Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2017 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 7-19-2017

Computer-Adaptive Surveys (CAS) as a Means of Answering Questions of Why

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

Lesley A. Gardner
The University of Auckland, l.gardner@auckland.ac.nz

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

Computer-Adaptive Surveys (CAS) as a Means of Answering Questions of Why

Completed Research Paper

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

Traditional surveys are excellent instruments for establishing the correlational relationship between two constructs. However, they are unable to identify reasons why such correlations exist. Computer-Adaptive Surveys (CAS) are multi-dimensional instruments where questions asked of respondents depend on the previous questions asked. Their principal advantage is they allow the survey developer to input a large number of potential causes. Respondents then roll down through the causes to identify the one or few significant causes impacting a correlation. This study compared a café satisfaction CAS to a traditional survey of the same item bank to test whether CAS performs its intended task better than a traditional survey. Our study demonstrates that when one is trying to find root cause, CAS achieves a higher response rate, requires fewer items for respondents to answer, has better item discrimination, and has a higher agreement among respondents for each item.

Keywords: *Computer-Adaptive Surveys, traditional surveys, conclusion validity*

Introduction

Computer-Adaptive Surveys (CAS) are most useful for the situation where there are two or more constructs which are correlated and one desires to understand the reason why those constructs are correlated. Consider two different scenarios. In one scenario, a technology is perceived as something desirable to adopt, because it is useful and easy to use, i.e., adheres to TAM/UTUAT (Bagozzi 2007; Gefen et al. 2003; Legris et al. 2003). However, as researchers, we want to know what characteristics of the technology make it useful or easy to use- questions existing survey instruments cannot answer (Pinsonneault and Kraemer 1993). In a second scenario, a café owner wants to know not only what aspect of customer service he could improve on, but also how he can improve (Fundin and Elg 2010; Sampson 1998; Wisner and Corney 2001). However, with traditional survey techniques, he would be forced to give customers a huge survey to identify the salient issues for improvement.

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 et al. 2013; Hayes 1992; Sinclair et al. 1993). If the survey is too long, respondents will suffer from a fatigue effect, and not answer questions properly (Berdie 1989; Deutskens et al. 2004). If the survey is too short, it will not provide enough information to adequately capture respondent sentiment (Hayes 1992). Finally, long surveys can be broken up over time (Goodman et al. 1992). However, these often still are unable to identify root causes of issues (Hayes 1992; Peterson and Wilson 1992).

One way of handling these issues is through a Computer-Adaptive Survey (CAS), which is a new kind of survey. 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 of the questionnaire items- unfilled questions cannot be treated as non-responses. Unfortunately, methods for validating CAS are under researched- to the point where they are nonexistent.

This paper attempts to validate that CAS outperforms a traditional survey in identifying one or a few narrowly defined constructs that respondents as a whole have the greatest or least affiliation to. To do this, we compare a customer satisfaction CAS against its equivalent traditional survey. We find that CAS has several advantages over the traditional survey, as CAS has a higher response rate, requires fewer items for respondents to answer, has better item discrimination in that respondents tend to provide more extreme scores in CAS, and has a higher agreement among respondents for each item.

The paper is constructed in the following manner. The next section introduces the related literature, describing CAS and its design. We then compare the results of CAS with a traditional survey of the same item bank. Finally, findings are presented and discussed.

Computer Adaptive Surveys (CAS) vs. Traditional Surveys

CAS and traditional surveys differ, as traditional surveys are about establishing a correlation between two or more constructs, whereas CAS investigates why those constructs are correlated (i.e., identify root cause). As an example, a traditional survey establishes a high negative correlation between food quality and satisfaction. A CAS is used to identify whether customers are unhappy with the taste, temperature, texture, arrangement, etc. of the food.

To identify root cause, it is often necessary to ask many questions. Most surveys are not designed to be very long. Therefore, they are not especially designed to be informative or diagnostic for identifying root cause (Goodman et al. 1992; Hayes 1992; Peterson and Wilson 1992). Also, in most surveys, the respondent is intended to answer the majority of questions. With a large 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 et al. 2004). Finally, the effort to complete a CAS grows logarithmically with the length of the CAS. In contrast, effort grows linearly with the length of a traditional survey. This is because questions in a CAS are represented in a tree and the respondent navigates down a branch of the tree instead of doing the whole survey.

Several techniques have been used to obtain more information from traditional surveys (Calinescu et al. 2012; Calinescu and Schouten 2016; Groves and Heeringa 2006; Schouten et al. 2013). One is open-ended questions where respondents write down their feelings (Hayes 1992). However, open-ended survey responses are often time-consuming to analyse (Jackson and Trochim 2002). In most cases, interpretation of responses to open-ended questions is difficult (Jones and Lin 1997). In addition, response rates for open-ended questions are lower than for multiple choice ones (Krosnick, 1999).

Follow-up contacts have been consistently reported as being the most powerful technique for increasing response rates, both in mail and online surveys (Wiley et al. 2009). Follow-up surveys or interviews are used to obtain more in depth information (Locker 2000; Woodruff 1997). As an example, Mirror Wave Surveys are three-question surveys that are followed up every time a new product is purchased. This way, a customer’s purchasing patterns can be monitored (Payne and Frow 2013). When the goal is to drill down into individuals’ perceptions, numerous follow-ups may be required. However, studies show when multiple follow-ups are used, the response rates decreased after the second follow-up (Cook et al. 2000; Zhang 2000). Repeated follow ups have diminishing returns and may be treated as spam, hence irritating potential respondents (Fan and Yan 2010; Montgomery and Cutler 2013; Porter 2004; Rogelberg and Stanton 2007).

Another way to get more precise responses is to increase survey length. However, survey fatigue, which is also called over-surveying, survey disillusionment, or survey saturation can cause nonresponse or false responses. There are several studies showing that length has a negative effect on participation rates (Galesic and Bosnjak 2009; Krosnick 1999). As an example, in one study, two surveys were given where the long version took approximately 30–45 minutes to finish, while the short version could be completed within 15 to 30 minutes. The results indicated that the response rates were significantly higher in the shorter version (Deutskens et al. 2004).

Computer-Adaptive Surveys

Computer-Adaptive Surveys (CAS) are multi-dimensional instruments where questions asked of respondents depend on the previous questions asked. Its principal advantage is 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 our case, the things about the café the respondent is least satisfied with. 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 et al. 2004).

