg, 2010; Sampson, 1998; Wisner and Corney, 2001). Surveys must strike a balance between trying to understand every respondent's individual views and opinions, and not exhausting the respondent with too many questions (Haschke, Abberger, Wirtz, Bengel, and Baumeister, 2013; Hayes, 1992; Sinclair, Clark, and Dillman, 1993). If the survey is too long, respondents will suffer from a fatigue effect, and not answer questions properly (Berdie, 1989; Deutskens, de Ruyter, Wetzels, and Oosterveld, 2004). If the survey is too short, it will not provide enough information to adequately capture respondent sentiment (Hayes, 1992).

Computer-Adaptive Surveys (CAS) are multi-dimensional instruments where questions asked of respondents depend on the previous questions asked. CAS offers certain advantages over traditional surveys. Its principal advantage is that it allows the survey developer to include a large number of questions. The only questions the respondent answers are the ones most salient to the issue being addressed. In contrast, if the same number of questions were asked on a traditional survey, the respondent is likely to encounter fatigue and quit before providing critical information (Galesic and Bosnjak, 2009; Groves, 2006; Groves et al. 2004; Heerwegh and Loosveldt, 2006; Porter, Whitcomb and Weitzer, 2004).

A CAS can be thought of as a hybrid of a traditional perception survey and a Computer-Adaptive Test. Computer-Adaptive Tests are designed to efficiently assess and evaluate a respondent's ability or performance by administering questions dynamically based on answers to the questions the participant answered previously (Thompson and Weiss, 2011). CAT uses Item response theory (IRT) to do this. IRT models describe in probabilistic terms the relationship between a person's response to a question and his or her level of the 'latent variable' being measured by the scale (Reeve & Fayers, 2005). There are two broad kinds of IRT, dichotomous and polytomous. Dichotomous IRT models require each item to be scored either correct or incorrect. For example, the Graduate Management Admission Test (GMAT) asks the respondent to answer language and mathematical questions in increasing order of difficulty (Stricker, Wilder, and Bridgeman, 2006). The next question asked of a respondent depends on whether the previous questions were answered correctly. In contract, in polytomous IRT models, an item can have two or more response categories. For example, a 5-point Likert type scale (Reeve & Fayers, 2005). Polytomous models are more common in surveys, such as measuring political knowledge (Montgomery & Cutler, 2013), or measuring workplace bullying (Ma, Chien, Wang, Li, and Yui, 2014).

CAS is comparable to CAT, but their goals are very different. CAS aims to identify the child constructs most salient to a particular group of respondents (e.g., which things did you like the least or most), while CAT assesses an ability or performance (Hol, Vorst, and Mellenbergh, 2008; Merrell and Tymms, 2007). Traditionally, the goal of the typical CAT is to produce a score evaluating ability or performance on a single or few constructs. In contrast, the goal of CAS is typically to identify which of many child constructs are perceived by the respondent as most relevant to them.

These dissimilarities in goals result in structural incongruities between the two kinds of surveys. CAT relies on Item Response Theory (IRT) functions to determine further questions to ask respondents (Embretson and Reise, 2000; Lord, 1980; Thompson and Weiss, 2011; Thorpe and Favia, 2012). CAS, in contrast, uses an adaptive version of branching to arrange the questions. Lowest or highest scores on a set of questions causes the system to retrieve related, but more precise questions. The question structures are also different. As an example, on the GMAT, which is based on IRT, the "correct" answer adds a point to the score, while an incorrect one deducts from 0.25 to 0.20 from one's score. In contrast, items in CAS are more akin to those on traditional psychometric instruments that are designed to "load" on a construct.

3 Building and Assessing a Threshold for Q-Sorting for CAS Items

Traditionally, construct validity in surveys is performed using two methods. The first is factor loading, "which is the correlation between the original variables and the factors" (Hair et al., 1998). However, this method is problematic for use with CAS (Diamantopoulos and Winklhofer, 2001; Jarvis, MacKenzie, and Podsakoff, 2003; Mackenzie, Podsakoff, and Podsakoff, 2011). In the CAS parent-child relationship, the parent will correlate highly with at least one of the children, but is unlikely to correlate with all. For example, if a respondent answers they are dissatisfied with the food quality, the respondent might be unhappy about the way the food was prepared but be satisfied with the portion size.

