USE OF GENERIC INTERNET-BASED SURVEY PROGRAM FOR THE COLLECTION OF FOOD SENSORY DATA: A VALIDATION STUDY

by

SARA JOELLE SHERIDAN

(Under the Direction of Ruthann B. Swanson)

ABSTRACT

Internet-based sensory scorecards administered through Qualtrics (Qualtrics, Provo, UT), a generic survey software program, on iPads are a potential alternative to time-consuming and error-prone paper-based sensory data collection. Using a counter-balanced, repeated measures design, the concurrent validity (using paper scorecards as the criterion measure), usability, error rate, and time requirements of Qualtrics scorecards were assessed for 9-point structured (n=113) and unstructured (n=108) hedonic line scales, 15-point unstructured descriptive line scale (n=6), ranking test (n=107), triangle test (n=98), and tetrad test (n=108). Statistical analyses (paired t-tests, Friedman's Test, Wilcoxon Signed Rank tests) indicated that panelists' responses between paper and Qualtrics scorecards were not significantly different ($p > 0.05$). Panelists rated the usability of the Qualtrics scorecards favorably. Total time requirements were reduced by 68.8% with the use of the Qualtrics scorecard; however, the error rate was not reduced. These results indicate that Qualtrics scorecards are a valid sensory data collection method.

INDEX WORDS: Sensory evaluation, Affective sensory tests, Discrimination sensory tests, Descriptive sensory analysis, Electronic sensory data collection

USE OF GENERIC INTERNET-BASED SURVEY PROGRAM FOR THE COLLECTION OF FOOD SENSORY DATA: A VALIDATION STUDY

by

SARA JOELLE SHERIDAN

B.S.F.C.S., University of Georgia, 2013

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2015

USE OF GENERIC INTERNET-BASED SURVEY PROGRAM FOR THE COLLECTION OF FOOD SENSORY DATA: A VALIDATION STUDY

by

SARA JOELLE SHERIDAN

Major Professor:	Ruthann Swanson
Committee:	Barbara Grossman
	Judy Harrison

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

ACKNOWLEDGEMENTS

They say it takes a village to raise a child. I do not know anything about raising a child, but I can say that it takes a village to conduct a research project and write a thesis. This project would not have been possible without the help of many people.

Dr. Swanson, thank you for your support, patience (especially when I had my "deer in headlights" look), and expertise. You gave me the perfect combination of guidance and independence throughout this project, and as a result, I have become a more capable and confident student, researcher, and future professional. Dr. Grossman and Dr. Harrison, thank you both for your invaluable guidance and feedback over the past two years. Writing and defending a thesis is a daunting task, but you both provided advice and humor that made the process much less intimidating.

Laura Leigh Richardson and Sharon Hanson, I am so grateful for both your creativity and help over the past two years; you two worked out most of the technical kinks for this project before I could even think of them myself. It has been such a joy to get to know both of you.

Dorothy Dupree, you are a saint for giving up so much of your free time to help me prepare and run my sensory panels. Thank you, Rosemarie Rooker, Anna Cheely, Betsy Britt, and Emily Pleasance, for your advice and help over the past two years. I could not have asked for a better lab group.

Finally, thank you to my family, especially Mom, Dad, and Lara. You three have always been my biggest supporters. Mom, thank you for always reminding me to pray when I felt

overwhelmed and for your endless patience and understanding. Also, thank you for driving to Athens with 14 packages of oatmeal cookies for my panel after I had a meltdown in the Publix parking lot. Dad, thank you for your reminders to "stay the course" and for providing an example of dedicated work ethic and leadership. Lara, thank you for the countless movie nights and for being a "frequent flyer" at my sensory panels, even if it was just for the free SmartPop and ice cream.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1: INTRODUCTION

History and definition of sensory evaluation

Before the advent of systematic sensory evaluation in the 1940's, one individual, such as a brewmaster, wine taster, dairy judge, or food inspector, determined the presence or absence of desirable sensory characteristics (Kemp and others 2009; Lawless and Heymann 2010). Modern sensory evaluation emerged during World War II in an effort to develop food products acceptable to American soldiers (Pangborn 1981 as cited in Meilgaard and others 2007). Dove (1946) was one of the first to explore the relationship between the food industry and the acceptability of food products by consumers, calling acceptability the link that connected consumers with food. According to Dove, the importance of food acceptability was a result of the gradual shift towards commercial agriculture and quantity production of processed foods for the consumer market. Prior to this shift, families were self-sufficient. Family members determined quality food characteristics, and the preparation methods that produced these standards were passed to each subsequent generation. However, the growth of the commercial food market increased the awareness of food acceptability due to the shift from family-wide to consumer population-wide food acceptance.

Since the formal establishment of the need for the science of food acceptance by Dove, sensory evaluation has been developed into a formal methodology made up of different classes of sensory tests. The Institute of Food Technologists defines sensory evaluation as "a scientific method used to evoke, measure, analyze, and interpret those responses to products as perceived

through the senses of sight, smell, touch, taste, and hearing" (Anonymous 1975 as cited in Kemp and others 2009). Lawless and Heymann (2010) break this definition down into four parts: evoke, measure, analyze, and interpret. Evoke refers to the guidelines of sensory evaluation regarding the controlled conditions under which preparation and presentation of samples are executed. These conditions include sending subjects to individual testing booths, coding samples with random three-digit codes, presenting samples in a random order to prevent sequencing bias, and setting standards for testing room temperature, time between samples, and the number of samples presented. Measure refers to the quantitative methods used to determine relationships between sensory characteristics and human perception. In sensory evaluation, measurement methods are based on the principles used in behavioral research. Analysis involves the use of appropriate experimental design and statistical tests to draw conclusions from data collected from test subjects. Finally, interpretation refers to drawing conclusions from statistical analyses with consideration of the hypotheses and background literature.

Sensory evaluation is an important part of product development and can provide information on sensory attributes of a particular product category that influences consumer behavior including acceptance (Kemp and others 2009). It is important in decisions regarding the effects of ingredient substitution or modification, processing methods, and scaling up pilot samples to large-scale productions on sensory quality and/or product acceptability. Sensory testing can also be used in quality assurance procedures to determine shelf life and product variability in order to prevent sub-standard products from reaching the marketplace. Data from sensory testing can also be used to inspire marketing campaigns that focus on desirable sensory characteristics (Kemp and others 2009; Lawless and Heymann 2010).

Sensory attributes

Food sensory testing is concerned with appearance, odor and aroma, texture, flavor, and in some cases, sound. A sensory test may focus on a specific sensory attribute, or focus on multiple attributes. Because these attributes overlap during the consumption of food, most consumers will not be able to give an accurate evaluation of a single sensory attribute (Meilgaard and others 2007).

Appearance

Appearance is the first sensory attribute perceived by a consumer, and is often the basis for choosing to purchase or consume a food product (Meilgaard and others 2007). Appearance includes color, size/shape, surface texture, and clarity, all of which are common appearance characteristics consumers use to evaluate a product's overall appearance. Color includes both the perception of different wavelengths, as well as the evenness of color in a food. Size and shape refers to the length, thickness, width, particle size and geometric shape of a food. Surface texture helps a consumer anticipate whether a product will be wet, dry, crisp, tough, or soft. Clarity refers to the opacity of translucent solids or liquids.

Odor/Aroma

The "smell" of a food product includes both its odor and aroma. Odor is detected when volatiles of the food product enter the nasal passage and are sensed by the olfactory receptors on the cilia covering the nasal cavity (Kemp and others 2009). Aromatics are the volatiles that are perceived during the mastication of food (Meilgaard and others 2007). The temperature and texture of a food product, as well as enzymatic reactions, play roles in the release of volatiles from food. The release of volatiles increases exponentially with the increase of the temperature of a food; in addition, volatiles escape more easily from soft, porous foods than hard, dry foods.

In some instances, an enzymatic reaction must take place at a freshly cut surface in order for the volatiles to be percieved.

Touch/Texture

Touch includes somesthesis, kinesthesis, and chemesthesis (Kemp and others 2009). Somethesis refers to the detection of sensations by surfaces of the oral cavity related to touch such as force, particle size, and temperature. Kinesthesis involves nerve fibers in muscles, joints, and tendons that sense tension and relaxation, which allows for the perception of sensory attributes such as hardness. Chemethesis involves the stimulation of trigeminal nerves in the skin, mouth, and nose by chemical substances. The stimulation of these nerves results in hot, burning, cooling, tingling, or astringent sensations, all of which are considered mouth-feel attributes.

Texture is defined as "the sensory and functional manifestation of the structural, mechanical, and surface properties of foods detected through the senses of vision, hearing, touch, and kinesthetic" (Szczesniak 2002). It is separated into three properties: mechanical, geometrical, and moisture (Kemp and others 2009). Mechanical properties include hardness, cohesiveness, adhesiveness, denseness, and springiness. Geometrical properties refer to the perception of the size, shape, and orientation of particles in a food. These properties are described using words such as smoothness, grittiness, graininess, and fibrousness. Moisture properties refer to the perception of water, oil, and fat as measured by tactile means. These properties include descriptors such as moistness, or the amount of wetness or oiliness present, juiciness, oiliness, and greasiness.

Flavor

According to Amerine and others (1965) flavor is defined as "the sum of perceptions resulting from stimulation of the sense ends that are grouped together at the entrance of the alimentary and respiratory tracts." Thus, flavor includes the aromatics, the tastes, and the chemical feelings perceived in a food product.

Factors affecting sensory measurements

Humans are prone to context effects and bias, and therefore psychological and physiological factors must be controlled during sensory evaluation to eliminate or reduce these factors (Lawless and Heymann 2010). Context effects refer to instances in which the judgment of a product shifts due to factors such as other products that are tested during the same evaluation session. Biases are inaccurate reflections of a panelist's sensory experience resulting from observations or other influences on a panelist's responses.

Psychological factors

Expectation error, suggestion error, distraction error, stimulus error, logical error, proximity error, attribute dumping, order effect, contrast effect, central tendency error, and motivation error are all types of psychological factors that can affect panelists' responses during sensory evaluation (Kemp and others 2009; Meilgaard and others 2007).

Expectation error describes the phenomenon in which panelists' preconceived ideas of a product, formulated by knowledge of experiment objectives or other information, affect panelists' responses in sensory evaluation (Meilgaard and others 2007). Providing panelists with only the minimum amount of information needed to perform the test and coding samples prevents this type of error.

Suggestion error refers to instances where comments or noises made by a panelist affect the judgment of other panelists (Kemp and others 2009). Placing panelists in individual sensory booths and discouraging noises or discussion during evaluations prevents suggestion error. Panelists can also be distracted by noises inside and outside of the testing environment. This is referred to as distraction error and can be avoided by creating a testing environment that is quiet, comfortable, and encourages professional behavior.

Stimulus error is when a panelist knows or presumes to know the identity of the stimulus, and therefore bases his/her judgment on that stimulus (Lawless and Heymann 2010). Logical error refers to the tendency for panelists to form associations between two or more characteristics being evaluated in samples (Meilgaard and others 2007). Colored lights, nose clips, and masking sample characteristics and presentation consistent throughout an evaluation can help reduce these types of error.

Proximity error, or the halo effect, describes how the judgments of a sensory attribute in one product may affect the judgments of other products assessed in the same evaluation session (Kemp and others 2009). This type of error is more likely to occur with untrained consumer panelists, and thus the use of trained panelists can help eliminate the halo effect. Additionally, evaluating one sensory attribute at a time, or randomizing the order of attribute evaluations can help control for this error.

Dumping effect refers to instances where panelists are not given the opportunity to evaluate all of the sensory attributes they perceive in a product. This results in panelists "dumping" their observations on the scales that are available to them (Lawless and Heymann 2010; Kemp and others 2009). This can be prevented be allowing panelists to evaluate all varying attributes in a product, however there are always instances where a change in a specific

attribute was not observed in preliminary testing, or the attribute was not deemed relevant for consumer testing (Lawless and Heymann 2010).

Order effect describes how the assessment of a sample can be influenced by the sensory attributes of the sample preceding it (Kemp and others 2009). Contrast effect is a type of order effect, and refers to instances where samples that are noticeably different may be rated in an exaggerated manner. Central tendency error refers to both the tendency of panelists to rate samples using only the middle of the scale and the tendency for panelists to prefer samples presented in the middle of a set (Kemp and others 2009; Meilgaard and others 2007). Randomizing and balancing the order of sample presentation can avoid all of these types of order effects.

Finally, motivation plays an important role in producing reliable, consistent responses (Kemp and others 2009; Meilgaard and others 2007). It is important to create an atmosphere that is both comfortable and professional in order to instill motivation in panelists.

Physiological factors

Adaptation and enhancement are two physiological factors that can influence sensory evaluation. Adaptation occurs with continued exposure to a stimulus and results in a reduction in sensitivity to that stimulus, and potentially other stimuli (Kemp and others 2009). Adaptation in panelists can be prevented by limiting the number of samples presented, allowing an appropriate amount of recovery time between samples, and providing palate cleansers for panelists to consume after each sample. Enhancement can occur as a result of the interaction between various stimuli, resulting in an increased or decreased perception of one or more sensory characteristics (Meilgaard and others 2007). The interaction of stimuli can also result in the increased perception of both stimuli, or a synergy of the interacting stimuli. These effects can be

avoided by understanding how the various stimuli present in the samples interact with one another.

Rationale

Traditionally, sensory data are collected using paper-based data collection methods. However, despite its simplicity and low initial costs in equipment, training, and support (Wilcox and others 2012), paper-based data collection requires manual data entry that is time-consuming and subject to human error (Shapiro and others 2004; Fritz and others 2012). Potential alternative data collection methods include using tablets and Internet-based forms. Studies in clinical settings revealed that these methods are user-friendly, more time-efficient, and more accurate than paper-based methods (Wilcox and others 2012; Shapiro and others 2004). Research concerning the use of iPad and Internet-based data collection in food sensory analysis is limited, but computers have produced consistent panelist responses (Plemmons and Resurreccion 1999) using an intuitive evaluation system (Armstrong and others 1997). This project updates research regarding the use of electronic data collection in food sensory analysis through the innovative use of Internet-based scorecards delivered via iPads.

Research question

Are Internet-based sensory scorecards delivered on Qualtrics software via iPads a valid method of data collection and organization for food sensory analysis?

Objectives

The overall hypothesis was tested with both consumer and trained panels in order to validate all three classes of sensory tests (discrimination, descriptive, and affective). The study population included men and women from the UGA campus aged 18 and older. The objectives of this study were:

1. To assess the concurrent validity of Internet-based sensory scorecards using paper-based scorecards as the criterion measure. *Because of the flexibility and modifiability allowed in the development of Internet-based scorecards, it was hypothesized that responses generated by this method correlated and were consistent with responses generated by paper.*
2. To determine how panelists rate the usability of Internet-based scorecards as an evaluation instrument. *Because of consumer familiarity with Apple iOS on iPhones and iPods (Mintel 2013), it was hypothesized that panelists would report that iPads were easy to use as an evaluation instrument.*
3. To determine the error rate and time requirements of data collection and organization via Internet-based sensory scorecards when compared to paper-based scorecards. *Because Internet-based scorecards eliminate the need for manual data entry, it was hypothesized that this method would be more time-efficient and produce fewer data collection errors than paper-based scorecards.*

References

Amerine MA, Pangborn RM, Roessler EB. 1965. Principles of Sensory Evaluation of Food, New York: Academic Press. p. 549).

Anonymous. 1975. Minutes of division business meeting. Institute of Food Technologists – Sensory Evaluation Division, IFT, Chicago, IL. Cited in: Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell. p. 1.

Armstrong G, McIlveen H, McDowell D, Blair I. 1997. Sensory analysis and assessor motivation: can computers make a difference? Food Quality and Preference 8(1):1-7.