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 through questions which are dynamically assigned based on answers to the questions the participant answered previously (Thompson and Weiss, 2011). For example, the Graduate Management Admission Test (GMAT) asks the respondent to answer language and mathematical questions in increasing order of difficulty (Stricker et al. 2006). The next question asked of a respondent depends on whether the previous questions were answered correctly. Similarly, in the Merrell and Tymms test (2007), the aim is to understand the reading ability of students to provide better feedback and implement appropriate reading practices.

CAS and CAT are similar in that they both use a very large item bank, and rely on the idea that the response of one item directs the next item. They are different in several ways. One is that their goals are very different. CAS aims to identify the child constructs most salient to a correlation between two constructs for 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.

Second, they differ on how questions are administered. CAT uses Item Response Theory (IRT) that refers to a set of mathematical models which, 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 and Fayers 2005). In the IRT, there are two model(s), which can be either dichotomous or 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 and Fayers 2005). Polytomous models are more common in surveys, such as measuring political knowledge (Montgomery and Cutler 2013), or measuring workplace bullying (Ma, Chien, Wang, Li, and Yui, 2014). In contrast, CAS uses an adaptive version of branching to arrange the questions. The lowest or highest score on a set of questions triggers the retrieval of related, but more precise questions. In addition, the question structures are also different. 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.

Finally, initiation and termination in CAS and CAT function in specific ways. In most cases, respondents taking a particular CAT test all begin in the same way. In contrast, we could have the first 20 respondents taking a CAS begin with generic questions about the café. If we realise that most respondents are indicating issues with the service, the next 20 respondents might begin at a lower level of the hierarchy- on the service-related questions. Similarly, a CAT terminates when the CAT has enough information to perform a diagnosis, either when a fixed number of items have been answered (Babcock and Weiss 2009; Ho and Dodd 2012; Shin et al. 2012) or because further questions in the item bank provide no additional statistically meaningful information (Thompson and Weiss, 2011). In contrast, a CAS ends either when one has fully traversed a set number of branches of the hierarchy, or when the user reaches some threshold for a proxy for fatigue (e.g., user answers a certain number of questions).

CAS Implementation and Design

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. Child constructs are formative and represent more precise dimensions of their parent. To illustrate the process, consider the CAS we have developed, which is designed to elicit the problems customers have with cafés. The item bank consists of 175 survey questions. Most customer satisfaction surveys comprise 30 items or less- because of a lack of respondent patience, often only a single question is asked per construct (Heerwegh and Loosveldt 2006). There are five overarching constructs in our survey: (1) convenience, (2) service quality, (3) quality of food and drink, (4) price and value, and (5) ambiance.

How CAS Works

In CAS, respondents perform a depth-first traversal of the tree, where each stage of the traversal involves the respondent rating all items in the stage. Respondents then receive only child constructs associated with the lowest or highest rated constructs. Each respondent could traverse the CAS hierarchy in a different way. To illustrate, see Figure 1. If food is the area 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 menu choice, CAS 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 respondent continues to answer questions, CAS navigates deeper down the tree and questions roll down until the respondent hits one set of constructs with no children, for example, that there are insufficient vegetarian items on the menu. If most respondents agree there are insufficient vegetarian items on the menu, this would indicate lack of vegetarian items is the root cause for many customers' dissatisfaction. Of course, not every respondent would navigate the hierarchy the same way. Thus, aggregating the results from respondents allows the researcher to observe the multiple major problems across respondents. In addition, it allows managers of commercial enterprises to quickly find key issues to address.

However, it is possible there are more candidate constructs in that level of the hierarchy than allowed by a threshold. For example, assume initially that the respondent rated question1 "I am overall satisfied with the ambiance of the café" with a 1, and question3 "I am overall satisfied with the quality of food/drinks of the café" with a 1. CAS must retrieve constructs associated with the least scoring question. But CAS cannot determine which of the questions 1 or 3, to choose from. At this point, CAS asks the respondent which of questions 1 or 3 is most relevant to the respondent. If question 3 is chosen, then the constructs mapping to question 3 are retrieved by CAS. The process repeats until no subsidiary constructs can be selected, whereupon CAS stops.