The second method to assess construct validity is Q-sorting (Straub and Gefen, 2004). In the typical q-sort, independent raters are provided with a set of cards, where each card contains a single questionnaire item. Raters are then instructed to place the cards into groups, where the groups correspond to the constructs (Block, 1961). In some cases, the number of groups is pre-assigned (Segars and Grover, 1998). In others, grouping is left to the rater (McKeown and Thomas, 1988). Q- sorting may be one of the best methods to assess content and construct validity for constructs with parent-child relationships (Petter, Straub, and Rai, 2007).

In earlier research, we developed a methodology for performing q-sorts on CAS (Sabbaghan, Gardner, & Chua, 2016). In that methodology, we provided a framework where the hierarchies that independent raters develop are transformed into a quantitative form, and that quantitative form is tested to determine the inter-rater reliability of the individual branches in the hierarchy. The hierarchies are then successively transformed to test if they branch in the same way.

However, that article failed to identify suitable measures and "good enough" thresholds for demonstrating the similarity of two CAS trees. Most statistics-based research in information systems relies on these thresholds (Boudreau, Gefen, and Straub, 2001). For example, we regularly consider a p-value under 0.05 to be "good enough." The rest of this study outlines our strategy for determining this threshold.

To develop thresholds, we generate a hypothetical "perfect" hierarchy for CAS. We then systematically modify this and calculate how various measures of "sameness" are impacted by these modifications. This is similar to techniques others have employed to develop thresholds for other statistical techniques (Cheung and Rensvold, 2002; Hu and Bentler, 1998).

3.1 Swaps and Movements

There are two broad ways two hierarchies can differ, which we call a swap and a movement. A swap is where two constructs switch locations. A movement is where a construct moves and becomes a child of a new parent. We illustrate the various swaps and movements using the generic hierarchy Figure 2 as a basis.

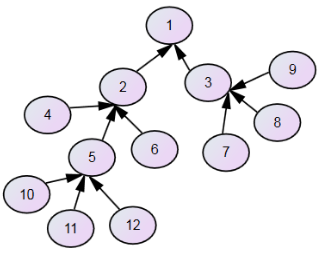
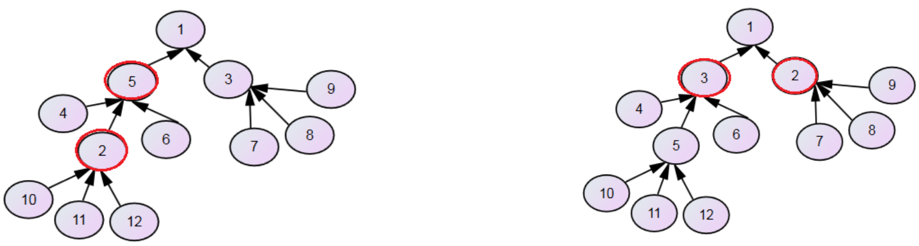


Figure 2. Example of hierarchy

Hierarchy Swap: a hierarchy swap is where a construct is swapped with a relative. A relative is the direct or indirect child or parent of the first construct. As an example in Figure 2, for construct 2, the relatives are constructs: 1, 4,5,6,10,11, 12. Figure 3(a) presents a hierarchy swap between constructs 5 and 2.

Level Swap: A level swap is where a construct is swapped with another construct of the same level. As an example in Figure 3(b), constructs 2 and 3 are on the same level and can be swapped.