Dove WE. 1946. Developing food acceptance research. Science 103: 187.

Fritz F, Balhourn S, Riek M, Breil B, Dugas M. 2012. Qualitative and quantitative evaluation of HER-integrated mobile patient questionnaires regarding usability and cost-efficiency. International Journal of Healthcare Informatics 81(5):303-13.

Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell.

Lawless HT, Heymann H. 2010. Sensory Evaluation of Food. New York, NY: Springer.

Lim J. 2011. Hedonic scaling: A review of methods and theory. Food Quality and Preference 22(8): 733-747.

Meilgaard MC, Civille GV, Carr BT. 2007. Sensory Evaluation Techniques, 4th ed. Boca Raton, FL: CRC Press.

Mintel. 2013. Tablets – Executive summary. Retrieved January 28, 2014 from Mintel Academic: http://academic.mintel.com.proxy-remote.galib.uga.edu/display/637690/.

Muñoz AM and Civille GV. 1998. Universal, product, and attribute specific scaling and the development of common lexicons in descriptive analysis. Journal of Sensory Studies 13(1): 57-75.

Pangborn RM. 1964. Sensory evaluation of food: A look backward and forward. Food Technology 18:1309. Cited in Meilgaard MC, Civille GV, Carr BT. 2006. Sensory Evaluation Techniques. p. 1.

Plemmons LE and Resurreccion AVA. 1999. Note on computerized data collection in consumer sensory evaluation. Journal of Sensory Studies14(1):19-34.

Shapiro JS, Bessette MJ, Baumlin KM, Ragin DF, Richardson LD. 2004. Automating research data collection. Academic Emergency Medicine 11(11):1223-28.

Szczesniak AS. 2002. Texture is a sensory property. Food Quality and Preference 13(4): 215-225.

Wilcox AB, Gallagher KD, Boden-Albala B, Bakken SR. 2012. Research data collection methods from paper to tablet computers. Medical Care 50(7):S68-73.

CHAPTER 2: REVIEW OF LITERATURE

Traditionally, paper has been the preferred method for data collection due to its ease in implementation (Wilcox and others 2012). Despite data collection's pivotal role in research, it is subject to human error, such as incorrect data entry and data loss (Shaprio and others 2004). New research has focused on the benefits of using mobile computing devices, like the iPad. Tablets are portable, enable the merging of electronic data from different sources, and are user-friendly (Wilcox and others 2012), and thus can reduce the use of paper during the data collection and analysis steps of research, including sensory evaluation.

Concurrent validity

Validity refers to a research instrument's ability to accurately measure what it is supposed to measure (Talmage and Rasher 1981). There are four types of validity: content, face, criterion, and construct. This project focused on criterion validity, or the capability to extrapolate information about an independent variable, or the criterion measure, from the data extracted from the dependent variable. In other words, criterion validity is the strength of the relationship between the original instrument or method and the new instrument or method. Criterion validity is either predictive or concurrent. Predictive validity measures how accurately an instrument predicts results or future behavior. Concurrent validity measures how a new instrument estimates the performance of current instrumentation. If a new instrument effectively

and accurately achieves this, then it has concurrent validity with the current instrument being used.

Validity of electronic data collection methods in research

Electronic data collection methods have been shown to generate responses that are consistent with their paper-based counterparts, especially as the use of computers, tablets, and other mobile, hand-held devices has become more widespread. Two older sensory analysis studies by Armstrong and others (1997) and Plemmons and Resurrection (1999) have indicated promising results.

Armstrong and others (1997) compared paper and computerized scorecards for various types of tests from all three classes of sensory evaluation tests, including triangle, ranking acceptance, and hedonic line scale tests. 12 panelists between the ages of 25-45 were selected through a screening exercise involving the completion of twenty discrimination tests over one month. Panelists were split into two groups; one group completed ten sensory tests using paper scorecards and the other group used the computerized scorecards, which were administered via PSA System 3 (version 20.5). The next day, the groups were reversed and sensory tests repeated.

Out of the various discrimination tests assessed, the ones requiring fewer samples, such as the triangle test, paired comparison test, and R-Index, had the greatest degree of accuracy (choosing the correct sample), although the percentage of accurate responses was lower for the computer responses in all three of these tests (Armstrong and others 1997). These tests also had the greatest consistency (the number of times when both methods produced the same response by each tester) in responses between the two scorecard methods (50%, 75%, and 71% for triangle,

paired comparison, and R-Index tests, respectively). Conversely, the ranking discrimination test had six samples with a low degree of accuracy and only 29% consistent responses between the paper and computerized scorecard. The researchers noted that number of samples, choice of samples, and training of panelists were all factors that affected the accuracy and consistency in responses between the two scorecard versions.

Consistency was generally low for the hedonic tests assessing appearance, aroma, texture, and flavor acceptability in beef products on a 15-point line scale, however the standard deviation for computer responses was lower than paper responses (Armstrong and others 1997). Although consistency in responses was low between the paper and computerized hedonic tests, the researchers argued that this was more acceptable in acceptance and preference tests since they are subjective measures. Consistency in responses for affective ranking tests was similarly varied. Individual responses for the acceptance and preference ranking tests varied in consistency (40% for rank acceptance and 75% for rank preference); however, the totaled rank responses corresponded closely.

Plemmons and Resurreccion (1999) evaluated the uniformity of results from computer (Compusense 5) and paper scorecards (potato chips, cereal, and canned okra) using a 9-point hedonic scale. 26 panelists with computer experience and 24 panelists without computer experience were separated into two groups in a crossover design. One group assessed appearance, flavor, and overall acceptability in the samples using the paper scorecards, and the other used the computer scorecards. After a 20-minute break, panelists repeated the evaluations with the alternate data collection method. Both groups viewed an instructional video explaining the procedures for the computerized sensory evaluation.

A paired samples t-test showed that there were no significant differences ($P > 0.05$) in appearance, flavor and overall acceptability for potato chips and okra between the two data collection methods. There was a significant difference in flavor and overall acceptability, but not in appearance, for cereal between the paper and computerized scorecard.

Swaney-Stueve and Heymann (2002) compared paper scorecards to computer scorecards administered via Compusense 5 (version 3.8) software to determine if scorecard type significantly affected experimental results, and to determine if paper scorecards could be substituted for computer scorecards in the event of computer failure. 12 trained panelists evaluated 34 sensory attributes in 12 brands of peanut butter. Panelists performed the experiment twice in order to use both scorecard versions. The amount of time between paper and computer sessions was not given. No significant scorecard effect was found in 22 of the 34 sensory attributes ($p > 0.05$), indicating that there was no significant difference in the way the attributes were scored. The scorecard*sample interaction was not significant for any attribute ($p > 0.05$), which demonstrated that the panelists scored the samples relatively similarly using both scorecard versions. Panelists tended to rate attributes higher when using the paper scorecards. Mann Whitney tests indicated that the replacement of any computer scorecards with paper scorecards would not significantly change the results of the experiment.

Usability of electronic data collection methods

In a study by Armstrong and others (1997), participants, including those with minimal computer experience, reported feeling comfortable using computers to record their responses for various food sensory analysis tests. Findlay and others (1986) found that panelists quickly adapted to the use of a computerized sensory analysis system (Compusense) in which a stylus-

type pen was used to indicate attribute intensity on an unstructured line scale, as well as indicate extent of liking on a 7-point structured hedonic scale. Participants in a study by Chuang and others (2008) reported that the paper-based survey and three different computer-based surveys were all relatively easy to use, rating two out of the three computer-based surveys higher than the paper-based survey.

Many clinical studies have reported that both participants preferred using electronic data collection methods to paper methods. Jeannic and others (2014) compared paper and electronic case report forms and found that participants were satisfied with both versions of the case report forms. When the participants were asked which version they preferred, the majority of participants indicated their preference for the electronic case report forms. 82% of participants in the study by Bliven and others (2001) preferred the electronic versions of the SAQ questionnaire and RAND-36 survey, which required the use of a touchscreen personal computer. Additionally, 89% found the electronic versions "very easy" or "easy" to use, and none rated the electronic versions as "very difficult" to use. Similarly, in a review by Lane and others (2006), 3 out of the 4 articles that evaluated patient preference for paper or electronic survey methods found that out of 91 patients, 54 preferred using portable handheld devices, while 19 preferred paper and pencil (20 patients had no preference).

A review by Shapiro and others (2004) found that Internet-based forms are easy to modify and can be formatted to prevent incomplete and incorrect answers. Survey administrators reported that iPads required little training and technical support compared to other electronic tools (Wilcox and others 2012). Administrators also noted that iPads were comparable to paper in regards to portability.

Error rate of electronic data collection methods

Manual data entry, which is a significant source of human error in paper-based data collection, is eliminated in electronic data collection (Shapiro and others 2004). Two studies compared the error rates of electronic and paper data collection, which included both incorrect or incomplete entries and errors made during data transfer (Fritz and others 2012; Chuang and others 2008). Fritz and others (2012) found fewer errors associated with electronic iPad-based patient questionnaires compared to the paper versions, partially due to the capability of the survey software to prevent panelists from accidentally skipping questions. There were also no errors associated with data transfer using the electronic method, however 17 out of 100 paper questionnaires were found to have missing or double answered questions. Similarly, there were no missing values for the computer-based methods in the study by Chuang and others (2008), but there were 23 (0.4%) missing values in 4 (16.7%) participants' paper questionnaires.

Time requirements of electronic data collection methods

Scorecard completion time and time required for the processing of data are both important considerations when weighing the advantages and disadvantages of paper and electronic data collection. On average, participants in the study by Fritz and others (2012) completed the three paper questionnaires faster than the iPad counterparts (8.22 minutes and 17.18 minutes, respectively), however the standard deviation for the iPad questionnaire completion time was 9.22, while the paper standard deviation was 2.55 minutes. According to the researchers, this suggests that outliers affected the mean completion time for the iPad questionnaires. Additionally, participants reported feeling no difference in the length of time it took them to complete the two versions of the questionnaires.

Similarly, participants in the study by Chuang and others (2008) needed the least amount of time to complete the paper questionnaire with a mean completion time of 15.83 minutes, compared to 17.91 minutes, 17.84 minutes, and 17.37 minutes for the three computer versions. However, the differences in completion time for the paper version and the three computer versions were not significantly different ($p > 0.05$).

Although participants in studies by Fritz and others (2012) and Chuang and others (2008) completed paper-based questionnaires faster than iPad and computer-based questionnaires, the time required to process the paper-based data ultimately made the electronic methods more efficient. Fritz and others (2012) reported an estimated 1.49 minutes per patient required to input and process data from clinical questionnaires, adding up to a total 61.5 hours per year for 2020 patients. In comparison, the questionnaires administered via iPads required no data input or processing. Chuang and others (2008) estimated 9.52 minutes per participant were required to process data from the paper questionnaires, while the computer-based questionnaires required no further data processing. In three out of nine articles reviewed by Lane and others (2006) electronic data collection methods reduce were found to data entry time, although only one article provided data to support that finding, reporting a reduction of 23% in data entry and transfer time.

Although the use of Qualtrics scorecards may decrease the time required for data organization and analysis in this study, an increase in the time required for the panelist to complete these forms may increase the time required to conduct sensory panels overall. This could result in additional material and labor costs, particularly if a panel is extended in order to accommodate the extra time needed to complete electronic scorecards.

Cost efficiency of electronic data collection methods

Despite high upfront costs, electronic data collection methods appear to be cost-effective over time due to reduced labor costs. Fritz and others (2012) compared the cost effectiveness of paper questionnaires and online questionnaires administered via iPads. To compute the costs associated with paper questionnaires, the prices of paper and staples were multiplied by the estimated number of patients seen per year and then added with the salaries of three different kinds of employees (student assistant, scientific assistant, medical documentation specialist) potentially responsible for inputting data. The costs of iPads were determined by adding the retail price of three iPads, iSkin cases, screen protectors, styluses, and depreciation costs for a lifespan of three years. Estimated costs for paper-based materials and associated personnel ranged from $1,093 - $2,187 per year depending on the type of personnel employed for data entry. Costs related to iPad-based data collection were estimated to be $944 per year due to the elimination of 61.5 hours of labor hours for data entry.

Potential issues associated with electronic data collection

Potential problems observed with the use of iPads mainly pertained to security of the data collected, stored, and transferred on the iPad, as well as the iPad itself. A study in the review by Wilcox and others (2012) addressed information security issues by encrypting the collected data before it was stored in the cloud server, then decrypting the data when downloading it at the research institution. In this study, connectivity problems were also avoided by using a local application on the iPad to implement surveys as opposed to connecting to the Internet to use a web-based survey. This method reduced productivity issues related to problems with connecting to the Internet, which Fritz and others (2012) experienced in their study. Wilcox and others

(2012) observed that system crashes and lengthy loading times created workflow issues; however, a second Wi-Fi access point easily resolved both issues.

Neither study addressed methods of preventing theft of the iPad, which could be a risk in a high volume sensory panel where panelists are hidden from view in individual sensory booths. However, the use of wall mounts with lockable cases to secure iPads to the walls of the sensory booths should eliminate this risk. This also will prevent food debris and liquid from damaging the iPads. Additionally, protection of the iPad screens with screen covers will allow the screens to be wiped clean.

Qualtrics

Qualtrics is a private research software company that specializes in online surveys and data collection and analysis. Qualtrics Research Suite (Provo, Utah) is an online survey platform that allows users to build, design, and implement online surveys; in addition, Qualtrics can organize and analyze data collected with the surveys, or collected data can be exported to other statistical software such as SPSS or SAS. Research Suite specializes in consumer, market, and academic insights and offers over 100 different types of survey questions, the ability to branch survey questions, as well as edit the display logic and randomize the order of survey items. Qualtrics also offers mobile and offline survey capabilities, giving this program an advantage over survey programs that are limited to desktop use, as well as programs that require an Internet connection.

Classes of sensory tests

Sensory tests are either objective or subjective, and fall into one of three classes: discrimination, descriptive, and affective tests. Objective tests include discrimination tests and descriptive tests. These tests provide objective information on the sensory properties of food products. Subjective tests include affective tests and are only used with consumer panels. These tests provide subjective data on acceptability, liking, or preference of food products. Each class of sensory test includes a number of different types of sensory tests. This project evaluated tests representative of each class of sensory test using both paper scorecards and Qualtrics scorecards administered via iPad. Triangle and tetrad tests represented discrimination tests, 15-point unstructured line scales as part of the Spectrum-like Descriptive Analysis Method™ represented descriptive tests, and ranking preference tests, as well as 9-point structured and unstructured hedonic scales represented affective tests. Each of these sensory tests and scales are described in the sections that follow.

Discrimination tests

Discrimination tests are employed to determine if there is a discernable difference between two or more products, or to see if two products are similar enough to be used interchangeably (Meilgaard and others 2007). Discrimination tests are often used to screen or train panelists for trained descriptive panels, but also can be used with untrained consumer panelists. In discrimination tests that assess whether the difference between samples is perceivable, the objective is to reject the null hypothesis (Lawless and Heymann 2010). Conversely, in discrimination tests that assess whether samples are similar enough to be interchangeable, the objective is to accept the null hypothesis.

There are two response formats that can be used during a discrimination test: forced choice and no difference. The forced choice format requires that panelists select a sample in response to the question, while the no difference format permits the panelists to respond that they find no difference in the samples. As described by Kemp and others (2009), discrimination tests are quick and should be used with samples that have unobvious differences. Along with their rapid execution, discrimination tests are also intuitive for panelists. The biggest limitation of discrimination tests is their inability to give information about the extent of the difference(s) between the samples; rather the results show only that a difference exists. In addition, discrimination tests cannot tell which of the samples is preferred.