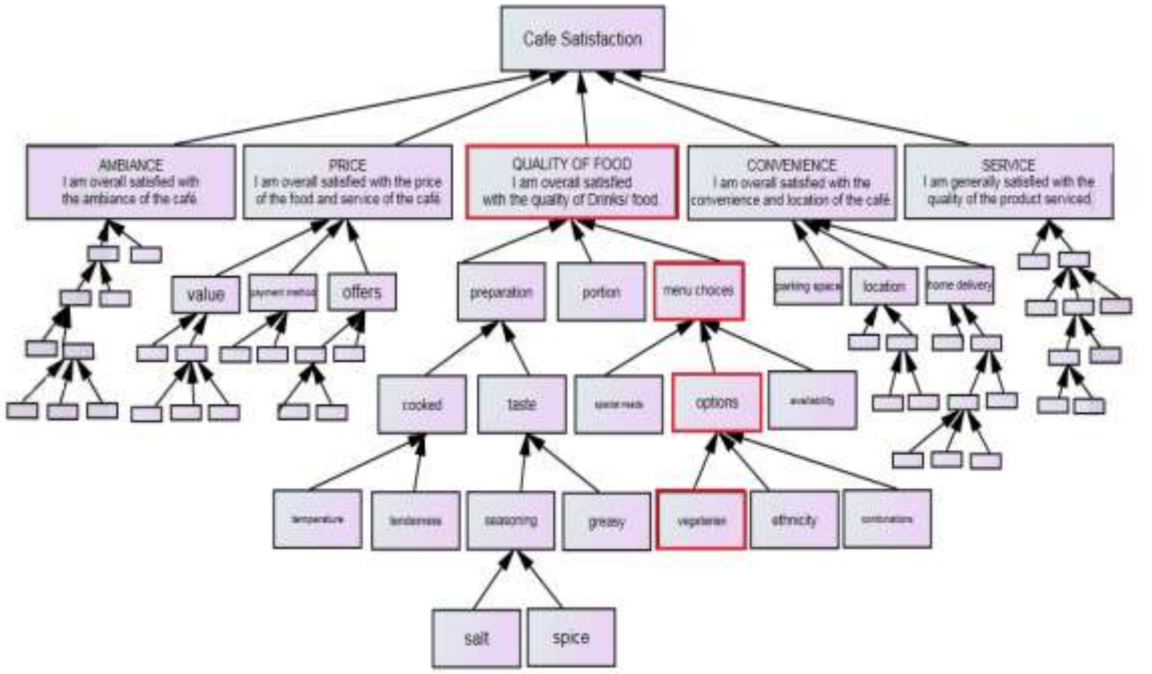

Figure 1. How CAS works for café satisfaction

Assessing the Results of CAS versus the Traditional Survey

Assessing the validity of CAS's results requires different techniques from traditional surveys for a number of reasons. One, unlike traditional surveys, a CAS cannot be used to find cause in the sense of there being an independent variable and dependent variable on the survey. In a CAS, cause is assumed; there is an implicit dependent variable (e.g., customer satisfaction) that the survey does not ask. Second, in CAS, there are child- parent relationships, where constructs that are children to other constructs in the hierarchy should represent some dimension of the parent construct. As a result, in CAS, we can expect high correlations between a parent and one child (e.g., if a respondent says they are dissatisfied with service, there is at least one subdimension of service they are unhappy about). However, because the subdimensions are orthogonal (as child constructs are formative on their parent), we expect low correlations across subdimensions (e.g., that someone is dissatisfied with the efficiency of service does not mean they are dissatisfied with the quality of service). Traditional statistical techniques cannot handle such complex correlations between items on a survey. Given the problems with traditional approaches, we argue for a new approach to assess the validity of CAS's results.

This paper solely focuses on assessing the conclusion validity of CAS. We test whether CAS is better at finding root cause than a traditional survey. There are clearly other forms of validity, such as conclusion, construct, internal, and external validity (Shadish et al. 2002). Testing these other forms of validity of CAS is the purview of other research. For example, in earlier research, we developed a methodology for evaluating the construct validity of CAS (Sabbaghan et al. 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. In addition, we developed techniques for evaluating the similarity of those hierarchies for evaluating the construct validity of CAS (Sabbaghan et al. 2017).

We employ café satisfaction as a proxy context for evaluating the efficacy of CAS. CAS typically identifies one or a few narrowly defined constructs that respondents as a whole have the greatest or least affiliation to. Hence for CAS to have credible results would mean that the "correct construct(s)" have been identified. Our hypotheses compare the café satisfaction CAS against a traditional survey on properties that suggest CAS is better for identifying these "correct constructs."

Many studies have tried to decrease nonresponse rates by using different techniques such incorporating incentives, or adaptive designs (Calinescu and Schouten 2016; Groves and Heeringa 2006; Peytchev 2013; Singer 2012; Wagner 2008). One supposed benefit of CAS is it will have a lower non-response rate, because respondents are required to answer fewer items to complete the survey. Therefore, we hypothesize:

H_1 Our café satisfaction CAS will have a higher response rate than the traditional survey.

In a CAS about dissatisfaction, a non-chosen item should mean the respondent was not dissatisfied with the parent of the item, and therefore the item itself. To test whether this is true, we compare the items unanswered by all respondents in CAS against responses in a traditional survey. A demonstration that the corresponding items in the traditional survey tend towards "satisfaction" rather than "dissatisfaction" on the Likert scale would provide evidence the CAS is performing as intended. Hence, we hypothesize:

H_2 For all non-chosen items in CAS, the corresponding item in the traditional survey will tend towards "satisfaction" rather than "dissatisfaction."

Respondents may not produce quality responses for various reasons, such as lack of motivation or cognitive overload (Krosnick 1991). Disengaged respondents may not provide quality responses, they may either leave items blank or provide false responses (Galesic and Bosnjak 2009; Groves 2006; Heerwegh and Loosveldt 2006; Porter et al. 2004; Tourangeau et al. 2004) . One example of false responses is answering a sequence of items in exactly the same way, i.e., straightlining (Fricker et al. 2005). As an example, a traditional survey with a 5-point Likert scale could have a straight line sequence of 5s, indicating the respondent filled those items without bothering to read the questions (Krosnick, 1991; Zhang and Conrad, 2013). Another example is choosing a neutral response options (for example, "No opinion" and "Don't know" answer options) instead of substantive options such as only choosing the value 3 from a 5-point Likert scale. In addition, false responses could also be produced when respondents randomly choose an answer to avoid the substantial cognitive effort required of processing each questionnaire item (satisficing) (Krosnick 1991). Other factors such as speeding (i.e., giving answers very quickly) (Zhang and Conrad 2013), fatigue and boredom (Galesic and Bosnjak 2009) accumulate throughout the survey, and decrease the willingness of respondents to invest in the effort needed for good quality answers.

CAS has two distinguishing features compared to traditional surveys. One is the role of choice questions. When a respondent indicates equal dissatisfaction with two things, CAS prompts the respondent to identify the item they are least satisfied with. Second is that items are administered so that previous questions determine the next questions, hence increasing the amount of interaction with respondents and distinguishing the branches which require more attention. Studies have shown that interactions such as conditional branching can not only assist with decreasing nonresponse items, but also increase respondents' attention to each question and increasing the quality of responses (Krosnick 1991; Manfreda, Batagelj and Vehovar 2002).

In this study, one of the aims is to assess which instrument (CAS or traditional survey) produces responses that better differentiate constructs associated with root cause. For this to be true, respondents should rate certain items more highly than others. We argue because respondents are more engaged with CAS, respondents will provide more truthful answers, and thus such discrimination between items will occur. Conversely, the length, and monotony of the traditional survey causes the respondent to switch off. The respondent no longer answers questions in a traditional survey in a focused way. Responses in a traditional survey vary less, because the respondent is less engaged. Thus, we hypothesize:

H_3 In the café satisfaction CAS, the standard deviation within an individual, across items will be higher than the traditional survey.

Response quality has been an important factor in survey research. As an example, in one study, the findings indicated that most peoples' answers to survey questions are completely random. In fact, if the same people are asked the same question in repeated interviews, only about half give the same answer (Zaller and Feldman 1992). In other studies, the findings indicated that in shorter surveys, respondents suffer less fatigue, hence there are less blank items and false responses (Deutskens et al. 2004; Galesic and Bosnjak 2009).