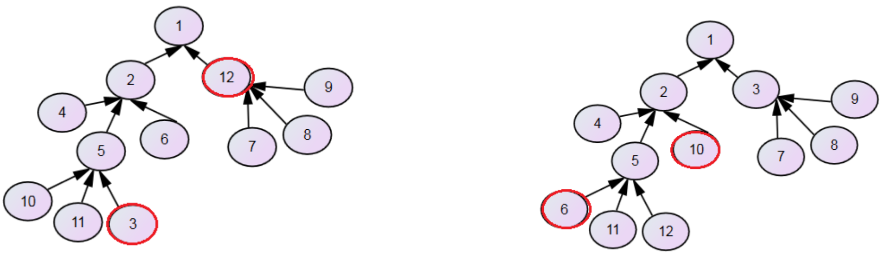


(a) hierarchy swap (b) level swap

Figure 3. Example of hierarchy swap and level swap

Diagonal swap: A diagonal swap is where a construct is identified and swapped with another. This second construct should not be in the same level as the first and the two constructs must not share the same direct parent. As an example, as illustrated in Figure 4(a), if construct 3 is the first construct, it can be swapped with either constructs 4, 5, 6, 10, 11, or 12.

Random swap: A random swap is where 2 constructs are identified and swapped. The second construct can be located at any position of the hierarchy. As an example, as presented in Figure 4(b), constructs 10 and 6 can be swapped.

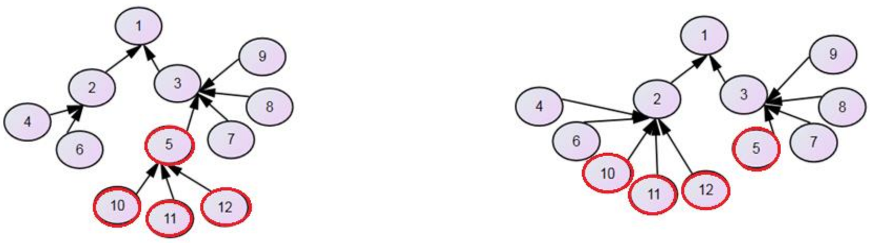


(a) diagonal swap (b) random swap

Figure 4. Example of diagonal and random swap

Movement with children: A movement with children is where a construct and all its children (if any) is selected and moved to another construct as the child of the second construct. Note that at all times a parent construct must have at least 2 children- otherwise the CAS will have no "branching" choice. As an example, as shown in Figure 5(a), construct 5 can move to under construct 3. Therefore, construct 3 is the new parent of construct 5.

Movement without children: A movement without children is where a construct without its children (if any) is selected and moved to another construct as the child of the second construct. The children of the first construct will go to the direct parent of the first construct. Note that at all times a parent construct must have at least 2 children. As an example consider Figure 5(b), where construct 5 has moved to under construct 3, however, constructs, 10, 11, and 12 have not moved with their parent and now are the children of construct 2.



(a) movement swap with children (b) movement swap without children

Figure 5. Example of movement swap with and without children

4 Data Collection

In our analysis, we have created a hierarchy with 121 constructs where each parent has exactly three children. We then performed each of the above movements and swaps on a random node one hundred times, recording each change from the perfect tree. We also have performed tests where random movements and swaps are performed 100 times. We then measure the correspondence of each tree against the "perfect" tree using 3 measures of association, Goodman and Kruskal's Lambda , Cohen's Kappa, and Goodman and Kruskal's Gamma (Goodman and Kruskal, 1954). We chose Lambda (λ) because it is the strength of association in a contingency table (Everitt, 1992) and has a meaning akin to r in a regression (Anderson and Gerbing, 1988). In addition, unlike the chi-square derivatives, the lambda coefficients provide a measure of the strength of relationship between two nominal variables and have proportional reduction in error interpretations (Goodman and Kruskal, 1954).

We chose Kappa, because it is widely used as a measure of association for contingency tables (Hambleton and Zaal, 2013; Rudick, Yam, and Simms, 2013; Sengupta and Te'eni, 1993; You, Xia, Liu, and Liu, 2012). According to Cohen (Cohen, 1968), Kappa is the observed proportion of agreement between the assigners after chance agreement is removed from consideration.