Discrimination tests are divided into two categories: overall difference tests and attribute difference tests (Kemp and others 2009). Overall difference tests show sensory differences between samples, if any exist, using all available information. Attribute difference tests focus on a single attribute and ask panelists to judge how that single attribute differs between samples (Meilgaard and others 2007). These tests can either be one-sided or two-sided. The objective of a one-sided attribute difference test is to confirm an increase or decrease in intensity of a specific attribute, whereas a two-sided attribute difference test is interested in either outcome.

Discrimination tests: Triangle test

A triangle test is a type of overall difference test, meaning it determines if an overall difference exists between samples (Kemp and others 2009) (Appendix D). The objective of a triangle test is to determine if there is a difference between two samples. Panelists are presented three coded samples, one different and two the same, and asked to choose the sample that is different. The samples are given different codes, and presented simultaneously. The number of panelists needed for a triangle test depends on the chosen significance level, and can be

determined using the Critical Number of Correct Responses in a Triangle Test (Meilgaard and others 2007). Data from triangle tests are analyzed by counting the number of accurate responses identifying the odd sample, as well as the total number of responses; "no difference" replies should not be counted (Meilgaard and others 2007). The correct responses are then compared to the appropriate cut-off point using the chosen significant level on the Critical Values Table for Sensory Differences (Meilgaard and others 2007).

Triangle tests are most effective when used to assess if an overall difference exists in a product resulting from changes in ingredients, processing, packaging, or storage, or to screen panelists according to their ability to discriminate differences in products (Meilgaard and others 2007). The major drawback of triangle tests is their inability to give any information other than whether or not a difference exists. Triangle tests also tend to induce sensory fatigue.

Discrimination tests: Tetrad test

Tetrad tests are a relatively new type of discrimination test and are thought to be a more powerful alternative to the triangle test (Ennis and Christensen 2014). During a tetrad test, panelists are presented four samples, comprised of duplicates of the two samples being evaluated (making two groups), and are asked to group the samples into two groups based on similarity (Rousseau and Ennis 2013) (Appendix E). Similar to triangle tests, data from tetrad tests are analyzed by counting the number of correct responses and comparing the resultant totals to the appropriate cut-off point using the chosen significant level on the Critical Number of Correct Responses in a Triangle Test.

Both triangle and tetrad tests have a guessing probability of 1/3, but tetrad tests have been found to be more powerful than triangle tests by a factor of 3 (Ennis 2012). However, it has

been suggested that the additional stimuli involved in tetrad testing may pose a risk for additional sensory fatigue.

Descriptive tests

Descriptive tests allow the quantitative and qualitative characterization of the sensory properties of a food product (Meilgaard and others 2007). Sensory judges quantify the qualitative aspects of a food product, which include all aroma, appearance, flavor, texure, aftertaste, and sound properties, and thus develop a description of the product's perceived attributes (Murray and others 2001). Because of this, descriptive panels are useful in determining sensory characteristics during the development and marketing of new food products (Kemp and others 2009). Along with giving full sensory descriptions of food products, descriptive tests can also help identify ingredient and process variables, and the sensory attributes of a product that affect acceptance (Lawless and Heymann 2010). Unlike discrimination tests, descriptive tests can quantify how much one product differs from another. Data from descriptive analysis can be related to both instrumental and consumer preference or acceptability measurements, which can help determine driving factors in product acceptability (Murray and others 2001; Kemp and others 2009).

Descriptive tests require 6-18 panelists, which is fewer than the number required for discrimination testing, however descriptive panelists must be trained to identify and quantify specific sensory attributes in food products (Kemp and others 2009). The qualitative aspect of descriptive analysis requires trained panelists to identify and describe the sensory characteristics they perceive in a food product, while the quantitative aspect requires panelists to differentiate and quantify the intensities of various sensory attributes in a food product (Meilgaard and others 2007). Additionally, the sensory characteristics agreed upon by trained panelists can also be

used for quality control purposes and to understand the shelf life of a product. The data collected from descriptive tests can be used to determine relevant sensory attributes to be included in consumer tests, as well as to help explain the results of consumer tests.

Descriptive tests: Spectrum™ Descriptive Analysis Method

Gail Civille created the Sensory Spectrum Method in the 1970's using many of the principles of the Texture Profile Method (Lawless and Heymann 2010). Qualitative and quantitative data can be generated from this method (Kemp and others 2009). Instead of generating their own terms to describe the sensory properties of a product, panelists choose from a standardized group of terms, which are used for other products in the same category (Lawless and Heymann 2010). This allows the comparison of data from different studies and allows the results to be applied to a general consumer population. Panelists are trained to use scales anchored with multiple reference points in an identical manner.

Panelist training is extensive and involves trying multiple products in the same product category and choosing terms from "lexicons" to describe the perceived sensory sensations associated with the product. For product assessment, panelists use 15-point numbered universal attribute intensity scales (Muñoz and Civille 1998). Once the panelists agree on the attributes of the product and the order of evaluation, each product is assessed individually (Kemp and others 2009). Reference points are used on the scales in order to "calibrate" the panelists (Lawless and Heymann 2010). The main disadvantage of the Spectrum Method™ is the length of time and effort associated with the training period. Additionally, it may not be possible to achieve absolute panelist calibration for certain attributes, such as bitterness and musk odors (Kemp and others 2009). A short version of the Spectrum Method™ may be employed when the research question focuses on specific attributes of the product (Meilgaard and others 2007) (Appendix F).

Affective tests

Affective tests gauge the preferences, attitudes, opinions and perspectives of consumers to a product, or specific product characteristics using qualitative and quantitative methods (Meilgaard and others 2007; Kemp and others 2009). These tests provide invaluable information that is used in product development, sensory-related brand strategy, and product advertising (Kemp and others 2009). Typically, 100 or more panelists are recruited for affective tests. Unlike descriptive tests, affective tests do not ask panelists to assess the intensity of specific sensory attributes; instead, affective tests determine whether a panelist prefers, likes, or finds a product acceptable. Affective tests can also determine liking, preference, or acceptability of specific sensory attributes. It is inappropriate to ask panelists completing descriptive test to provide affective information.

There are qualitative and quantitative affective tests. Qualitative affective tests measure consumer response to the sensory properties of products through discussion in interviews or small groups (Meilgaard and others 2007). These tests give researchers information about consumer needs, initial responses, terminology, and behavior regarding specific products. Quantitative affective tests are used to determine overall preference, liking, or acceptance of either one or more products overall, or specific sensory characteristics (Kemp and others 2009). There are two types: preference and acceptance. Preference tests are used to determine whether consumer panelists prefer one product versus another. Acceptance tests do not involve a choice between different products and can be used with one item (Lawless and Heymann 2010). Unlike preference tests, acceptance tests gather information about the sensory appeal of a product.

Affective tests: Ranking tests

Ranking tests ask consumer panelists to rank three or more products in descending or ascending order of preference, with no options for ties in the ranking (Lawless and Heymann 2010) (Appendix C). These tests indicate whether one product is preferred over another (Kemp and others 2009), although usually the focus is one or two samples from the set (Larmond 1977). Ranking can be used for attribute intensity or preference. 16 or more trained panelists should be used for greatest statistical power and if possible, the set of samples should be presented two or more times when ranking for attribute intensity (Meilgaard and others 2007). Samples should be presented in a random order, with each presentation order presented equally, and assessed in the order given. Panelists are asked to rank samples based on the specified attribute or for overall preference, using palate cleansers after each sample. Preference ranking is simple and intuitive for consumers; however ranking tests involving product sampling can be tiring for panelists.

Affective tests: Hedonic scales

Hedonic scales are used in affective tests with untrained consumer panelists. They are used to ascertain the level of liking in one or more products (Kemp and others 2009). Typically, a 9-point bipolar, balanced scale around a neutral center with four positive and four negative phrases on each side is used (Lim 2011). The phrases express degrees of effect and are arranged from most negative to most positive, forming a single continuum and enabling researchers to determine the mean value of responses in terms of liking or disliking. This method is simple and can be used for a variety of products with minimal instruction. Hedonic scales can be either structured and unstructured line scales (Appendices A and B). However, the lack of zero point on hedonic scales and the inequality of the scale intervals prevent this scaling method from

providing information about ratios of liking or disliking for products or attributes. Additionally, comparisons between individuals and groups cannot be made.

Measurement of sensory responses

Traditionally, paper has been the preferred method for data collection due to its ease in training, implementation, and support (Wilcox and others 2012). Consent forms, demographic questionnaires, food liking questionnaires, and sensory scorecards are all printed on paper and completed by panelists, and then data are manually compiled, coded, and entered into statistical programs. The three most commonly used methods of measuring sensory data are: classification, ranking, and scaling. These methods are used across all three classes of sensory tests. This project utilized ranking and scaling methods, which are described below.

Ranking

Ranking involves asking panelists to arrange three or more samples in order of intensity of an attribute or preference (Meilgaard and others 2007). This type of test is quick and requires little training, however panelists should be familiarized with the attributes being evaluated (Meilgaard and others 2007). Furthermore, ranking tests are not as effective as scaling techniques when more than three samples are evaluated.

Scaling

Scaling tests use numbers or words to describe the intensity of a sensory attribute or liking of one or more samples (Meilgaard and others 2007). Numbers can be attributed to the words in order to analyze data collected from the tests. Scaling tests can be used in descriptive and affective tests. When used in descriptive tests, the attributes used as reference points to express the intensity of an attribute in a sample can be based on one of three philosophies:

universal scaling, product-specific scaling, and attribute-specific scaling. All three of these philosophies require that panelists be trained to focus on specific attributes and indicate their perceptions using a standardized lexicon in order to establish a common frame of reference (Muñoz and Civille 1998).

With universal scaling, attribute intensities are based on a universal frame of reference that considers all product categories (Muñoz and Civille 1998). All attributes are compared to absolute intensity references (not to each other), which results in absolute intensity differences among the attributes. This scaling method is useful when comparing products in different product categories, or comparing similar products in the same product category. Product specific scaling is based on the idea that attribute intensities should be limited to the product category being assessed; in other words, a separate frame of reference is created for each product category. Because each product category has its own frame of reference, ratings using this type of scaling method are not absolute, and therefore samples in different product categories cannot be compared. With attribute specific scaling, each attribute in a product is rated separately using scales and intensity references specific to each attribute. Therefore, single attributes can be rated on a single scale across products. Like product specific scaling, this means that various attribute intensities in a product or across products cannot be compared, nor does it give any information about how truly strong or weak the intensity of an attribute is in a product.

The three main types of scales used with these scaling methodologies are category scales, line scales, and magnitude estimation scales (Meilgaard and others 2007). These types of scales can be used in both descriptive and affective sensory tests. This project utilized category and line scales. Category scales involve rating the intensity of a sensory attribute by labeling it with a value, or category, on a numerical scale. These scales do not measure how much the degree of

the intensity of an attribute in one product differs from another. Typically, 10-15 point category scales are used. Line scales ask panelists to express the intensity of a sensory attribute by making a mark on a horizontal line that equates to the amount perceived in the sample (Figure 2.2). Line scales can also be used to express liking, preference, and acceptability in affective tests (Figure 2.5). Line scales have anchors on either end (left end = 0 or none, right end = very strong intensity or liking), and can either be structured or unstructured. Unstructured line scales involve only the two end anchors, whereas structured scales involve additional anchors between the two end anchors. The marks made by panelists can be converted into numerical values by measuring the position of each mark on the line.

References

Armstrong G, McIlveen H, McDowell D, Blair I. 1997. Sensory analysis and assessor motivation: can computers make a difference? Food Quality and Preference 8(1):1-7.

Bliven BD, Kaufman SE, Spertus JA. 2001. Electronic collection of health-related quality of life data: Validity, time benefits and patient preference. Quality of Life Research 10(1):15-21.

Chuang Y, Chen L, Chuang M. 2008. Computer-based rating method for evaluating multiple stimuli on multiple scales. Computers in Human Behavior 24(5):1929-1946.

Ennis JM. 2012. 2012. Guiding the switch from triangle testing to tetrad testing. Journal of Sensory Studies 27(4):223-231.

Ennis JM and Christensen RHB. 2014. Precision of measurement in Tetrad testing. Food Quality and Preference 32(SI):98-106.

Findlay C. 1986. Computers and the Internet in sensory quality control. Food Quality and Preference 13(6):423-428.

Fritz F, Balhourn S, Riek M, Breil B, Dugas M. 2012. Qualitative and quantitative evaluation of HER-integrated mobile patient questionnaires regarding usability and cost-efficiency. International Journal of Healthcare Informatics 81(5):303-13.

Jeannic A, Quelen C, Alberti C, Durand-Zaleski I. 2014. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Medical Research Methodology 14(1):1-24.

Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell.

Lane SJ, Heddle NM, Arnold E, Walker I. 2006. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods of data collection. BMC Medical Informatics and Decision Making 6(23):23-33.

Larmond E. 1977. Methods for Sensory Testing: Preference Tests. In: Laboratory Methods for Sensory Evaluation of Food. Winnipeg, Canada: Kromar Printing Ltd. p. 55-59.

Lawless HT, Heymann H. 2010. Sensory Evaluation of Food. New York, NY: Springer.

Lim J. 2011. Hedonic scaling: A review of methods and theory. Food Quality and Preference 22(8):733-747.

Meilgaard MC, Civille GV, Carr BT. 2007. Sensory Evaluation Techniques. Boca Raton, FL: CRC Press.

Muñoz AM and Civille GV. 1998. Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. Journal of Sensory Studies 13(1):57-75.

Murray JM, Delahunty CM, and Baxter IA. 2001. Descriptive sensory analysis: Past, present, and future. Food Research International 34(6):461-471.

Plemmons LE and Resurreccion AVA. 1999. Note on computerized data collection in consumer sensory evaluation. Journal of Sensory Studies 14(1):19-34.

Rousseau B and Ennis JM. 2013. Importance of correct instructions in the tetrad test. Journal of Sensory Studies 28(4): 264-269.

Qualtrics. 2015. Qualtrics Research Suite. Retrieved June 13, 2014 from Qualtrics on the World Wide Web: http://www.qualtrics.com/research-suite/.

Shapiro JS, Bessette MJ, Baumlin KM, Ragin DF, Richardson LD. 2004. Automating research data collection. Academic Emergency Medicine 11(11):1223-28.

Swaney-Stueve M and Heymann H. 2002. A comparison between paper and computerized ballots and a study of simulated substitution between the two ballots used in descriptive analysis. Journal of Sensory Studies 17(6): 527-537.

Talmage H and Rasher SP. 1981. Validity and reliability issues in measurement instrumentation. Journal of Nutrition Education 13(3):83-84.

Wilcox AB, Gallagher KD, Boden-Albala B, Bakken SR. 2012. Research data collection methods from paper to tablet computers. Medical Care 50(7):S68-73.

CHAPTER 3: MATERIALS AND METHODS

General experimental design

Traditional paper and electronic scorecards created with Qualtrics Research Suite (Qulatrics, Provo UT) and administered on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation) were compared using a counterbalanced, repeated measures design. Six different sensory tests were conducted and each class of sensory test was represented. 9-point structured and unstructured hedonic line scales and a ranking test represented affective sensory tests. Triangle and tetrad tests represented discrimination tests. Descriptive sensory tests were conducted on 15-point unstructured line scales using predefined and standardized lexicons as described by the Spectrum Descriptive Analysis Method™ (Meilgaard and others 2007); trained panelists who participated in The Evaluation of Almonds Study (Cheely 2014) completed the Spectrum-like Descriptive Analysis Method™.

In accordance with a repeated measures design, each sensory test was conducted twice with the same panelists, once using paper scorecards and questionnaires and once using Qualtrics scorecards and questionnaires. Panelists who participated in the first session received extra credit in specific classes as well as a snack of their choice. Panelists were asked via email to return the following week to participate in the second session for which they received additional extra credit points and a snack of their choice. All sensory data from the individual sensory panels were collected over one day, except for the descriptive sensory data, which were collected

over two days. See Table 3.1 for the order of scorecard formats for each set of sensory tests. Panelists were allowed to participate once per set of sensory evaluation sessions.