When respondents are more engaged, they are able to focus more on the questions. This means we expect respondents will produce answers in CAS that are "closer to the truth." By implication, if a café is genuinely poor at a particular area of satisfaction, we should receive more consistent feedback to that effect. Conversely, if a café is satisfactory in an area of satisfaction, CAS should more closely reflect that. In a traditional survey, because the respondent is less engaged, there is more randomness in the answers, thereby increasing the error. Hence:

H_4 The café satisfaction CAS compared to the traditional survey will have a lower standard deviation within items across respondents.

Methodology

The remainder of this section describes how the sample, instrument, and data collection was conducted.

Sample

We selected the 5 most popular cafes in the university campus to compare and assess the results. The target population was customers of the 5 cafes- i.e., university students. Our sample was students from the Information Systems and Operation Management (ISOM) Department and Economics Department. We were limited to only two departments due to conditions imposed by our ethics committee. There were approximately 5700 undergraduates and 120 post graduates in both departments. An invitation to participate in our study was disseminated through the university student learning management system. As an incentive to participate, respondents were entered into a lucky draw worth 20 New Zealand dollars. 20 respondents won the lucky draw.

We collected our data in two waves from each instrument. We performed a non-response bias test across the two waves, with the second wave being a proxy for non-response. The test was performed by running an independent sample t-test for each item having at least 20 responses. For the café satisfaction CAS, we found the mean scores were not significantly different across all items and thus there was no evidence for non-response bias. However, for the traditional survey, we found that for three items, the mean scores were significantly different. However, at a p-value of 0.05, we expect 1 in 20 tests to be significant. Out of 175 items, 3 item were significant, which is well within the margin of error of a test of significance (McHugh 2011).

Instrument

The item bank for the café satisfaction CAS and the traditional survey consisted of 175 survey questions. It was developed as follows. First, we synthesized existing café satisfaction surveys (Hwang and Zhao 2010; Kim et al. 2005; Liang and Zhang 2009; Pizam and Ellis 1999; Pratten 2004; Ryu and (Shawn) Jang 2008; Saglik et al. 2014; Shanka and Taylor 2005). 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, as existing categories were saturated and items were explored and similarities were found. This continued until no more new conceptual categories emerged. As an example, if we found in an internet café forum, customers were complaining about payment methods, we would explore payment methods and explore different possible ways of payment in cafés (e.g., cash, credit card, vouchers). This would continue until we had covered all existing ways. 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. Constructs are mapped together in a hierarchy with constructs concerning higher level concepts linking to constructs with greater precision. Two blind and independent raters grouped and mapped the constructs together. The hierarchies were then transformed into a quantitative form, and that quantitative form was tested to determine the inter-rater reliability of the individual branches in the hierarchy. The hierarchies were then successively transformed to test if they branched in the same way. We have assessed the instrument for construct and content validity and items which did not "fit in" the hierarchy were dropped, leaving only 175 items (Sabbaghan et al. 2016, 2017).

Data collection

The data was collected in the following manner. First, respondents were presented a consent form and agreed with it. Respondents were informed of the length of the survey (i.e., 175 items). We did this, because research suggests informing subjects up front about this increases response rates (Marcus et al. 2007). If the respondent discovers a survey is longer than anticipated, trust placed in the researcher by the respondent is revoked, leading to a higher nonresponse or an increase of dropouts (Heerwegh and Loosveldt 2006).

Respondents then filled out some demographic details. They next chose one of the 5 mentioned cafés they wished to assess. Next, each respondent was assigned a random number (1 or 2) which determined which survey each respondent would be assigned to. The traditional survey took approximately 20-40 minutes to finish and approximately 3 to 5 minutes was required for the CAS. We preformed two rounds of data collection. In the first round, the preliminary results from CAS indicated that "price and value" was the construct respondents were most dissatisfied with. However, the items for "price and value" were situated in the bottom of the traditional survey, where there was a high item nonresponse rate. Hence, in the second round, we moved the items associated with "price and value" to the top of the traditional survey and left CAS with no alterations. Approximately 510 responses were collected from both instruments for all five cafés. In the first wave, 170 responses were collected for CAS and 142 responses for the traditional survey. In the second, which was conducted 3 months after the first wave, 105 responses were collected for CAS and 93 responses were collected from the traditional survey.

After assessing the results, the overarching constructs "convenience" and "ambiance" did not have sufficient data for further assessment. Hence these constructs were dropped from the analysis and the constructs "service quality," "food and drink quality," and "price and value" for all cafés were selected for further assessment. It should be noted that the CAS results for the overarching constructs which were dropped had high mean scores which suggest respondents were not dissatisfied with either construct across the five cafes. Thus, dropping those constructs did not materially impact our analysis.

Analysis

In this study, we compare a café satisfaction CAS against its equivalent traditional survey. First, we hypothesize that our café satisfaction CAS will have a higher response rate than the traditional survey. We have two data points- non-response for the CAS and non-response for the traditional survey. Hence, statistical analysis to support the hypothesis is not possible. It should be noted that a comparative statistical analysis of non-response is generally impossible. However, we feel that just because a question cannot be answered numerically does not mean the question should not be asked. Instead, we qualitatively analysed comments given by respondents who had quit either instrument. As there were only comments for the traditional survey, we went through the comments and grouped together comments which were similar. In addition, the first author's grouping was compared to the grouping of one independent rater blind to the study purpose to ensure inter-rater reliability. Kappa was 0.73, above the recommended threshold of 0.7 (Landis and Koch 1977). We then gave each grouping a label to ascertain the causes for quitting.

In our second hypothesis, we expect that for any item(s) in CAS that was not chosen by any respondent, the corresponding item in the traditional survey will tend towards "satisfaction" rather than "dissatisfaction." To test this hypothesis, we calculated the mean and standard deviation of all items in the traditional survey which had no corresponding score across all completed CAS. We then ran a one-tailed one-sample t-test against a mean of 3 (i.e., "average" on a 5-point Likert scale) to assess if the true mean of the sample is higher than the comparison value (3). We also counted the number of people who scored each item less than 3.

In our third hypothesis, we expect the standard deviation within an individual, across items will be higher than in the traditional survey. Hence in both instruments, we calculated the standard deviation of all items an individual filled across the entire survey. We then calculated the mean and standard deviation of the standard deviations. Finally, a two sample one-tailed t-test was executed to compare the café satisfaction CAS and traditional survey.