We chose Gamma (γ) because it is explicitly designed for data with ordinal values, and hierarchies are ordered data structures. Goodman and Kruskal interpret Gamma (γ) as "how much more probable it is to get like than unlike orders in the two classifications, when two individuals are chosen at random from the population" (Davis, 1967; Göktaş and İşçi, 2011; Goodman and Kruskal, 1954).

We calculated Lambda (λ), Kappa (ꝁ), and Gamma (γ) for each simulation run. Each run is then compared to the original hierarchy, which is the hypothetical "perfect" CAS with 121 constructs.

5 Analysis

Our preliminary analysis demonstrates Lambda (λ), Kappa (ꝁ), and Gamma (γ) change at different rates depending on the kind of movement and swap performed. Table 1 presents the mean change and standard deviations for the first 100 runs of each swap and movement. Preliminary results indicated that λ does not change when there are only swaps between two constructs. However, in cases where there are only movements of constructs, λ decreases by about 0.14. In contrast, Kappa (ꝁ) is more sensitive to swaps than movements. A level swap can cause a drop in Kappa with any value from 0.001 to 0.05. In contrast, Kappa in movement with children causes a drop from 0.001 to 0.008. As an example, Kappa in run 20, drops from 0.95 to 0.78 in level swap, while in movement with children, it drops from 0.99 to 0.86. Finally, Gamma appears to be the most stable across changes of the levels.

Type of Change	Mean Difference for Lambda	Lambda (SD)	Mean Difference For Kappa	Kappa (SD)	Mean Difference for Gamma	Gamma (SD)
Hierarchy Swap	0	0	0.005	0.023	0.007	0.035
Level Swap	0	0	0.004	0.029	0.002	0.013
Diagonal Swap	0	0	0.006	0.033	0.007	0.032
Random Swap	0	0	0.005	0.041	0.005	0.056
Movement w Children	0.002	0.017	0.002	0.019	0.002	0.023
Movement w/o Children	0.003	0.018	0.003	0.016	0.004	0.02
Randomly performed action	0.001	0.009	0.004	0.031	0.005	0.036

Table 1. Means and standard deviations for the first 100 runs of each swap and movement

6 Conclusion

For Computer-Adaptive Survey (CAS), we first created a new q-sorting methodology. Our q-sorting methodology can be beneficial in any area where the arrangement of constructs relies on perception and the constructs have a parent-child relationship. As an example, perceived usefulness and perceived ease of use can be redefined as CAS constructs to identify why people do not perceive a piece of IT as useful or easy to use.

In this method, raters are asked to sort items into constructs and map constructs together in a hierarchy. The hierarchies that independent raters develop are transformed into a quantitative form, and that quantitative form is tested to determine the inter-rater reliability of the individual branches in the hierarchy. The hierarchies are then successively transformed to test if they branch in the same way. To assess inter-rater reliability of the hierarchies of raters, a threshold(s) is necessary. To develop thresholds, we first generated a hypothetical "perfect" hierarchy for CAS which had 121 constructs. We then systematically modify this and calculated how various measures of "sameness" are impacted by these modifications. Hence we created four swaps and two movements. We ran a simulation to perform tests where random movements and swaps are performed 100 times. We then measured the correspondence of each tree against the "perfect" tree using 3 measures of association, Goodman and Kruskal's Lambda , Cohen's Kappa, and Goodman and Kruskal's Gamma (Goodman and Kruskal, 1954). In our preliminary results provided several insights. First, is that Lambda (λ) does not change when there are only swaps between two constructs. Second, is that Kappa (ƙ) is more sensitive to swaps than movements, as Kappa (ƙ) drops faster in swaps than in movements. Finally, Gamma (γ) appears to be the most stable across changes of the levels.

Our results therefore suggest that rather than employ a single statistic to calculate "good enough" construct validity, a mix of statistics might be ideal. This thinking is similar to Hu and Bentler (1998) who showed that a combination of CFI, SRMR and RMSEA were ideal for evaluating the goodness of fit for structural equation model.