All panelists across all sensory tests evaluated samples in individual sensory booths under white lighting. No direct interaction occurred between panelists or between researchers and the panelists. All samples were served on white trays with room temperature water in 9 ounce clear plastic cups (Sam's West, Inc., Bentonville, AR), a palate cleanser (baby carrots, unsalted saltine crackers, or seltzer water), and a napkin (Dixie Consumer Products, Atlanta, GA). All samples were given a random, three-digit code and shown in a balanced order of presentation.

During each individual panel, regardless of scorecard and questionnaire version, panelists were presented with a consent form (Appendices Q and R), one to three scorecards depending on the sensory test (Appendices A-F, I-N), a usability questionnaire (Appendices G and O), and a demographics questionnaire (Appendices H and P). The usability questionnaire asked panelists whether they agreed or disagreed with the statement, "The instructions on this scorecard are easy to understand and follow" on a 5-point Likert scale, where 1=Strongly agree and 5=Strongly disagree. This questionnaire also asked panelists to answer the question, "How comfortable did you feel when completing this scorecard?" also using a 5-point Likert scale, where 1=Very comfortable and 5=Very uncomfortable. The Qualtrics version of the questionnaire utilized the Likert, Single Answer Matrix Table question type. 5 scale points were used to represent the five possible answer choices. No labels were used and the Force Response option was enabled. The demographics questionnaire asked for panelists' age, sex, and how often they consumed the food product evaluated during the sensory panel. The Qualtrics version of the demographics questionnaire included three Single Answer Multiple Choice type questions. The number of

answer choices varied for each question. The answer choices were positioned vertically to match the paper scorecard. The Force Response option was enabled.

Paper scorecards

Panelists were asked to write their first and last names on each paper scorecard and questionnaire in order to match their responses across both sessions for each set of sensory tests. After returning the signed consent form, panelists were presented with samples. Three paper scorecards, one for each sample, were presented one at a time with their respective samples for 9-point structured and unstructured hedonic line scales. One paper scorecard was presented along with all samples for ranking, triangle, and tetrad tests. Panelists completed the usability and demographic questionnaires, modacially, after the samples were evaluated. The return of the demographic questionnaire signaled that panelists had completed the sensory evaluation.

Qualtrics scorecards

Panelists were first given instructions to sign a consent form, and then asked to type in their first and last name in a text entry box using their finger or a stylus provided in each booth. Additionally, panelists were asked to enter their booth number for blocking purposes (Figures 3.1 and 3.2). Panelists were then given instructions specific for each type of sensory test on a separate screen. The following screen(s) contained the evaluation portion of the scorecard. For the 9-point structured and unstructured line scales, panelists were asked to return their sample tray after marking four line scales for one sample, and then asked to press "Next" to access the next set of line scales for the following sample as the next sample tray was sent through the hatch. For the ranking, triangle, and tetrad tests, panelists were presented all samples at once and panelists were asked to complete the evaluation and then return the sample tray when finished. Panelists then completed the usability and demographics questionnaire. Panelists were asked to

signal their completion of the entire sensory evaluation by sliding a laminated piece of paper with the typed phrase “Booth Open” through the hatch at the end of every Qualtrics scorecard.

General Qualtrics set-up

Qualtrics Research Suite can be used to create simple surveys with little to no prior training. Qualtrics allows developers to create surveys from scratch using their quick survey builder, use premade survey templates, or use a copy of an existing survey in a user’s survey library (Figure 3.3).

Once a survey name is created and a folder is chosen for storage in the user’s library, users are taken to the survey builder page where they can work on the first block of their survey (Figure 3.4). Qualtrics surveys are made up of blocks and questions. Entire surveys can be made in one block, or questions can be separated into different blocks, and from there users can use branching and/or display logic to arrange the blocks to appear in specific orders based on survey participants’ answers or other qualifiers.

Users can either create a new question (or item) (Figure 3.5) or use a previously made item from their library; this includes using single questions from previous surveys or entire blocks from previous surveys. Items can include static content, standard questions, specialty questions, and advanced content. Static content includes descriptive text and graphics. Standard questions include multiple choice questions, matrix tables, text entry boxes, slider scales, ranking boxes, and side-by-side questions. Specialty questions include items such as graphic sliders, heat maps, and hot spots, while advanced content includes questions with timer settings (which can be applied as wait time between sensory assessments), file upload options, and Captcha verification. Users can further customize content and questions by choosing from different formats and

layouts; for example, there are 11 different types of rank order questions from which to choose (Figure 3.6).

Users have additional customization and editing options once they choose a question type. For example, for a slider scale question, the number of choices can be increased or decreased, as well as labeled. The scale can be edited to have different minimum and/or maximum values, gridlines can be removed or additional ones can be added. Users can decide on the number of decimal points recorded from each response, as well as whether or not to display the value to the survey participant (Figure 3.7).

There are many customization options available for all survey types, such as a force response option, skip logic, randomization of blocks and questions, and reordering of blocks and questions (Figure 3.8). More advanced survey options include editing the flow of the survey. This includes display logic and branch logic. Display logic can be used to display certain questions conditionally, based on previous survey answers (Figure 3.9) (Qualtrics 2015). This gives users the option to make their surveys more responsive and relevant to each survey respondent.

Branch logic is used to divert survey participants to different survey flow based on different variables such as question answers (Figure 3.10). While display logic is best for a small number of questions, branch logic is more appropriate when users want to selectively show an entire block of questions (Qualtrics 2015).

Qualtrics Research Suite uses ongoing survey responses to create an initial report, which includes minimum and maximum response values, mean response value, the standard deviation, and the number of responses for each question in the survey. The initial report can be exported into Microsoft Word, Excel, Powerpoint, or as a PDF file (Figure 3.11). Raw data can also be

downloaded as various file types and analyzed using statistical software such as SPSS or SAS. Users can choose which questions and data to include in the download file.

iPad and Qualtrics set-up

Qualtrics scorecards were administered on iPads, which were mounted in each of the six individual sensory booths. The iPads were encased in plastic, lockable cases that included cutouts for charger cords. The cases were installed on the right side of each sensory booth approximately 24 cm above the surface on which panelists evaluated samples in order to prevent damage from spilled liquids or food debris. The iPads were oriented horizontally as opposed to vertically so that the size and length of the line scales on the Qualtrics scorecards would better match those on the paper scorecards. During Qualtrics sessions, the iPads were kept charged using extra long charging cables (Amazon.com, Inc., Seattle, WA) plugged in to outlets above each hatch. Panelists completed all Qualtrics scorecards through Google Chrome web browser. All Qualtrics scorecards were set to the Qualtrics stock layout, "Minimalist." Panelists were not permitted to return to pages already submitted while completing a Qualtrics session.

Qualtrics features employed during sensory evaluation

Balanced order of presentation was achieved with the Qualtrics scorecards by employing the block and branch logic features in Qualtrics Research Suite. Blocks are groups of questions that are part of the overall survey flow, which allows users to customize the order in which question blocks and other elements of a survey appear to participants. All Qualtrics scorecards in this research project followed the same basic survey flow (Figure 3.12). First, all panelists saw instructions regarding extra credit information and filling out consent forms, followed by a block of questions entitled "Name", where panelists entered their first and last names and their booth

number. The Name block also included any general instructions specific for completing the sensory test being conducted during that particular session. Depending on the sensory test being conducted, the next one to three blocks contained scorecard elements specific to each type of sensory test. The contents of these blocks will be discussed in the following sections specific to each sensory test. The last block all panelists saw was entitled "End", which contained the usability and demographic questionnaires, as well as instructions for submitting the scorecards and questionnaires and utilizing the "Open Booth" sign. A customized end survey element was used to refresh the scorecard when a panelist submitted his/her completed scorecard.

Branch logic allows users to show different variations of the survey flow to participants based on their responses to certain questions. This feature allowed the presentation of samples to correspond to the order in which the Qualtrics scorecard for each sample appeared on the iPads. For the majority of the sensory tests conducted, the booth number entered by each panelist was used as the variable, or "condition" that determined which block order, and therefore which sample order, each panelist saw while completing the Qualtrics scorecards. Pieces of paper with the corresponding sample order were taped next to each sensory booth hatch on the researchers' side, which allowed the researchers to present the samples in the same order that they appeared on each scorecard.

Other Qualtrics features used consistently across all sensory tests included force response, page break, and total duration embedded data field. The force response option enables users to prevent participants from advancing to the next screen until a response is entered for a question. The force response option was applied to all questions in all Qualtrics sensory tests. Page breaks allow users to break up blocks of questions into multiple pages. Page breaks were utilized in blocks with multiple questions and descriptions to keep panelists from having to scroll

down the page. The total duration embedded data field was added to all Qualtrics sensory tests in order to collect the amount of time it took each panelist to complete the Qualtrics session. Because the time was recorded at the start of each Qualtrics session, the total duration data field was not accurate when there was an extended period of time between panelists at a particular iPad.

Affective sensory tests

9-point structured hedonic line scale

Appearance, texture, flavor, and overall acceptability were assessed in three oatmeal cookies (Psst!, The Kroger Co., Cincinnati, OH; Homekist, South Street Bakery Co., Massillon, OH; Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario) using a 9-point structured hedonic line scale, where 1=extremely dislike and 9=extremely like (Appendices A and I) (n=113). Individual cookies were presented in sealed Ziploc snack bags (S.C. Johnson & Son, Inc., Racine, WI) to panelists one at a time. Samples were prepared one day prior to panels and stored at room temperature. Panelists received baby carrots (The Kroger Co., Cincinnati, OH) in addition to room temperature water as a palate cleanser.

Each of the four scales for each sample on the paper scorecard contained nine numbered boxes (1-9) with anchors "Extremely dislike" and "Extremely like". Instructions to panelists were, "Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by marking the box that best reflects how much you like this product."

The Qualtrics scorecard utilized the draggable slider question format; each scale represented one question (Table 3.2). Two labels, "Extremely dislike" and "Extremely like" were added to each scale as the anchors. Eight gridlines were added to each scale to mimic the

boxes on the paper scorecard. The snap to grid function was enabled to limit panelists to whole numbers, as was done with the paper scorecard. The minimum and maximum values were 1 and 9, respectively, and the number of decimals recorded was 0. No custom start position was used and the force response option was enabled. Each attribute scale for each sample was shown on the iPad screen separately using the page break feature to keep panelists from having to scroll down the page to complete the four line scales for each sample. The ability to see all attribute scales at once for each sample was the biggest visual difference between the two scorecard versions. The instructions for panelists were, "Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by tapping the point on the line scale that best represents your feelings towards the product."

Panelists evaluated the oatmeal cookies using paper scorecards during the first session and the Qualtrics scorecards during the second session (Table 3.1). For the paper scorecards, the consent forms, three 9-point structured hedonic line scale scorecards (one for each sample), usability questionnaire, and demographic questionnaire were paper-clipped together. The order of the three scorecards in the packets was alternated and used to determine sample presentation order and to achieve a balanced order of presentation. For the Qualtrics scorecard, each of the four attribute scales for each sample was contained in their own separate block (Figure 3.13). The resulting survey flow was Name Block, the three sample blocks, which were presented to panelists in alternating orders, and End Block. Each sample was presented first, second, or third an equal number of times through the use of branch logic with booth number as the condition.

9-point unstructured hedonic line scale

Appearance, texture, taste, and overall acceptability were assessed in Greek, low fat, and Carbmaster vanilla yogurt (The Kroger Co., Cincinnati, OH) using a 9-point unstructured

hedonic line scale, where 1=extremely dislike and 9=extremely like (Appendices B and J) (n = 108). Responses were measured to the nearest hundredth decimal place. Yogurt samples were portioned with a #70 scoop into 2-ounce translucent plastic cups with lids (Daily Chef, Sam's West, Inc., Bentonville, AR) one day prior to panels and stored in a refrigerator (Hobart Corporation, Troy, OH) overnight. Samples were removed from the refrigerator 1-3 minutes prior to testing. Samples were presented one at a time, each with a taster spoon (Solo Cup Company, Highland Park, IL). Unsalted top crackers (Kroger, The Kroger Co., Cincinnati, OH) were used as a palate cleanser.

The instructions and design of the paper and Qualtrics scorecards were based on those described by Meilgaard and others (2007) and Kemp and others (2009). Each of the four scales for each sample on the paper scorecard contained an unstructured line with the anchors "Dislike Extremely" and "Like Extremely". Instructions to panelists were, "Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by using a pen to mark the spot on the line scale that best represents how much you like this product."

The Qualtrics scorecard utilized the draggable slider question format; each scale represented one question (Table 3.2). Two labels, "Extremely dislike" and "Extremely like" were added to each scale as the anchors. The scales contained no grid lines in order to make it unstructured. The snap to grid function was disabled to allow panelists to mark any number between 1 and 9, as was done with the paper scorecard. The minimum and maximum values were 1 and 9, respectively, and the number of decimals recorded was 2. No custom start position was used and the force response option was enabled. Each attribute scale for each sample was shown on the iPad screen separately using the page break feature to keep panelists from having to scroll down the page to complete the four line scales for each sample. The ability to see all

attribute scales at once for each sample was the biggest visual difference between the two scorecard versions. The instructions for panelists were, "Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by tapping the point on the line scale that best represents your feelings towards the product."

Panelists evaluated the yogurt samples with Qualtrics scorecards during the first session and paper scorecards during the second session (Table 3.1). The same survey flow used to achieve a balanced order of presentation for the paper and Qualtrics versions of the 9-point structured hedonic scale scorecards also were used for the 9-point unstructured scale scorecard (Figure 3.13).

Ranking test

Three creamy peanut butter samples (Kroger, The Kroger Co., Cincinnati, OH; Jif, The J.M. Smucker Company, Orville, OH; Peter Pan, ConAgra Foods, Inc. Omaha, NE) were ranked according to preference, where 1 = least preferred and 3 = most preferred (Appendices C and K) (n = 107). Samples were portioned using a #70 scoop into 2-ounce translucent plastic cups with lids (Daily Chef, Sam's West, Inc., Bentonville, AR) one day prior to panels and stored overnight at room temperature. Samples were presented all at once with three unsalted top crackers (Kroger, The Kroger Co., Cincinnati, OH) and three taster spoons (Solo Cup Company, Highland Park, IL) for tasting and/or spreading the peanut butter on the crackers. Seltzer water (Kroger, The Kroger Co., Cincinnati, OH) was used as a palate cleanser.

The design and instructions for the paper and Qualtrics scorecards were based on those described by Lawless and Heymann (2011). Panelists were asked to "Please assess the samples in the order presented, from left to right. Rank the samples from most preferred to least

preferred using the following numbers: 1 = least preferred, 3 = most preferred (Ties are NOT allowed)." Panelists wrote 1, 2, or 3 in the spaces provided next to each sample code.

The Qualtrics scorecard utilized the Rank Order question type with 3 items (or samples) (Table 3.3). Text entry boxes were enabled to allow panelists to enter a 1, 2, or 3 by each sample code. The Must Rank Between tool was used to ensure that panelists could only enter the numbers 1, 2, or 3. The items (sample codes) were positioned vertically to match the paper scorecard. Additionally, the force response option was enabled. The instructions for the Qualtrics scorecard matched those on the paper scorecard.

Panelists evaluated the peanut butter samples with paper scorecards during the first session and Qualtrics scorecards during the second session (Table 3.1). The order of the sample codes on the scorecards was kept constant across all panelists for both scorecards. Each sample was presented in the first, second, and third position an equal number of times. This was achieved by assigning sample presentation orders to each of the six sensory booths. Because the order of the sample codes on the scorecards did not change, Branch Logic was not used for the Qualtrics version of this scorecard. The resulting survey flow was Name Block, the sample block, which contained the ranking question, and End Block (Figure 3.14).