Our final hypothesis is that the across-individual, within-item standard deviation in a café satisfaction CAS should be lower than the equivalent item in a traditional survey. Thus, we first calculated the standard deviation within item across individuals for both instruments for all cafés. This produced two standard deviations per questionnaire item, one for CAS, one for the traditional survey. It was not possible to perform a traditional parametric statistical analysis with this sample size. Instead, we performed a sign test (Dixon and Mood 1946). We subtracted the standard deviation of the traditional survey from CAS from each item, and assessed the extent the signs were positive or negative. An overwhelming number of negative signs would suggest items in the traditional survey tended towards a higher standard deviation.

Results

In our first hypothesis, we expected that our café satisfaction CAS would have a higher response rate than the traditional survey. For CAS, out of 275 respondents, 31 quit without providing a reason. For the traditional survey of 236 respondents, 70 quit. 40 respondents quit the traditional survey with a reason as tabulated in Table 1 and 30 respondents quit without providing a reason.

As noted earlier, statistical analysis is not possible given the data. Nevertheless, our analysis of the qualitative responses given is telling. Notably, 60% of respondents in the traditional survey quit the survey due to the large number of the items in the survey. The next most common reason given was only given by 7.5% of respondents. This suggests support for H1.

Table 1. Reasons for quitting the traditional survey

Reason for quitting	percentage	example
Number of items	60%	175 questions is too long
Long time	7.5%	The survey will take too long for me to do.
repetitive	7.5%	too long, a lot of repetition, never ending
Have put enough effort	7.5%	I have already answered 50 ....
Not sufficient motive	7.5%	I am not gonna answer 175 questions
boring	5%	I’m bored it’s too long mate
Incentive not sufficient	5%	too many questions, not worth a voucher

Table 1. Presents the reasons for quitting the traditional survey

In our second hypothesis, we expected that any item not chosen by any respondent doing CAS would have corresponding scores in a traditional survey indicating that respondents were satisfied with the item. The number of items no respondent answered in the CAS for cafés 1 to 5, are respectively, 64, 48, 54, 47, and 71. Consider Figure 1 which presents the corresponding items to the ones no respondent answered in the CAS, for the traditional survey for cafés 1-5. The red shades represent the responses with the value of 1, the purple shades responses with the value of 2, yellow 3, blue 4, and green 5 respectively. As illustrated, the majority of responses for those items are equivalent to three and above.

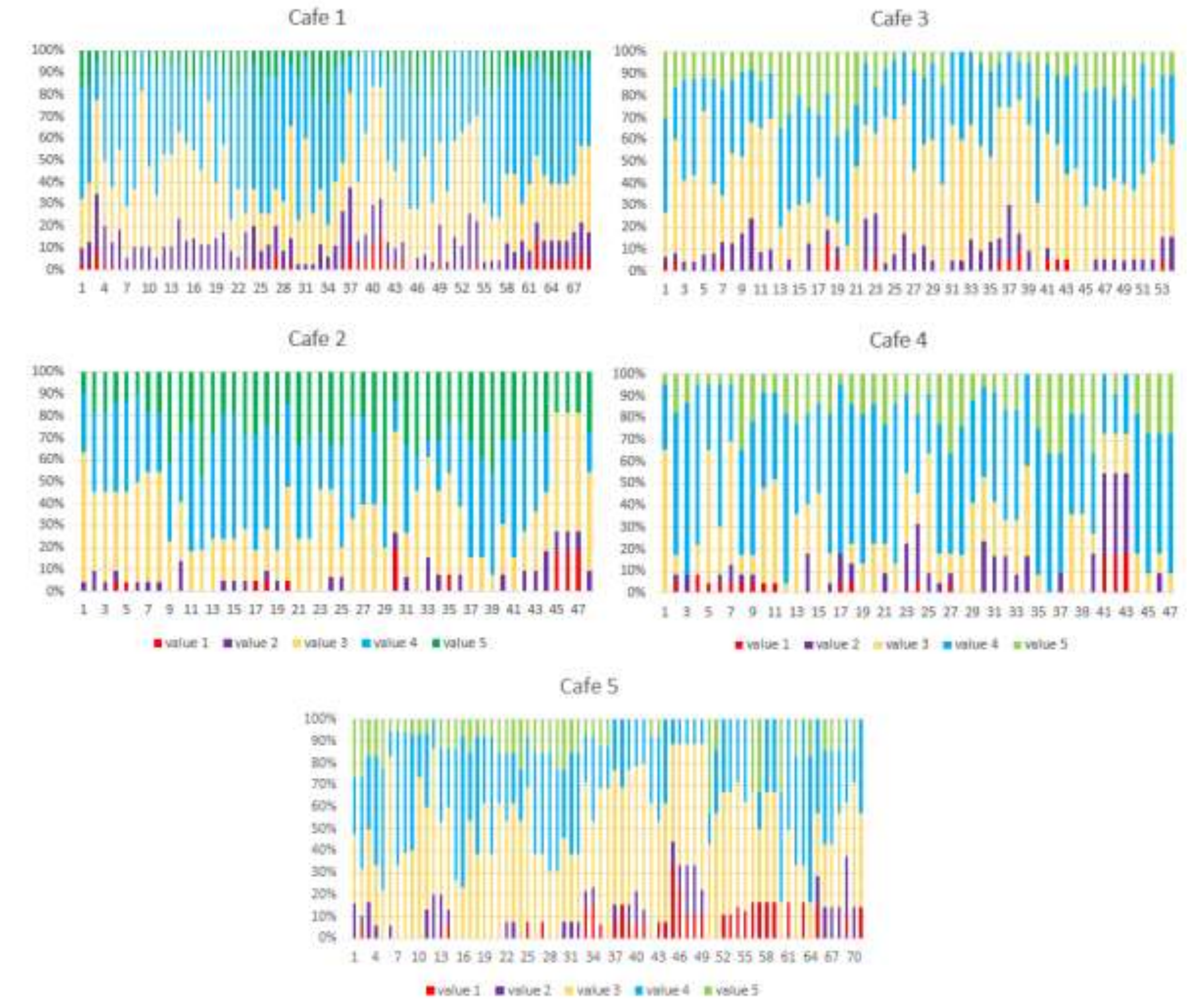

Figure 1. Presents the corresponding items no respondent answered in the café satisfaction CAS, for the traditional survey for café 1-5.