In our future work, we hope to explore and assess CAS in several areas. One is to continue our work with Lambda (λ), Kappa (ƙ), and Gamma (γ) to provide a "good enough" threshold for construct validity for CAS. Second, we hope to use the 3 measures of association for identifying the reasons of differences in the trees developed by independent raters. Third, we hope to assess the generalizability and credibility of the results of CAS against an external criterion, such as online customer reviews.

To do all of these things, we are developing a café satisfaction CAS to identify the constructs customers are least satisfied with. In this CAS, we have a total of 175 questions. However, the regular respondent only has to answer an average of 20 questions for the CAS to identify which issue (e.g., food quality, price) the respondent is least satisfied with. The item bank for the café satisfaction CAS consisted of 175 survey questions. It was developed as follows. First, we synthesized existing café satisfaction surveys (Hwang & Zhao, 2010; Kim, Moreo, & Yeh, 2005; Liang & Zhang, 2009; Pizam & Ellis, 1999; Pratten, 2004; Ryu & (Shawn) Jang, 2008; Saglik, Gulluce, Kaya, & Ozhan, 2014; Shanka & Taylor, 2005). From these existing surveys, we first, identified the overarching constructs which were the most prominent, hence, in our café satisfaction CAS, there are five overarching constructs: (1) convenience, (2) service quality, (3) quality of food and drink, (4) price and value, and (5) ambiance. Most surveys comprised 30 items or less- because of a lack of respondent patience, often only a single question is to be asked per construct. In addition, the first author trawled internet café forums to identify common complaints. New items were developed based on those complaints. Here, principles from grounded theory (Strauss and Corbin, 1994) specifically, axial coding, guided us. Hence, approximately 400 items were collected. Items across the surveys and from the forums were then compared and duplicates were discarded. Fewer than 300 items remained after this step and two independent raters blinded to the study's purpose went through the items and marked items that were either vague or repetitive. Approximately 60 items were dropped here. Next, we rearranged and reorganized the questions into a hierarchy. We assessed the instrument for construct and content validity and items which did not "fit in" the hierarchy were dropped, leaving only 175 items (Sabbaghan, Gardner, and Chua, 2016). The 175 items only focus on café related issues, hence if one were to use this for another context such as restaurants, the content and construct validity would need to be re-evaluated.

Finally, we hope to assess CAS in other fields, such as developing our own technology acceptance model (TAM) instrument to determine why the constructs encourage the intention to use technology.