Discrimination sensory tests

Triangle test

Perception of difference between regular and Baked Crunchy Cheetos® (Frito-Lay, Inc., Plano, TX) was assessed using a triangle test (Appendices D and L) (n = 98). Three Cheetos® of similar size and shape were distributed into 5-ounce translucent plastic cups with lids (Solo Cup Company, Forest Lake, IL) one day prior to panels and stored overnight at room temperature.

Samples were presented all at once with seltzer water (Kroger, The Kroger Co., Cincinnati, OH) as a palate cleanser. Each of the Cheetos® samples in a triangle was also of similar size and shape.

The instructions and design for the paper and Qualtrics scorecards were based on those by Meilgaard and others (2007) and Kemp and others (2009). The instructions for panelists were, "Circle the number of the sample that is different." The Qualtrics scorecard utilized a single answer multiple choice type question in which each of the three sample codes was an answer choice (Table 3.3). For ease of coding during data analysis, the code responses function was used to recode the odd sample, or answer choice, as 1. The two similar samples, or answer choices, were recoded as 2. The answer choices were arranged horizontally to match the paper scorecard. The sample code labels were positioned above the Multiple Choice buttons. The force response option was enabled.

Panelists evaluated the Cheetos® samples with Qualtrics scorecards during the first session and paper scorecards during the second session (Table 3.1). A balanced order of presentation was achieved by assigning each of the six sensory booths a different set of sample codes. When panelists returned for the second session, they were assigned to a booth with the same odd sample, either in the same position or a different position, as the triangle they assessed during the first session. The Qualtrics responses for the Name Block, which included the booth number that each panelist entered during the first session, was used to ensure that each panelist was sent to a booth with the same odd sample as they had during the first session. Each of the six possible triangles was contained in their own separate block and was assigned to one of the six sensory booths using the branch logic feature. The resulting survey flow was Name Block, the assigned triangle block, and End Block (Figure 3.15). For the paper scorecards, the consent

form, the triangle test scorecard, the usability questionnaire, and the demographic questionnaire were all paper clipped together. Each booth had its own triangle test scorecard with a unique set of sample codes.

Tetrad test

Perception of difference between undiluted and diluted (20% dilution) orange juice from concentrate (Kroger, The Kroger Co., Cincinnati, OH) using a tetrad test (Appendices E and M) as demonstrated in the methodology by Ishii and others (2014) was assessed (n = 108). 60 mL samples were dispensed into 5-ounce translucent plastic cups with lids (Solo Cup Company, Forest Lake, IL) one day prior to panels and stored in a refrigerator overnight (Hobart Corporation, Troy, OH). Samples were removed from the refrigerator 1-3 minutes prior to presentation and were presented all at once with three unsalted top crackers (Kroger, The Kroger Co., Cincinnati, OH) as palate cleansers.

The instructions and design of the paper and Qualtrics scorecards were based on those described by Rousseau and Ennis (2013). The paper scorecards asked panelists to "Group the samples into two groups of two based on similarity. Write the sample codes for the two groups in the spaces below." The Qualtrics scorecard employed the drag and drop version of the pick, group, and rank type question (Table 3.3). Each of the four sample codes was one item, which panelists dragged and dropped into one of two groups. The columns option was enabled, as was the force response option. The stack items (sample codes) options were disabled. Finally, the "each group contains" option was enabled to ensure that panelists dragged and dropped two sample codes into each of the two groups.

Panelists evaluated orange juice samples with paper scorecards during the first session and Qualtrics scorecards during the second session (Table 3.1). Like the triangle tests, a

balanced order of presentation was achieved by assigning each of the six sensory booths a different set of sample codes. Because there is no odd sample in a tetrad test, panelists were not assigned to a specific booth when they returned for the second session. For the paper scorecards, the consent form, the tetrad scorecard, the usability questionnaire, and the demographics questionnaire were all paper clipped together. Each booth had its own tetrad test scorecard with a unique set of sample codes. Similar to the Qualtrics triangle test scorecard, each of the six possible tetrads, all with unique sample codes, was contained in their own separate block and was assigned to one of the six sensory booths using the branch logic feature. The resulting survey flow was Name Block, the assigned tetrad block, and End Block. The survey flow for the Qualtrics triangle test also illustrates the complete survey flow used for the Qualtrics tetrad test (Figure 3.15).

Descriptive sensory test

15-point unstructured descriptive line scale

6 trained panelists from The Evaluation of Almonds Study (Cheely 2014) evaluated two dry medium-roasted Nonpareil almond samples (The Almond Board of California, Modesto, CA) held under different combinations of temperatures, humidity levels, and packaging materials for 16 months using 15-point unstructured descriptive line scale scorecards (Appendices F and N. One almond sample was held at 25°C at 65% relative humidity (RH) in a polypropylene bag and stored in a HotPack (SP Industries, Warminster, PA) environmental chamber. The second almond sample was held at 15°C without humidity control in a high barrier bag and stored in a Thermo Scientific Incubator (Thermo Fisher Scientific, Inc., Waltham, MA). Additionally,

panelists were given a warm-up roasted/unsalted almond sample (The Fresh Market, Inc., Greensboro, NC) at the start of the evaluation session.

Almond samples were evaluated in two separate phases. During the first phase, panelists assessed hardness and crunchiness as defined by Meilgaard and others (1991) and sweetness as defined by Meilgaard and others (1991). Each sample contained five almonds in 2-ounce clear, translucent cups with lids (Daily Chef, Sam's West, Inc., Bentonville, AR**)** and was served with baby carrots (Kroger, The Kroger Co., Cincinnati, OH**)** as a palate cleanser. Panelists also received reference scales. During the second phase, panelists received a set of samples that included five almonds and 3 odor reference samples in order to evaluate odor and flavor rancidity attributes, along with baby carrots as a palate cleanser. The Universal Intensity Scale was used for reference during the second phase. Panelists waited one minute between phases, and 7 minutes between samples.

Odor samples were prepared at least two hours, but no more than six hours, prior to evaluation. Whole roasted almonds were placed in a mini-food processor (Proctor Silex 1.5C Mini Food Chopper, Hamilton Beach Brand, Inc., Southern Pines, NC) and then pulsed and shaken simultaneously for 10 seconds. The chopped almonds were then sifted through a 2-mm sieve to remove fine almond dust. The chopped almonds remaining in the sieve were then sifted through a 5-mm sieve. The remaining almonds that measured between 2 and 5-mm were used for the odor samples. Larger almond pieces were pulsed again in the mini food processor. 15.0 grams of the chopped almonds were presented in 5-ounce clear, translucent cups with a lid (Solo Cup Company, Forest Lake, IL). Three odor samples were presented with each whole almond sample during the second phase, one cup for each oxidation factor evaluated.

The same settings used for the Qualtrics version of the 9-point unstructured hedonic scale were used for the Qualtrics version of the 15-point unstructured descriptive scale, except the minimum and maximum values, which were set to 0 and 15, respectively (Table 3.2). The labels (anchors) were 0 = Not perceptible and 15 = High Intensity.

Panelists evaluated the almond samples using paper scorecards during the first session and the Qualtrics scorecards during the second session (Table 3.1). Panelists completed the second session 1 – 24 hours after the first session. For the paper scorecards, the consent forms, 15-point unstructured line scale scorecards (one for each phase; two for each sample), usability questionnaire, and demographic questionnaire were paper-clipped together. Every panelist evaluated the warm-up sample first, but the order of presentation was balanced for the 15°C high barrier bag sample and the 25°C/65% RH polypropylene bag sample.

For the Qualtrics scorecard, each phase for each sample was contained in a separate block. Page breaks were not used to separate scales in each block, and therefore panelists had to scroll down the screen to complete every scale in a block. The resulting survey flow was Name Block, Warm-up Block – Phase 1 (hardness, sweetness, and crunchiness), Warm-up Block – Phase 2 (odor and flavor rancidity attributes), 2565 Phase 1, 2565 Phase 2, HBB Phase 1, and HBB Phase 2 (2565 and HBB were alternated for a balanced order of presentation), and End Block (Figure 3.16). It should be noted that the two phases for each sample could be contained in the same block for each sample for a more streamlined survey flow. A balanced order of presentation was achieved by using branch logic with booth number as the condition. Panelists were oriented to the Qualtrics scorecard before they entered the sensory booth to evaluate the samples, just as they were oriented to the paper scorecard and Universal Intensity Scale at the start of The Acceptability of Almonds Study.

Statistical analysis

All results were analyzed using SAS University Edition (SAS Inc., Cary, NC). Usability and demographics questionnaires were analyzed by PROC FREQ to profile each set of panels.

Affective and descriptive sensory tests

PROC UNIVARIATE was used to produce normality plots in order to confirm normal distribution of data and equal variance. Data that lacked normality or equal variance was transformed to meet the necessary assumptions for appropriate analysis. Outlier data points were removed when necessary. PROC GLM was used to determine whether or not scorecard type, sample, or an interaction between the two had a significant effect on panelists' responses ($p \leq 0.05$). Least-square means and standard errors were generated and PDIFF was used for means separation. PROC TTEST with the PAIRED statement was used to determine whether there were significant differences between panelists' responses using the paper scorecard and the Qualtrics scorecard ($p \leq 0.05$).

Ranking and discrimination tests

Friedman's Test was used to test for significant differences in the distribution of ranking scores for each scorecard version ($p \leq 0.05$). Significant differences in panelists' responses using the two scorecard versions were determined using Wilcoxon Signed Rank test ($p \leq 0.05$).

Panelists' ability to detect a difference between the samples in the triangle and tetrad tests was determined with the Critical Values Table for Sensory Differences (Meilgaard and others 2007) ($p \leq 0.05$). Significant differences in panelists' responses using the two scorecard versions were determined using Wilcoxon Signed Rank test ($p \leq 0.05$).

Usability questionnaire

PROC UNIVARIATE was used to generate descriptive statistics. PROC FREQ determined frequency of responses in order to profile panelists' opinions of the paper and Qualtrics scorecards.

Table 3.1: Counterbalanced order [a] of paper and Qualtrics[b] sensory scorecards in a study comparing traditional paper sensory scorecards to electronic scorecards administered via Qualtrics on iPads

Sensory Tests	Food Product	Session 1[c]	Session 2[c]
9-point structured hedonic line scale [d]	Oatmeal cookies	Paper	Qualtrics
9-point unstructured hedonic line scale [e]	Greek, low-fat, and Carbmaster vanilla yogurt	Qualtrics	Paper
Ranking test [f]	Peanut butter	Paper	Qualtrics
Triangle test [g]	Regular and Baked Crunchy Cheetos®	Qualtrics	Paper
Tetrad test [h]	Undiluted and 20% diluted orange juice	Paper	Qualtrics
15-point unstructured descriptive line scale [i]	Roasted almonds	Paper	Qualtrics

[a] Traditional and electronic sensory scorecards were compared using a counterbalanced, repeated measures design

[b] Qualtrics Research Suite, Provo, UT

[c] Each sensory test was conducted twice with the same panelists. Panelists completed one version of a scorecard during session 1, then was asked via email to return one week later to complete the other version of a scorecard during session 2

[d] Appearance, texture, taste, and overall acceptability was evaluated in three oatmeal cookies (Psst!, The Kroger Co., Cincinnati, OH; Homekist, South Street Bakery Co., Massillon, OH; Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario) using a 9-point structured hedonic line scale where 1=extremely dislike and 9=extremely like

[e] Appearance, texture, taste, and overall acceptability was evaluated in Greek, low fat, and Carbmaster vanilla yogurt (The Kroger Co., Cincinnati, OH) using a 9-point unstructured line scale where 1=extremely dislike and 9=extremely like

[f] Three brands of peanut butter (Kroger, The Kroger Co., Cincinnati, OH; Jif, The J.M. Smucker Company, Orville, OH; Peter Pan, ConAgra Foods, Inc. Omaha, NE) were ranked based on preference where 1=least preferred and 3=most preferred

[g] Perception of difference between Regular and Baked Crunchy Cheetos® (Frito-Lay, Inc., Plano, TX) was assessed using triangle tests where each panelists assessed one triangle

[h] Perception of difference between undiluted and 20% diluted orange juice samples (The Kroger Co., Cincinnati, OH) using tetrad tests where each panelists assessed one tetrad

[i] Texture and odor and flavor rancidity attributes in dry, medium-roasted Nonpareil almonds (Almond Board of California) stored in either polypropylene bags at 25°C with 65% relative humidity or high barrier bags at 15°C for 16 months using a 15-point unstructured descriptive line scale and predefined and standardized lexicons represented descriptive tests, as part of the Spectrum-like Descriptive Analysis Method™

Table 3.2 Qualtrics settings for 9-point structured and unstructured hedonic line scale and 15-point unstructured descriptive line scale scorecards[a] in a study comparing traditional paper sensory scorecards to electronic scorecards administered via Qualtrics on iPads

Sensory Test	Question Type	Items	Labels	Grid lines	Min and Max Value	Type	Position	Decimal Points	Force Response	Other
	----------Qualtrics[b] Settings----------									
9-point structured hedonic line scale[c]	Draggable Slider	1	2	8	Min: 1 Max: 9	Slider	N/A	0	Yes	Snap to Grid: Yes Show Value: No
9-point unstructured hedonic line scale[d]	Draggable Slider	1	2	0	Min: 1 Max: 9	Slider	N/a	2	Yes	Snap to Grid: No Show Value: No
15-point unstructured descriptive line scale[e]	Draggable Slider	1	2	0	Min: 0 Max: 15	Slider	N/A	2	Yes	Snap to Grid: No Show Value: No

[a] Traditional and electronic sensory scorecards were compared using a counterbalanced, repeated measures design. Each sensory test was conducted twice with the same panelists. Panelists completed one version of a scorecard during session 1, then was asked via email to return one week later to complete the other version of a scorecard during session 2

[b] Qualtrics Research Suite, Provo, UT

[c] Appearance, texture, taste, and overall acceptability was evaluated in three oatmeal cookies (Psst!, The Kroger Co., Cincinnati, OH; Homekist, South Street Bakery Co., Massillon, OH; Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario) using a 9-point structured hedonic line scale where 1=extremely dislike and 9=extremely like

[d] Appearance, texture, taste, and overall acceptability was evaluated in Greek, low fat, and Carbmaster vanilla yogurt (The Kroger Co., Cincinnati, OH) using a 9-point unstructured line scale where 1=extremely dislike and 9=extremely like

[e] Texture and odor and flavor rancidity attributes in roasted almonds (Almond Board of California) stored in either polypropylene bags at 25°C with 65% relative humidity or high barrier bags at 15°C for 16 months were evaluated using a 15-point unstructured descriptive line scale and predefined and standardized lexicons represented descriptive tests, as part of the Spectrum-like Descriptive Analysis Method™

Table 3.3 Qualtrics settings for ranking, triangle, and tetrad test scorecards[a] in a study comparing traditional paper sensory scorecards to electronic scorecards administered via Qualtrics on iPads

Sensory Test	Question Type	Items	Labels	Grid lines	Min and Max Value	Type	Position	Decimal Points	Force Response	Other
	----------Qualtrics[b] Settings----------									
Ranking test[c]	Rank Order	3	N/A	N/A	N/A	Text Box	Vertical	N/A	Yes	Must Rank Between: Must rank at least 1 and no more than 3
Triangle Test[d]	Multiple Choice	3	N/A	N/A	N/A	Single Answer	Horizontal	N/A	Yes	Label Position: Above Recode Values: Odd sample = 2; other samples = 1
Tetrad Test[e]	Pick, Group, and Rank	4; 2 groups	N/A	N/A	N/A	Drag and Drop with columns	N/A	N/A	Yes	Stack Items: No Stack Items in Groups: No Each Group Contains: At least 2 and no more than 2