A one-tailed one-sample t-test against a mean of 3 (i.e., “average” on a 5-point Likert scale) of all the items which were null in CAS confirms this. In addition, we identified the number of items with a mean less than 3 for each café. About 5-11 percentage of items had a mean score of less than 3, the number dependent on the café analyzed. Given the significance of the test results and relative small percentage of responses indicating dissatisfaction, H2 is supported.

Table 2. The t-test of all the non-chosen items in CAS for all traditional surveys for the five cafés.

Cafe	Number of blank items in CAS	number of items with mean value < 3	Mean for traditional survey Items	Std. Deviation	Test Value = 3		
					t	df	Sig. (2-tailed)
cafe1	70	4	3.4798	0.29189	13.15	69	<0.001
cafe2	48	4	3.7887	0.35517	15.385	47	<0.001
cafe3	54	0	3.5264	0.29615	13.061	53	<0.001
cafe4	47	3	3.7028	0.41698	11.555	46	<0.001
cafe5	71	8	3.3646	0.40286	7.625	70	<0.001

Table 2. Presents the t-test of all the items which were non-chosen in CAS for all traditional surveys for the five cafés.

In our third hypothesis, we expected the standard deviation across all items answered by an individual will be higher in CAS than the traditional survey. Figure 2 presents the responses for both instruments for café 1. The grey shades represent the responses which are either item non-response in the traditional survey or have not been chosen by respondents in the café satisfaction CAS. The red shades represent the responses with the value of 1, the purple shades responses with the value of 2, yellow 3, blue 4, and green 5 respectively. Cafés 2-5 share a similar pattern with café 1. As the Figure demonstrates, only a fraction of the items is responded to in CAS. The construct "price and value" is indicated as one respondents are least satisfied with in both instruments as it has been answered by most respondents. In addition, in the café satisfaction CAS, certain items of the construct "price and value" show larger patches of red and purple compared to the traditional survey, which means respondents were very unhappy with these items. The graph suggests CAS will have a higher cross-item standard deviation.

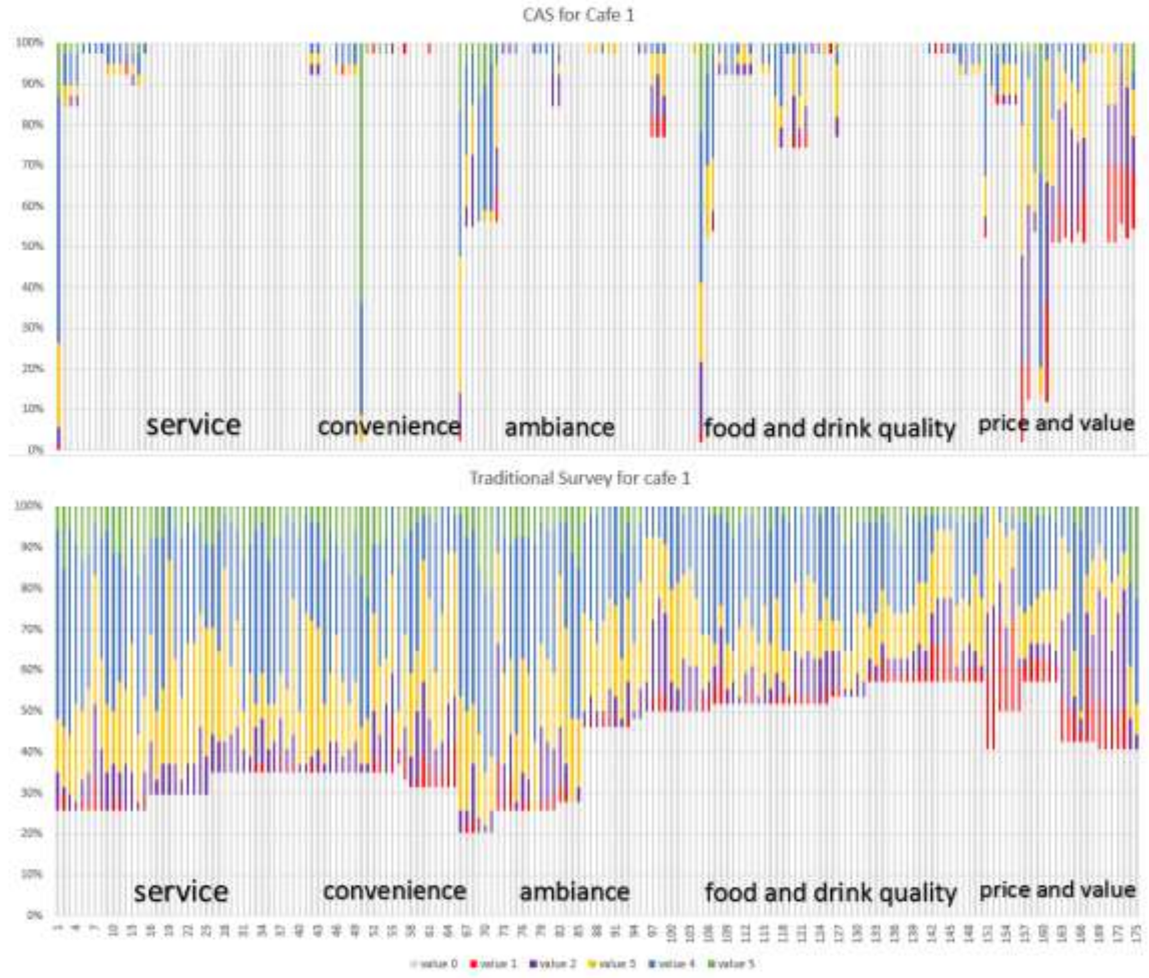

Figure 2. Presents the responses for both instruments for café 1.

We ran a two sample one-tail t-test of the mean standard deviation of every individual for all five cafes. According to Table 3, throughout the overarching constructs "service quality", "food and drinks quality" and "price and value" café satisfaction CAS has a higher standard deviation across items. Hence H3 is supported.

Table 3. Compares the standard deviation		
Construct	Overarching Constructs	
Type	*CAS*	Traditional survey
Mean	1.0113	0.656
Standard dev.	0.083	0.042
Observations	15	
Hypothesized Mean Difference	0	
t Critical one-tail	1.7081	
p (one-tail)	0.0003	

Table 3. Compares the standard deviation of individuals across items of CAS with the items of the traditional survey for the 5 cafes.