References

Anderson, J. and D. Gerbing (1988). “Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach.” *Psychological Bulletin*, 103(3), 411–423.
Bagozzi, R. P. (2007). “The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift”. *Journal of the Association for Information Systems*, 8 (4), 244–254.
Berdie, D. R. (1989). “Reassessing the value of high response rates to mail surveys”. *Marketing Research,* 1 (September), 52–65.
Block, J. (1961). The Q-sort Method in Personality Assessment and Psychiatric Research. 3–26.
Boudreau, M.C., Gefen, D. and D. W. Straub (2001). “Validation in Information Systems Research.” *MIS Quarterly*, 25(1), 1-15.
Cheung, G. W. and R. B. Rensvold (2002). “Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance.” *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255.
Cohen, J. (1968). “Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit.” *Psychological Bulletin,* 70(4), 213–220.
Davis, J. A. (1967). “A Partial Coefficient for Goodman and Kruskal’s Gamma.” *Journal of the American Statistical Association,* 62(317), 189–193.
Diamantopoulos, A. and M. Winklhofer (2001). “Formative Indicators : Scale.” *Journal of Marketing Research,* 38(2), 269–277.
Deutskens, E., de Ruyter, K. Wetzels, M., and P. Ooserveld (2004). “Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study“, *Marketing Letters*, 15 (1), 21–36.
Embretson, S. E. and S. P. Reise (2000). “Item response theory for psychologists.” *Quality of Life Research,* 4(3), 1–371.
Everitt, B. S. (1992). *The Analysis of Contingency Tables, Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.
Fundin, A., and M. Elg (2010). “Continuous Learning Using Dissatisfaction Feedback in New Product Development Contexts,” *International Journal of Quality & Reliability Management* (27:8), 860–877.
Galesic, M. and M. Bosnjak (2009). “Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey.” *Public Opinion Quarterly*, 73(2), 349–360.
Göktaş, A. and Ö. İşçi (2011). “A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation.” *Metodoloski Zvezki*, 8(1), 17–37.
Goodman, L. A. and W. H. Kruskal (1954). “Measures of association for cross classifications.” *Journal of the American Statistical Association*, 49(268), 732–764.
Groves, R. M. (2006). “Nonresponse Rates and Nonresponse Bias in the Household Surveys.” *Public Opinion Quarterly*, 70(5), 646–675.
Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. and R. Tourangeau (2004). *Survey Methodology, Statistics*. Edited by J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, and T. W. Smith. Wiley-Interscience (Wiley series in survey methodology).
Hair, J. F., Anderson, R. E., Tatham, R. L. and W. C. Black (1998). “Multivariate Data Analysis.” *International Journal of Pharmaceutics*. Edited by J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black. Prentice Hall (Journal of Econometrics).
Haschke, A. Abberger, B., Wirtz, M., Bengel, J., and H.Baumeister (2013). “Development of short form questionnaires for the assessment of work capacity in cardiovascular rehabilitation patients.” *International journal of occupational medicine and environmental health,* 26 (5), 742–50.
Hayes, B. E. (1992). *Measuring Customer Satisfaction*. Milwaukee, WI: ASQC Quality Press.
Heerwegh, D. and G. Loosveldt (2006). “An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys.” *Journal of Official Statistics*, 22(2), 191–210.

Hol, a. M., Vorst, H. C. M. and G. J. Mellenbergh (2008). “Computerized Adaptive Testing of Personality Traits.” *Journal of Psychology,* 216(1), 12–21.

Hu, L. and P. M. Bentler (1998). “Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification.” *Psychological Methods*, 3(4), 424–453.

Hwang, J., and J. Zhao (2010). “Factors Influencing Customer Satisfaction or Dissatisfaction in Restaurtant Business Using Answertree Methodology,” *Journal of Quality Assurance in Hostpitality & Tourism* (11:2), 93–110.

Jarvis, C. B., MacKenzie, S. B. and P. M Podsakoff (2003). “A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research.” *Journal of Consumer Research*, 30(September 2003), 199–218.

Jones, C. A., and B. Lin (1997). “Some Issues in Conducting Customer Satisfaction Surveys,” *Journal of Marketing Practice: Applied Marketing Science,* (3:1), 4–13.

Kim, Ay.-S. Moreo, P. J. and R. J.M. Yeh (2005). “Customers’ Satisfaction Factors Regarding University Food Court Service“. In: *Journal of Foodservice Business Research,* 7 (4), 97–110.

Liang, X. and S. Zhang (2009). “Investigation of customer satisfaction in student food service“. In: *International Journal of Quality and Service Sciences*, 1 (1), 113–124.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems, Applied Psychological Measurement*. Hillsdale, N.J.: L. Erlbaum Associates.

Ma, S. C., Chien, T. W., Wang, H. H., Li, Y. C., and M. S. Yui (2014). “Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bully ing.” *Journal of medical Internet research*, 16(2), e50.

Mackenzie, S. B., Podsakoff, P. M. and P. M Podsakoff (2011). “Construct Measurement and Validation Procedures in Mis and Behavioral Research : Integrating New and Existing Techniques.” *MIS Quarterly,* 35(2), 293–334.

McKeown, B. F. and D. B. Thomas (1988). *Q Methodology* (Quantitative Applications in the Social Sciences series, Vol. 66). Newbury Park, CA: Sage.

Merrell, C. and P. Tymms (2007). “Identifying reading problems with computer-adaptive assessments.” *Journal of Computer Assisted Learning,* 23, 27–35.