[a] Traditional and electronic sensory scorecards were compared using a counterbalanced, repeated measures design. Each sensory test was conducted twice with the same panelists. Panelists completed one version of a scorecard during session 1, then was asked via email to return one week later to complete the other version of a scorecard during session 2

[b] Qualtrics Research Suite, Provo, UT

[c] Three brands of peanut butter (Kroger, The Kroger Co., Cincinnati, OH; Jif, The J.M. Smucker Company, Orville, OH; Peter Pan, ConAgra Foods, Inc. Omaha, NE) were ranked based on preference where 1=least preferred and 3=most preferred

[d] Perception of difference between Regular and Baked Crunchy Cheetos® (Frito-Lay, Inc., Plano, TX) was assessed using triangle tests where each panelists assessed one triangle

[e] Perception of difference between undiluted and 20% diluted orange juice samples (The Kroger Co., Cincinnati, OH) using tetrad tests where each panelists assessed one tetrad

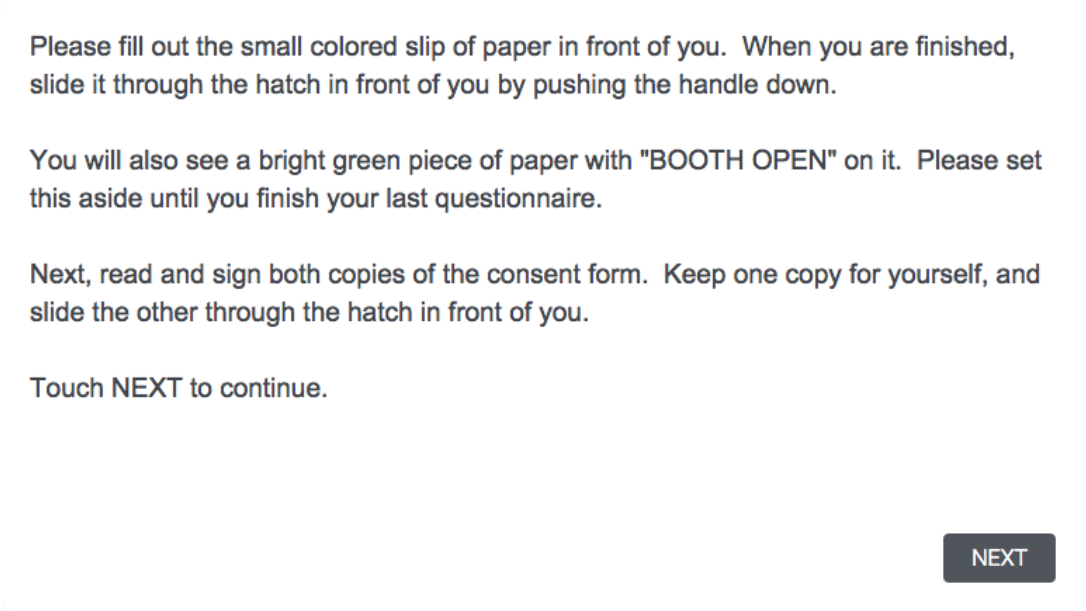


Figure 3.1 Booth instructions for panelists during Qualtrics (Qualtrics Research Suite, Provo, UT) sensory evaluation sessions

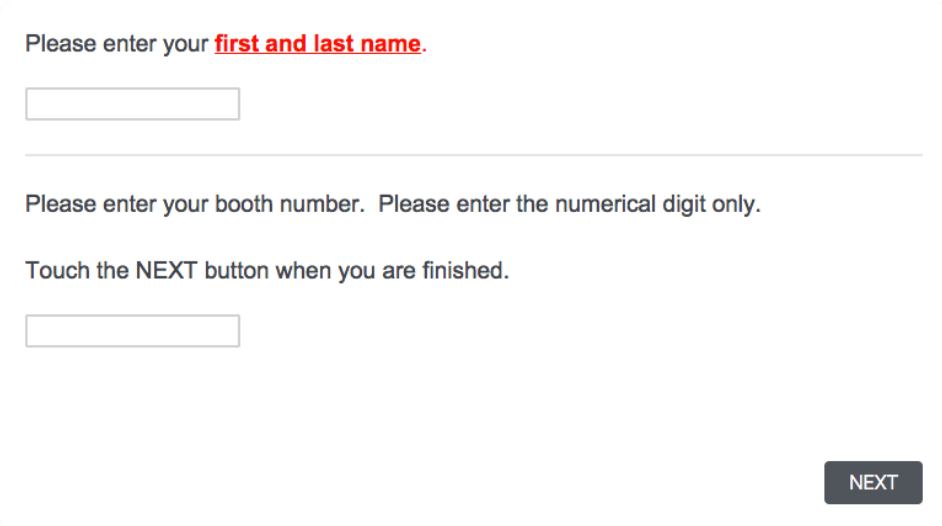


Figure 3.2 Name and booth number questions completed by panelists during Qualtrics (Qualtrics Research Suite, Provo, UT) sensory evaluation sessions