In our fourth hypothesis, we expected the café satisfaction CAS would have a lower within item standard deviation than the traditional survey, which reflects the error the traditional survey encourages. Table 4 presents the results of the sign test for the standard deviation of each survey question (175 items) for both café satisfaction CAS and the traditional survey. As the number of negative sign counts are overwhelmingly higher for each cafe, H4 is clearly supported.

Table 4. Sign Test				
Café		Positive sign count	Negative sign count	*p*-value
Café 1	CAS	11	164	<0.0001
	Traditional Survey			
Café 2	CAS	8	167	<0.0001
	Traditional Survey			
Café 3	CAS	4	174	<0.0001
	Traditional Survey			
Café 4	CAS	2	173	<0.0001
	Traditional Survey			
Café 5	CAS	13	162	<0.0001
	Traditional Survey			

Table 4. Sign Test for the standard deviation of each item of CAS with each item of the traditional survey for the 5 cafes

Discussion and Conclusion

This study compared a café satisfaction CAS to a traditional survey of the same item bank. As our study demonstrates, CAS has certain advantages over the traditional survey. First, given the same item bank, CAS tends to have a higher response rate as respondents are required to respond to fewer items (Galesic and Bosnjak 2009). Second, the CAS instrument can function as intended to identify constructs which respondents have the greatest or least affiliation to. When respondents do not fill out CAS items, it really means those items are unimportant for further analysis. Third, the cross-item standard deviation in CAS is higher than in a traditional survey, which means it is easier to identify salient items in CAS than in a traditional survey. Finally, CAS has a much lower random error rate, possibly because of reduced respondent fatigue, than a traditional survey. The CAS within-item standard deviation is much lower than in a traditional survey.

All of this suggests CAS can be applied as a useful instrument to further a number of areas of IS research. Traditional IS survey research has been useful for developing causal relationships between constructs, but has been poor at "unpacking" constructs to develop a rich understanding of how things work. As an example, 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. CAS provides a quantitative tool for doing this. 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. Just as structural equation modelling enabled IS researchers to study complex causal structures, CAS enables IS researchers to begin quantitatively exploring questions of why, thereby furthering theory (Sutton and Staw 1995).

As future research, we hope to explore and assess CAS in several areas. One, we want to assess CAS in other fields. This study compared a café satisfaction Computer-Adaptive Surveys (CAS) against a traditional survey of the same item bank. We chose café satisfaction instead of a more salient IS topic (e.g., TAM) because of the availability of existing, relevant instruments. If we were to develop our own TAM/UTUAT instrument, the potential quality of the questionnaire items would be a confound. By using existing items employed in traditional surveys, this confound is eliminated. Future work will apply CAS to develop TAM/UTUAT to determine why the constructs encourage intention to use. Two, we continue developing techniques, strategies, and algorithms to increase the accuracy and efficiency of CAS. This study used an adaptive version of branching for respondents to move from one set of items to another set according to a pre-defined criterion. Future study aims to assess and explore other possible options for determining the next set of question(s).

References

Babcock, B., and Weiss, D. J. 2009. "Termination Criteria in Computerized Adaptive Tests : Variable-Length CATs Are Not Biased," in *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*.

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), pp. 244–254.

Berdie, D. R. 1989. "Reassessing the Value of High Response Rates to Mail Surveys," *Marketing Research*, (1:September), pp. 52–65.

Calinescu, M., and Schouten, B. 2016. "Adaptive Survey Designs for Nonresponse and Measurement Error in Multi-Purpose Surveys," *Survey Research Methods*, (10:1), pp. 35–47.

Calinescu, M., Schouten, B., and Bhulai, S. 2012. "Adaptive Survey Designs that Minimize Nonresponse and Measurement Risk," *Statistics Netherlands*.

Cook, C., Heath, F., and Thompson, R. L. 2000. "A Meta-Analysis of Response Rates in Web-or Internet-Based Surveys," *Educational and Psychological Measurement*, (60:6), pp. 821–836.

Deutskens, E., de Ruyter, K., Wetzels, M., and Oosterveld, P. 2004. "Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study," *Marketing Letters*, (15:1), pp. 21–36.

Dixon, W. J., and Mood, A. M. 1946. "The Statistical Sign Test," *Journal of the American Statistical Association*, (41:236), pp. 557–566.

Fan, W., and Yan, Z. 2010. "Factors Affecting Response Rates of the Web Survey: A systematic review," *Computers in Human Behavior*, (26:2), Elsevier Ltd, pp. 132–139.

Fricker, S., Galesic, M., Tourangeau, R., and Yan, T. 2005. "An Experimental Comparison of Web and Telephone Surveys," *The Public Opinion Quarterly*, (69:3), pp. 370–392.

Fundin, A., and Elg, M. 2010. "Continuous Learning Using Dissatisfaction Feedback in New Product development contexts," *International Journal of Quality & Reliability Management*, (27:8), pp. 860–877.

Galesic, M., and Bosnjak, M. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey," *Public Opinion Quarterly*, (73:2), pp. 349–360.

Gefen, D., Karahanna, E., and Straub, D. W. 2003. "Trust and TAM in Online Shopping: An Integrated Mode," *MIS Quarterly*, (27:1), pp. 51–90.

Goodman, J. A., Broetzmann, S. M., and Adamson, C. 1992. "Ineffective - That's the Problem With Customer Satisfaction Surveys," *Quality Progress*, (25:5), pp. 35–38.

Groves, R. M. 2006. "Nonresponse Rates and Nonresponse Bias in the Household Surveys," *Public opinion quarterly*, (70:5), pp. 646–675.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. 2004. *Survey Methodology Statistics*, Wiley series in survey methodology, (J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, and T. W. Smith, eds.), (Vol.

2nd), Wiley-Interscience.

Groves, R. M., and Heeringa, S. G. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs," *Journal of the Royal Statistical Society*, (169:3), pp. 439–457.

Haschke, A., Abberger, B., Wirtz, M., Bengel, J., and Baumeister, H. 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), pp. 742–50.

Hayes, B. E. 1992. *Measuring Customer Satisfaction*, Milwaukee, WI: ASQC Quality Press.