Montgomery, J.M. and J. Cutler (2013). “Computerized Adaptive Testing for Public Opinion Sur- veys.” *Political Analysis*, 21, 141–171.

Petter, S., Straub, D. and Rai, A. (2007). “Specifying Formative Constructs in Information Systems Research.” *MIS Quarterly,* 31(4), 623–656.

Pinsonneault, A., and K. L. Kraemer (1993). “Survey Research Methodology in Management Information Systems: An Assessment.” *Journal of Management Information Systems,* (10:2), 75-105.

Pizam, A., and T. Ellis (1999). “Customer Satisfaction and its Measurement in Hospitality Enterprises.” *International Journal of Contemporary Hospitality Management,* (11:7), 326-339.

Porter, S. R., Whitcomb, M. E. and W. H. Weitzer (2004). “Multiple surveys of students and survey fatigue.” *New Directions for Institutional Research*, 2004(121), 63–73.

Pratten, J. D. (2004). “Customer Satisfaction and Waiting Staff.” *International Journal of Contemporary Hospitality Management,* (16:6), 385–388.

Reeve, B.B. and P. Fayers (2005). “Applying item response theory modelling for evaluating question naire item and scale properties.” *Assessing quality of life in clinical trial: methods and practice,* 2, 55–73.

Rudick, M. M., Yam, W. H. and Simms, L. J. (2013). “Comparing countdown- and IRT-based approaches to computerized adaptive personality testing.” *Psychological assessment,* 25(3), 769–79.

Ryu, K. and S. (Shawn) Jang(2008). “DINESCAPE: A Scale for Customers’ Perception of Dining Environments“. *Journal of Foodservice Business Research*, 11 (1), 2–22.

Sabbaghan, S., and Gardner, L., and C. E. H. Chua (2016). “A Q-Sorting Methodology for Computer-Adaptive Surveys“. *ECIS.*

Saglik, E., Caglar Gulluce, A., Kaya, U., and , K. C. Ozhan (2014). "Service Quality and Customer Satisfaction Relationship : A Research in Erzurum Ataturk University." *American International Journal of Contemporary Research,* 4(1), 100–117.

Sampson, S. E. (1998). "Gathering Customer Feedback via the Internet: Instruments and Prospects." *Industrial Management & Data Systems,* 98(2), 71–82.

Segars, A. H. and V. Grover (1998). "Strategic Information Systems Planning Success: An Investigation of the Construct and Its Measurement." *MIS Quarterly,* 22(2), p. 139-163.

Shanka, T., and R. Taylor, (2005). "Assessment of University Campus Café Service: The Students' Perceptions." *Asia Pacific Journal of Tourism Research,* (10:March 2015), 329–340.

Sinclair, M. D., Clark, J. R., and D. A. Dillman (1993). "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rate for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly,* (57), 289–304.

Straub, D. and Gefen, D. (2004). "Validation Guidelines for IS Positivist Research." *Communications of the Association for Information Systems*, 13, 380–427.

Strauss, A., and J. Corbin (1994). "Grounded Theory Methodology," *Handbook of Qualitative Research*, 273–285.

Stricker, L. J., Wilder, G. Z. and B. Bridgeman (2006). "Test Takers' Attitudes and Beliefs about the Graduate Management Admission Test." *International Journal of Testing*, 6(3), 255–268.

Thompson, N. and Weiss, D. (2011). "A framework for the development of computerized adaptive tests." *Practical Assessment, Research & Education*, 16(1).

Thorpe, G. L. and A. Favia (2012). "Data Analysis Using Item Response Theory Methodology : An Introduction to Selected Programs and Applications." *Psychology Faculty Scholarship*, 20.

Wisner, J. D., and W. J. Corney (2001). "Comparing Practices for Capturing Bank Customer Feedback." *Benchmarking: an International Journal*, (8:3), 240–250.

You, W., Xia, M., Liu, L. and D. Liu (2012). "Customer knowledge discovery from online reviews." *Electronic Markets*, 22(3), 131–142.