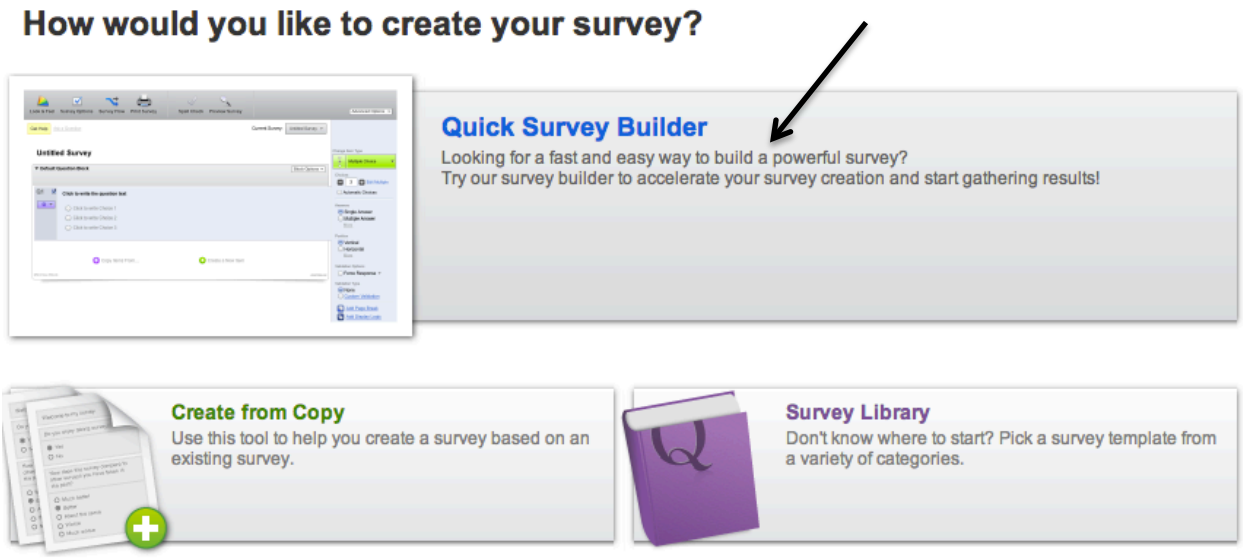


Figure 3.3 Survey builder in Qualtrics Research Suite

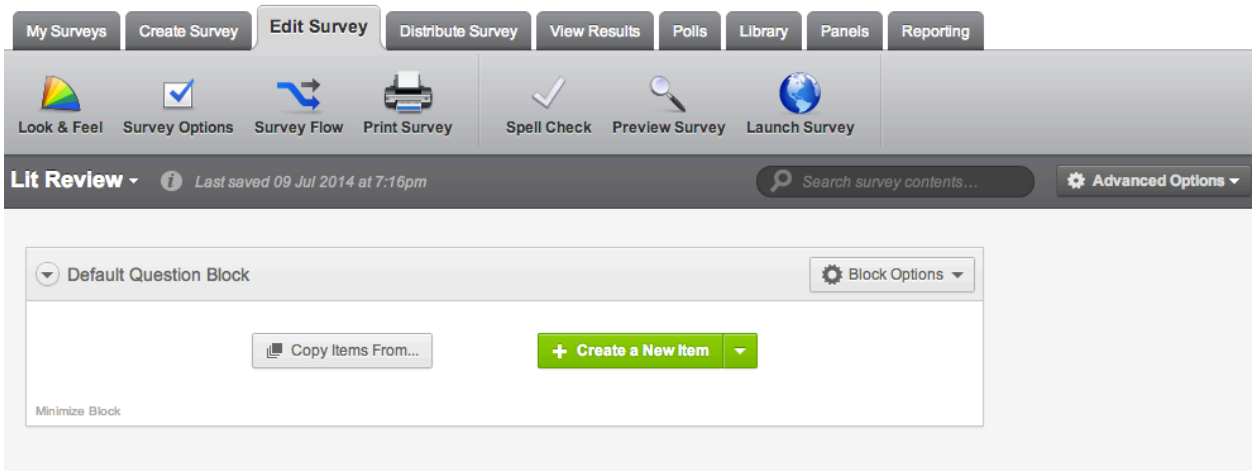


Figure 3.4 Default question block in Qualtrics

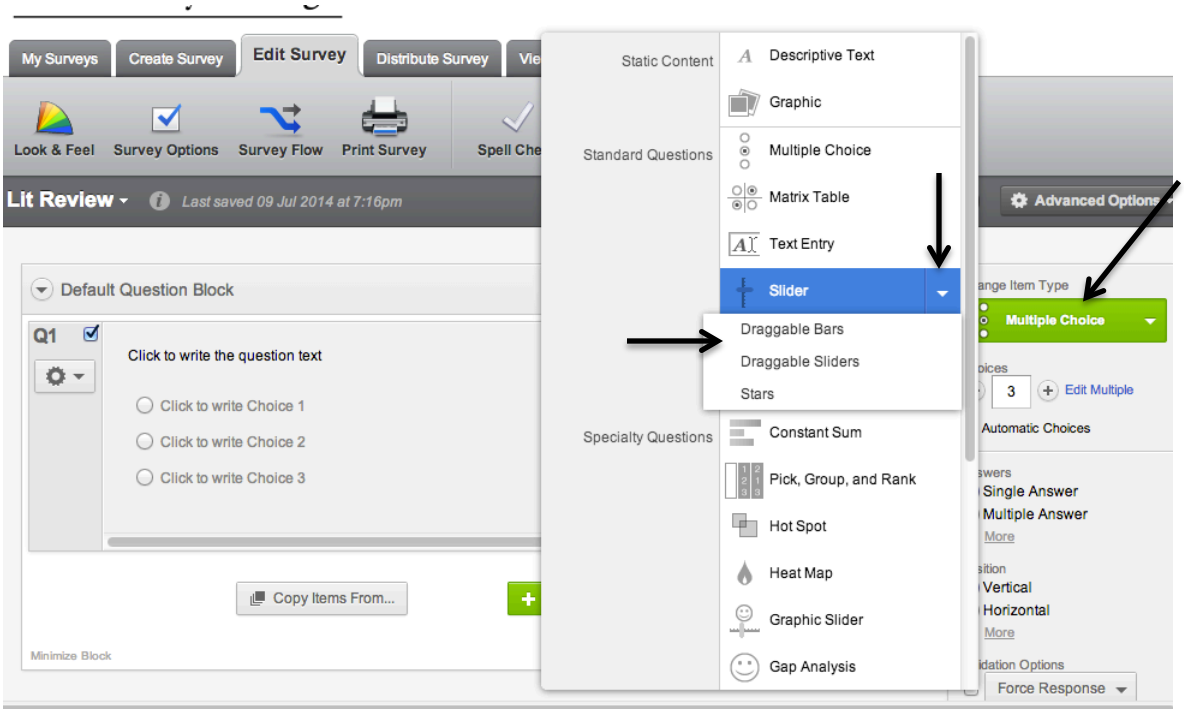


Figure 3.5 Creating a new question in Qualtrics

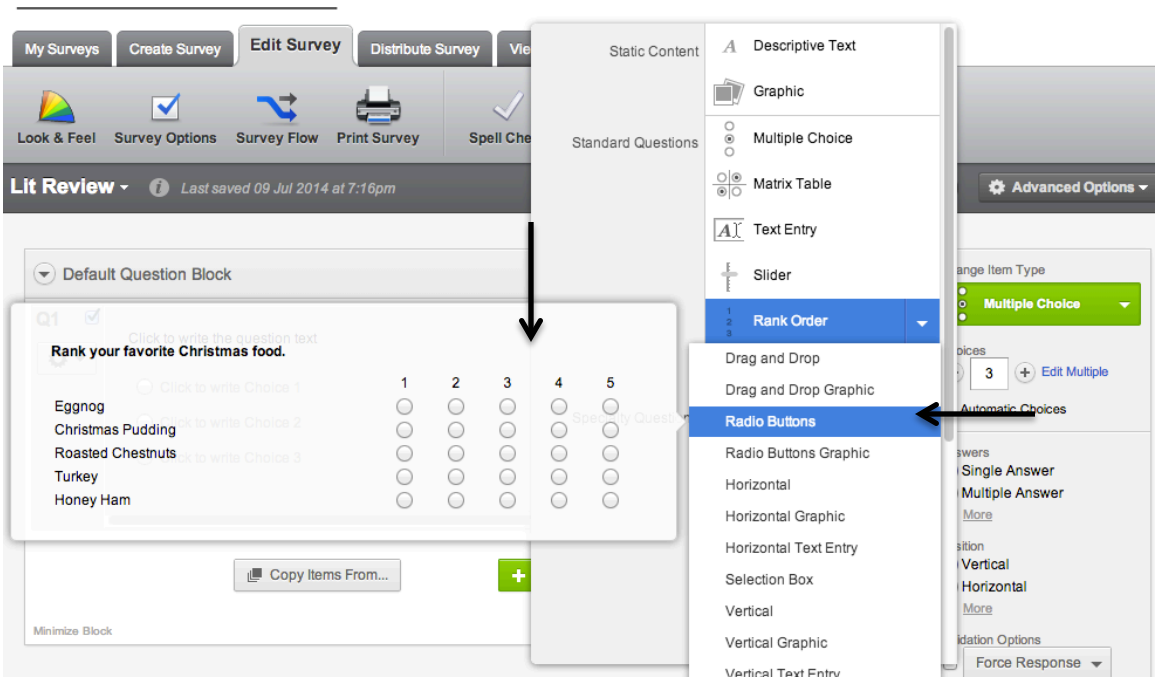


Figure 3.6 Question types available for use in Qualtrics

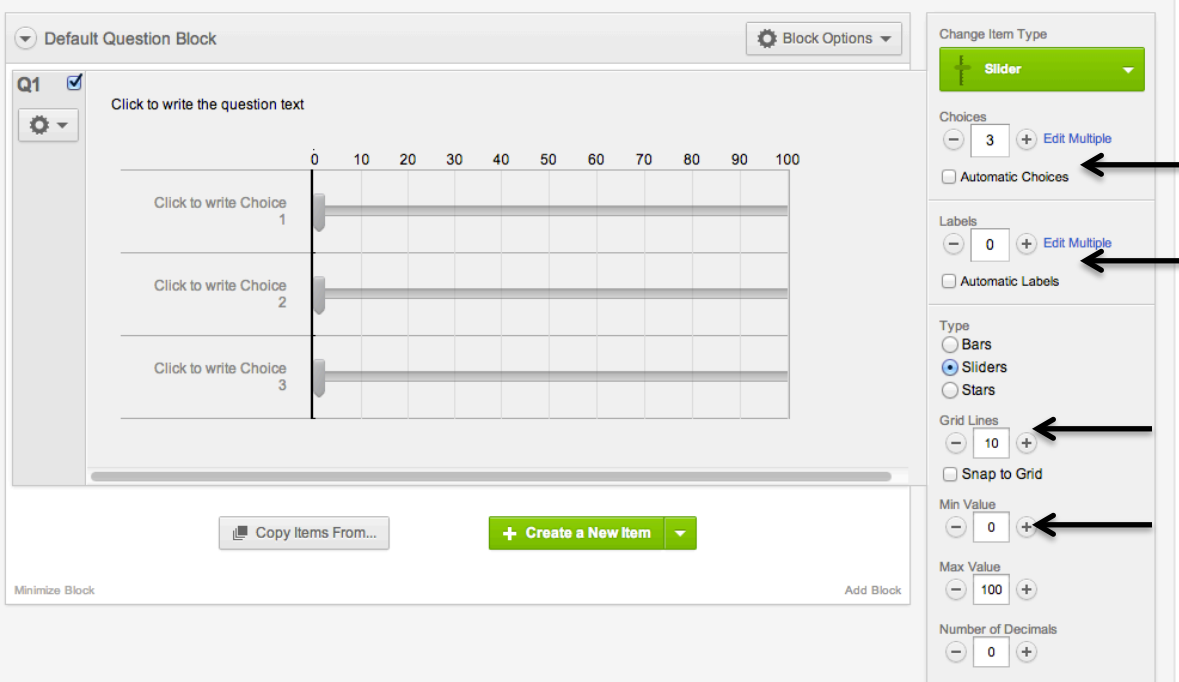


Figure 3.7 Customizing questions in Qualtrics

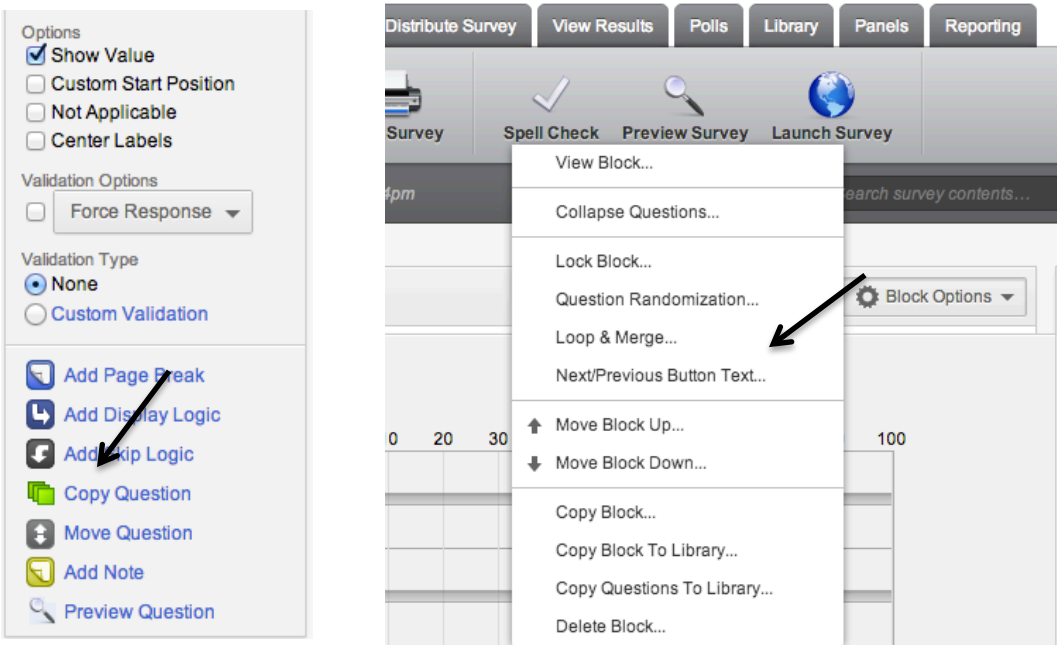


Figure 3.8 Block and question options in Qualtrics

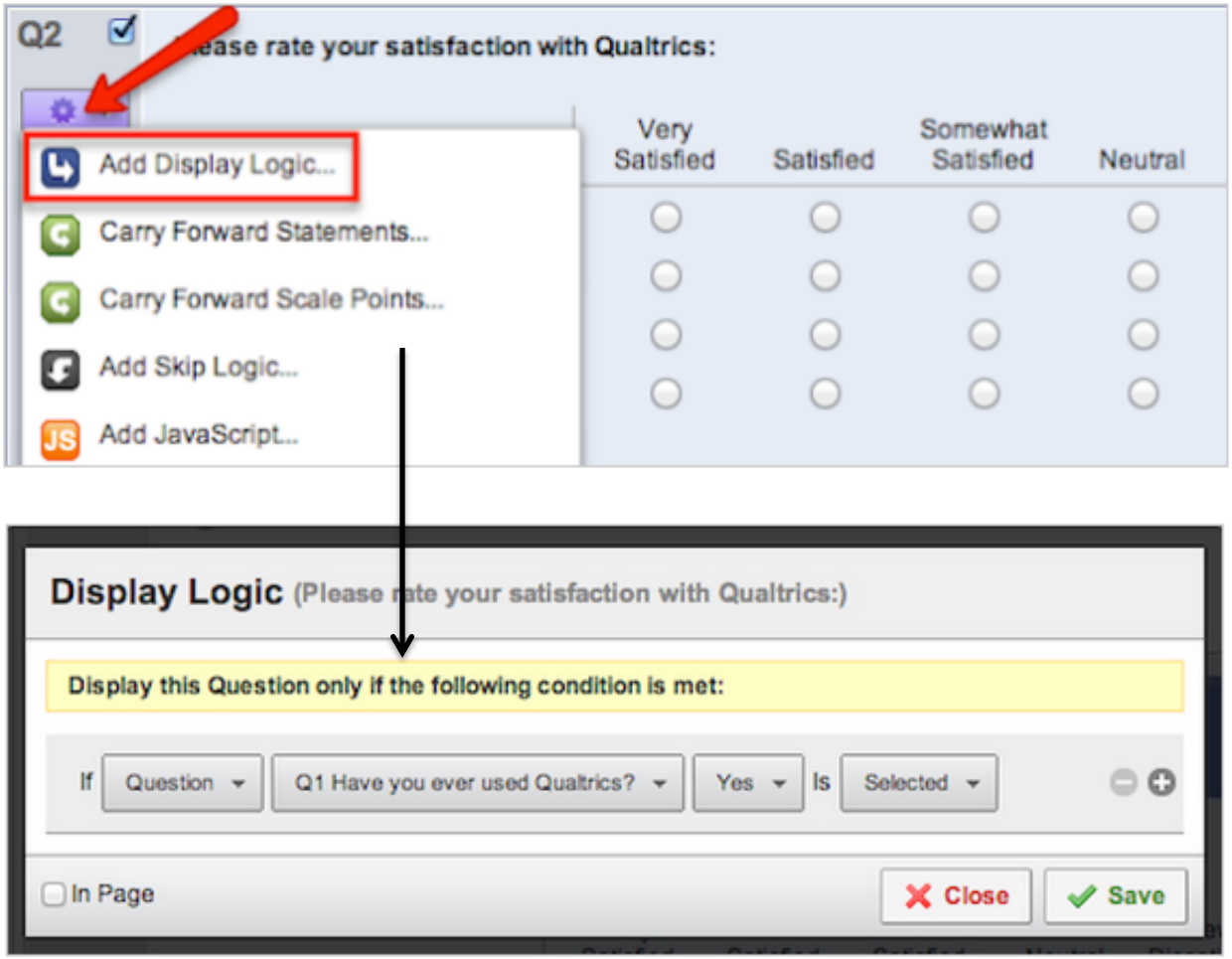
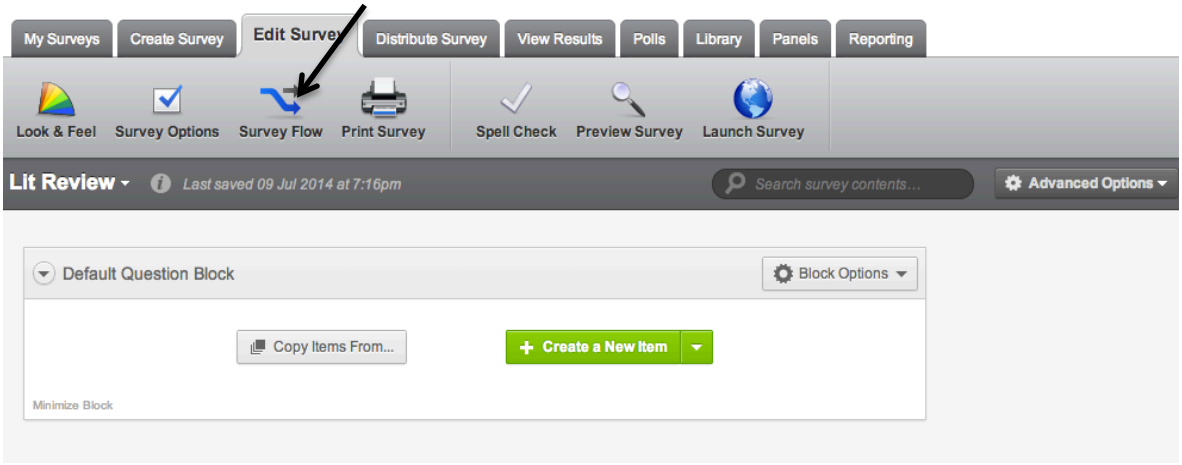


Figure 3.9 Display Logic in Qualtrics



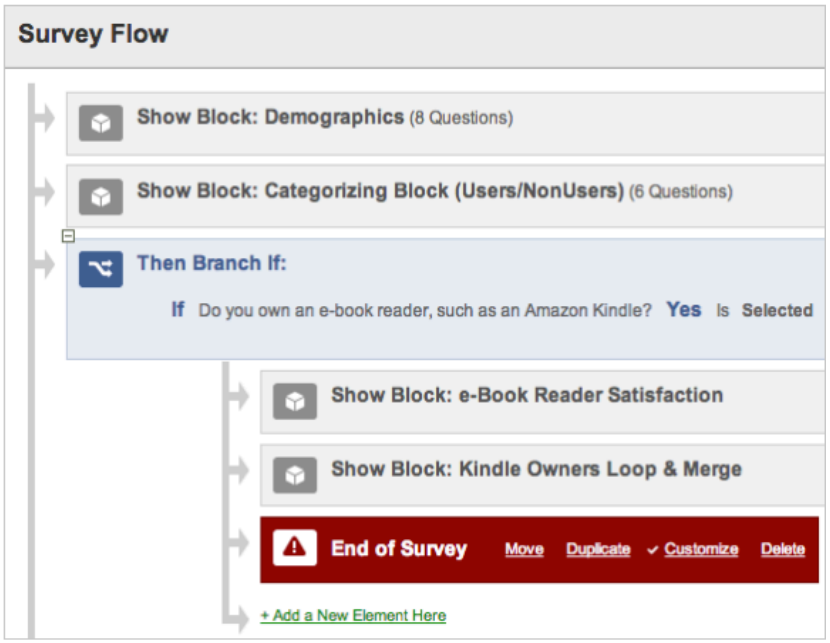


Figure 3.10 Editing Survey Flow using branch logic in Qualtrics

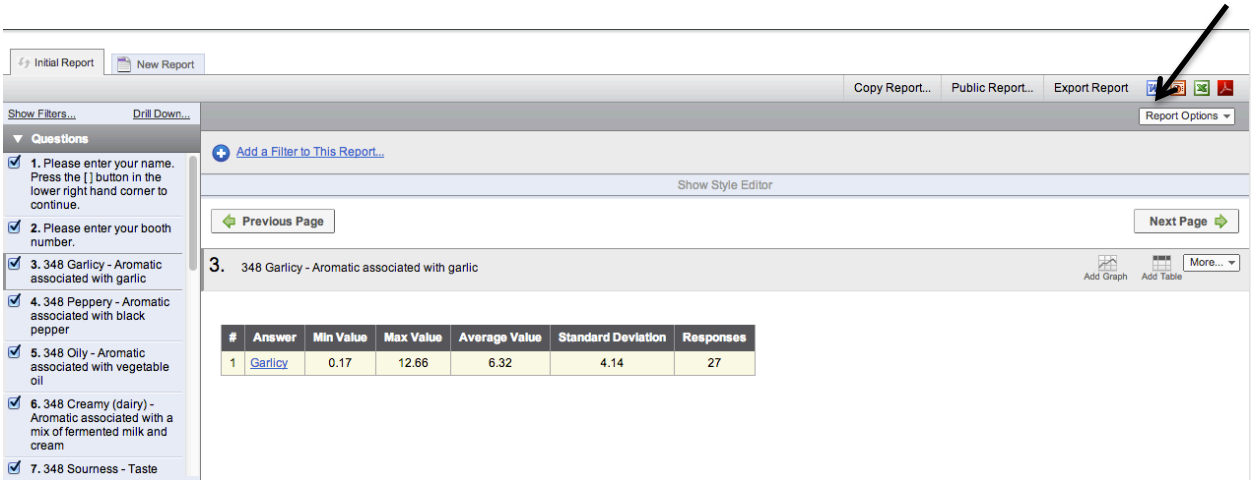


Figure 3.11 Initial report in Qualtrics

Name Block

Sensory Scorecards/Sample Blocks

End Block

End of Survey (Customized to reroute to back to Name Block for following panelist)

Figure 3.12 Basic survey flow for Qualtrics (Qualtrics Research Suite, Provo, UT) scorecards

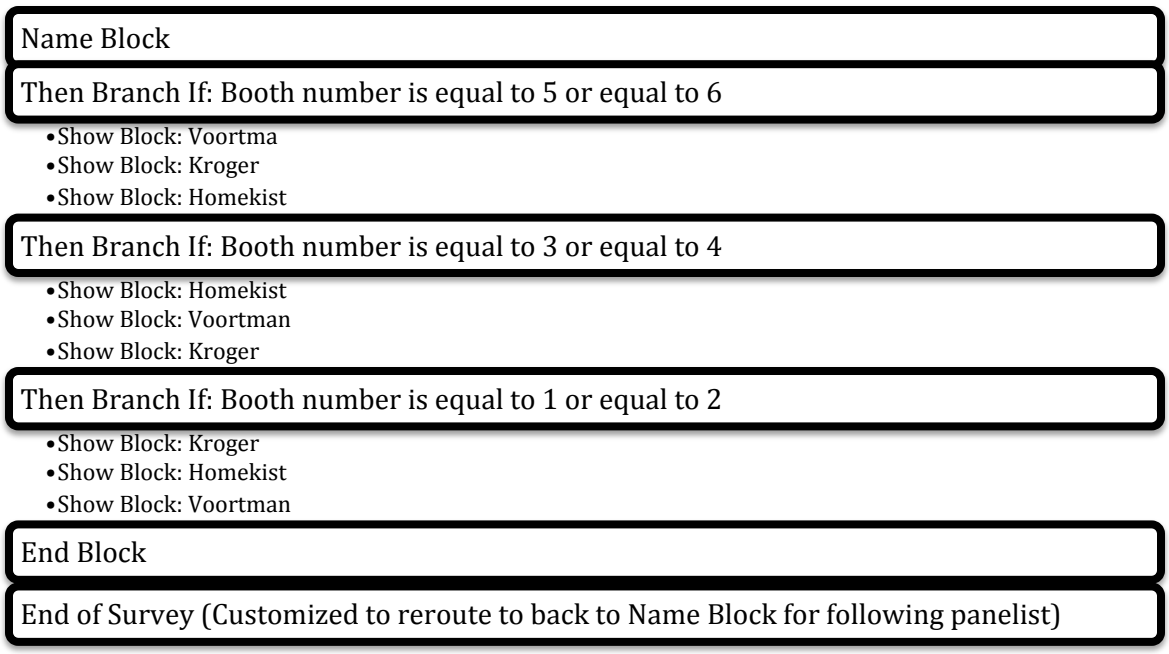


Figure 3.13 Survey flow for Qualtrics (Qualtrics Research Suite, Provo, UT) version of 9-point structured hedonic scale scorecard used to assess appearance, texture, taste, and overall acceptability in three oatmeal cookies (Psst!, The Kroger Co., Cincinnati, OH; Homekist, South Street Bakery Co., Massillon, OH; Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario) where 1=extremely dislike and 9=extremely like