Heerwegh, D., and Loosveldt, G. 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), pp. 191–210.

Ho, T.-H., and Dodd, B. G. 2012. "Item Selection and Ability Estimation Procedures for a Mixed-Format Adaptive Test," *Applied Measurement in Education*, (25:4), pp. 305–326.

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

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

Jackson, K. M., and Trochim, W. M. K. 2002. "Concept Mapping as an Alternative Approach for the Analysis of Open-Ended Survey Responses," *Organizational Research Methods*, (5:4), pp. 307–336.

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

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

Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, (5:3), pp. 213–236.

Krosnick, J. A. 1999. "Survey research.," *Annual Review of Psychology*, (50), pp. 537–67.

Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, (33:1), pp. 159–174.

Legris, P., Ingham, J., and Collerette, P. 2003. "Why do People Use Information Technology? A Critical Review of the Technology Acceptance Model," *Information & Management*, (40:3), pp. 191–204.

Liang, X., and Zhang, S. 2009. "Investigation of Customer Satisfaction in Student Food Service," *International Journal of Quality and Service Sciences*, (1:1), pp. 113–124.

Locker, D. 2000. "Response and Nonresponse Bias in Oral Health Surveys.," *Journal of Public Health Dentistry*, (60:2), pp. 72–81.

Ma, S.-C., Chien, T.-W., Wang, H.-H., Li, Y.-C., and Yui, M.-S. 2014. "Applying Computerized Adaptive Testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying.," *Journal of medical Internet research*, (16:2), pp. 1-14.

Manfreda, K. L. M., Batagelj, Z., and Vehovar, V. 2002. "Design of Websurvey Questionnaires: Three Basic Experiments," *Jounal of Computer-Mediated Communication*, (7:3), p. 0

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., and Schutz, a. 2007. "Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys," *Social Science Computer Review*, (25), pp. 372–383.

McHugh, M. L. 2011. "Multiple Comparison Analysis Testing in ANOVA," *Biochemia medica*, (21:3), pp. 203–9.

Merrell, C., and Tymms, P. 2007. "Identifying Reading Problems with Computer-Adaptive Assessments," *Journal of Computer Assisted Learning*, (23), pp. 27–35.

Montgomery, J. M., and Cutler, J. 2013. "Computerized Adaptive Testing for Public Opinion Surveys," *Political Analysis*, (21:2), pp. 141–171.

Payne, A., and Frow, P. 2013. *Strategic Customer Management Integrating Relationship Marketing and CRM*, Cambridge: Cambridge University Press.

Peterson, R. A., and Wilson, W. R. 1992. "Measuring Customer Satisfaction: Fact and Artifact," *Journal of the Academy of Marketing Science*, (20:1), pp. 61–71.

Peytchev, A. 2013. "Consequences of Survey Nonresponse," *The ANNALS of the American Academy of Political and Social Science*, (645:1), pp. 88–111.

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

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

Porter, S. R. 2004. "Pros and Cons of Paper and Electronic Surveys," *New Directions for Institutional Research*, (2004:121), pp. 91–97.

Porter, S. R., Whitcomb, M. E., and Weitzer, W. H. 2004. "Multiple Surveys of Students and Survey Fatigue," *New Directions for Institutional Research*, (2004:121), pp. 63–73.

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

Reeve, B. B., and Fayers, P. 2005. "Applying Item Response Theory Modelling for Evaluating Questionnaire Item and Scale Properties," *Assessing Quality of Life in Clinical Trial: Methods and Practice*, (2), pp. 55–73.

Rogelberg, S. G., and Stanton, J. M. 2007. "Introduction: Understanding and Dealing With Organizational Survey Nonresponse," *Organizational Research Methods*, (10:2), pp. 195–209.

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

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

Sabbaghan, S., Gardner, L., and Chua, C. E. H. 2017. "A Threshold for a Q-Sorting Methodology for Computer-Adaptive Surveys," in *ECIS*.

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

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

Schouten, B., Calinescu, M., and Luiten, A. 2013. "Optimizing Quality ofResponse through Adaptive Survey Designs," *Survey Methodology*, (39:1), pp. 29–58.

Shadish, W. R., Cook, T. D., and Campbell, D. T. 2002. "Experimental and Quasi-Experimental for Generalized Designs Causal Inference," Wadsworth Cengage learning.

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

Shin, C. D., Chien, Y., and Way, W. D. 2012. "A Comparison of Three Content Balancing Methods for Fixed and Variable Length Computerized Adaptive Tests.,"

Sinclair, M. D., Clark, J. R., and Dillman, D. A. 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), pp. 289–304.

Singer, E. 2012. "The Use of Incentives to Reduce Nonresponse in Household Surveys," *Survey Methodology Program*, p. 11.

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

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

Sutton, R. I., and Staw, B. M. 1995. "What Theory is Not," *Administrative Science Quarterly*, (40:3), pp. 371–384.

Thompson, N., and Weiss, D. 2011. "A Framework for the Development of Computerized AdaptiveTtests," *Practical Assessment, Research & Education*, (16:1), pp. 1–9.

Tourangeau, R., Couper, M. P., and Conrad, F. 2004. "Spacing , Position , and Order Interpretive Heuristics for Visual Features of Survey Questions," *Public Opinion Quarterly*, (68:3), pp. 368–393.

Wagner, J. 2008. "Adaptive Survey Design to Reduce Nonresponse Bias," *ProQuest*.

Wiley, J. B., Han, V., Albaum, G., and Thirkell, P. 2009. "Selecting Techniques for Use in an Internet Survey," *Asia Pacific Journal of Marketing and Logistics*, (21:4), pp. 455–474.

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

Woodruff, R. B. 1997. "Customer Value: the Next Source for Competitive Advantage," *Journal of the Academy of Marketing Science*, (25:2), pp. 139–153.

Zaller, J., and Feldman, S. 1992. "A Simple Theory of the Survey Response : Answering Questions versus Revealing Preferences," *American Journal of Political Science*, (36:3), pp. 579–616.

Zhang, C., and Conrad, F. G. 2013. "Speeding in Web Surveys : The Tendency to Answer Very Fast and its Association with Straightlining," *European Survey Research Association*, (8:2), pp. 127–135.

Zhang, Y. 2000. "Using the Internet for Survey Research: A Case Study," *Journal of the American Society for Information Science*, (51:1), pp. 57.