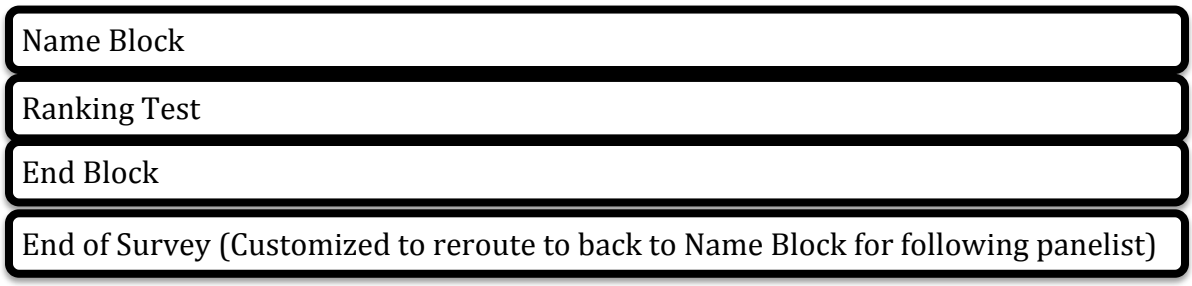


Figure 3.14 Survey flow for Qualtrics (Qualtrics Research Suite, Provo, UT) version of affective ranking scorecard used to assess preference of three peanut butter samples (Kroger, The Kroger Co., Cincinnati, OH; Jif, The J.M. Smucker Company, Orville, OH; Peter Pan, ConAgra Foods, Inc. Omaha, NE) where 1=most preferred and 3=least preferred

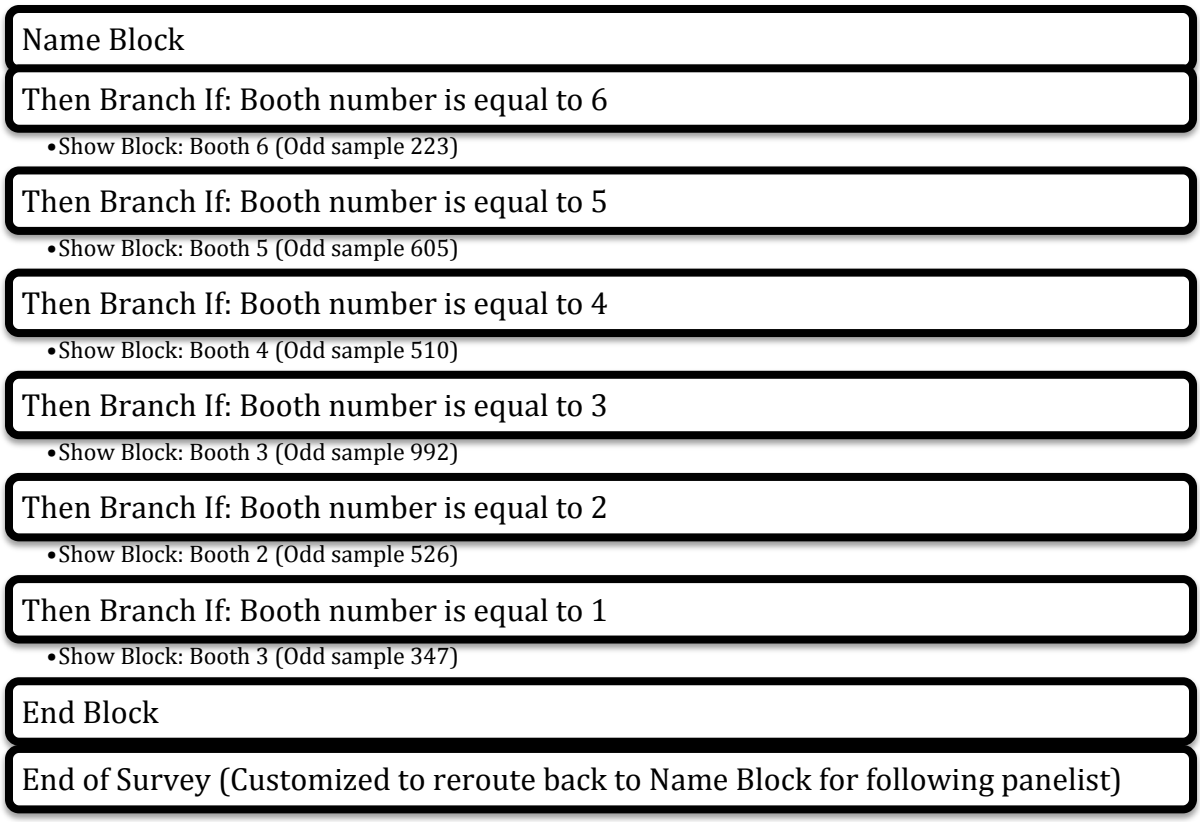


Figure 3.15 Survey flow for Qualtrics (Qualtrics Research Suite, Provo, UT) version of triangle test scorecard used to assess perception of difference between Regular and Baked Crunch Cheetos® (Frito Lay, Inc., Plano, TX)

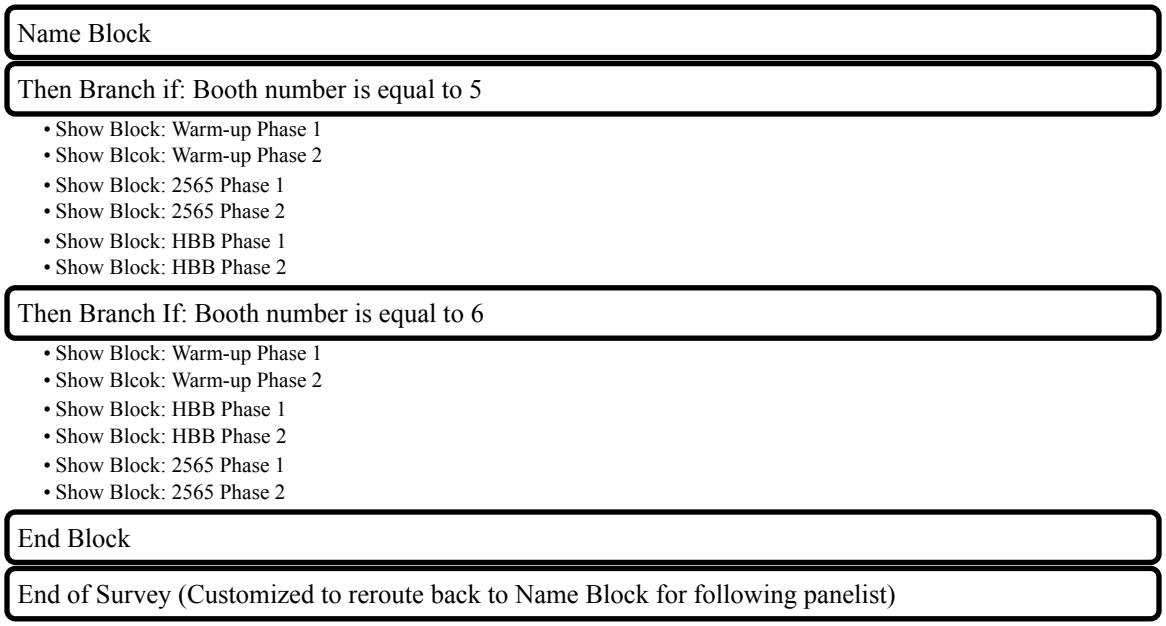


Figure 3.16 Survey flow for Qualtrics (Qualtrics Research Suite, Provo, UT) version of 15-point unstructured descriptive line scale scorecard used to assess texture and odor and flavor rancidity attributes in dry, medium-roasted Nonpareil almond stored in either polypropylene bags at 25°C with 65% relative humidity or high barrier bags at 15°C for 16 months using predefined and standardized lexicons, as part of the Spectrum-like Descriptive Analysis Method™

References

Cheely AC. 2014. Consumer Acceptability of Dry Medium-Roasted Almonds Stored Under Varying Storage Conditions (Master's Thesis). University of Georiga.

Ishii R, O'Mahoney M, Rousseau B. Triangle and tetrad protocols: Small sensory differences, resampling and consumer relevance. Food Quality and Preference 31:49-55.

Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell.

Lawless and Heymann. 2011. Sensory Evaluation of Food. New York, NY: Springer.

Meilgaard M, Civille GV, Carr BT. 1991. Sensory Evaluation Techniques. 2nd ed. Florida: CRC Press. 354 p.

Meilgaard MC, Civille GV, Carr BT. 2007. Sensory Evaluation Techniques. Boca Raton, FL: CRC Press.

Qualtrics. 2015. Qualtrics University Support and Training. Retrieved June 13, 2014 from Qualtrics on the World Wide Web: http://www.qualtrics.com/university/researchsuite/?utm_campaign=Help%20Button&utm_medium=Product&utm_source=MySurveysSection.

Rousseau B and Ennis JM. 2013. Importance of correct instructions in the tetrad test. Journal of Sensory Studies 28(4): 264-269.

CHAPTER 4: RESULTS AND DISCUSSION

Sensory techniques are used to assess acceptability (hedonic scales, ranking tests), to describe products (descriptive analysis), and to discriminate between or among products (triangle, tetrad tests). A counterbalanced, repeated measures study design was used to compare two versions of 9-point structured and unstructured hedonic line scales, a 15-point unstructured descriptive line scale, a ranking test, a triangle test, and a tetrad test: a traditional paper version and a Qualtrics (Qualtrics Research Suite, Provo, UT) version administered on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA) were completed. Data collection was repeated after a one-week interval. Order of data collection method varied with test. Volunteer panelists assessed the food products, then completed usability and demographic questionnaires.

Concurrent validity of Qualtrics scorecards

9-point structured hedonic line scale

113 panelists assessed appearance, texture, flavor, and overall acceptability of three commercial oatmeal cookies (Psst!, The Kroger Co., Cincinnati, OH; Homekist, South Street Bakery Co., Massillon, OH; Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario), first with a paper scorecard, and then one week later with the Qualtrics scorecard. The majority of the panelists were female (88.5%). 79.7% of panelists were between the ages 18-23.

Panelists rated the acceptability of all sensory attributes between 4 and 6 on the 9-point scale, where 9 is extremely like and 1 is extremely dislike (Table 4.1). There were no significant differences due to data collection method across all brands for any of the sensory attributes ($p > 0.05$). Across both data collection methods, the acceptability of all sensory attributes differed significantly due to cookie brand ($p \leq 0.05$), which is expected in a hedonic test. Additionally, a significant interaction (PROC GLM) existed between data collection method and cookie brand for acceptability of appearance and texture ($p \leq 0.05$). Homekist cookies were the most acceptable in regards to appearance, while Kroger cookies were the most acceptable in regards to texture, flavor, and overall acceptability. The difference in the LS-means for acceptability ratings between data collection methods for each oatmeal cookie brand within sensory attribute all are less than 1 unit apart and fell within the moderately acceptable range of the hedonic scale. This indicates that, from session 1 to session 2, panelists rated cookie samples similarly regardless of data collection method.

Paired samples t-test (PROC TTEST) showed no significant differences between data collection methods for 10 out of 12 attribute pairs within brand name ($p > 0.05$) (Table 4.2). Acceptability ratings for appearance and texture in Voortman Sugar Free oatmeal cookies differed significantly between data collection methods, however the LS-means differed less than 1 unit and both fell in the "moderately like" range. Plemmons and Resurreccion (1999) found few significant differences in acceptability ratings between paper and computerized scorecards as well. They found no significant differences ($p > 0.05$) in appearance, flavor, and overall acceptability ratings between scorecard versions for two out of the three food products assessed, but did find significant differences in flavor and overall acceptability in the third product assessed. The mean ratings for flavor and overall acceptability differed less than 0.5 units and

fell within the same acceptability range. Panelists without computer experience tended to rate some sensory attributes for potato chips and cereal higher than did panelists with computer experience. Although experience with iPads was not assessed, data collection method did not have a significant effect on the way panelists rated flavor and overall acceptability in the oatmeal cookie samples ($p > 0.05$), however the brand of the oatmeal cookie did have a significant effect on panelists' ratings, which is expected in a hedonic test.

9-point unstructured hedonic line scale

108 panelists assessed appearance, texture, flavor, and overall acceptability in three vanilla yogurt samples (Carbmaster, Greek, and low-fat vanilla yogurt, The Kroger Co., Cincinnati, OH). In accordance with the counterbalanced, repeated measures study design, panelists completed the Qualtrics scorecard first, and were asked to return one week later to repeat the evaluation using paper scorecards. The majority of panelists were female (88.9%). 90.7% of the panelists were between the ages of 18-23.

Panelists rated the acceptability of all sensory attributes between 4 and 7 on the 9-point scale, where 9 is extremely like and 1 is extremely dislike. No significant interactions (PROC GLM) existed between data collection method and vanilla yogurt type ($p > 0.05$). The main effects for data collection method and vanilla yogurt type were significant for acceptability of appearance ratings ($p \leq 0.05$) (Table 4.3). Harrison and others (1971) postulated that one reason panelists' preferences may change over subsequent hedonic evaluations relates to the "conflictual state" they experience during the initial hedonic judgment involving panelists' met or unmet expectations and surprises from the food product. Subsequent hedonic judgment allows panelists to evaluate the same food product outside of the "conflictual state" and more thoroughly assess the complexity of the food product. In the case of the vanilla yogurt, the

appearance of the three varieties was noticeably different, despite the fact that the samples were presented monodically. The varying white and off-white colors of the yogurt samples may not have met panelists' expectations, and thus triggered the "conflictual state" for panelists. Because panelists were asked to return for a second session for each sensory test, there also was likely self-training by the panelists. This is a limitation of repeat assessments, particularly with consumer panelists.

The main effect for data collection method across all yogurt types were not significant for texture, flavor, or overall acceptability ($p > 0.05$). The main effect for vanilla yogurt type was significant for appearance, texture, flavor, and overall acceptability ($p \leq 0.05$). Carbmaster yogurt was the most acceptable in regards to appearance and texture, whereas low-fat yogurt was found to be the most acceptable in regards to flavor and overall acceptability. The difference in the LS-means for acceptability ratings between data collection method for each vanilla yogurt type was within 0.75 units and fell within the moderately acceptable range of the hedonic scale. This indicates that, from session 1 to session 2, panelists rated yogurt samples similarly regardless of data collection method.

Paired samples t-test (PROC TTEST) showed no significant differences between data collection methods for 10 out of the 12 attribute pairs within yogurt type ($p > 0.05$) (Table 4.4). Overall acceptability of Greek vanilla yogurt and appearance of low-fat vanilla yogurt were significantly different between data collection methods, however the LS-Means differed by less than 1 unit for both pairs and both still fell in the "moderately like" range.

Ranking test

107 panelists ranked the acceptability of three brands of creamy peanut butter (Kroger, The Kroger Co., Cincinnati, OH; Jif, The J.M. Smucker Company, Orville, OH; Peter Pan,

ConAgra Foods, Inc., Omaha, NE), where 1 = most preferred and 3 = least preferred. Panelists evaluated the samples using the Qualtrics scorecard during the first session and the paper scorecard during the second session. 92.2% of the panelists were female, and the majority of panelists were between the ages of 18-23 (88.2%).

According to Friedman's Test, there was no significant difference in the distribution of ranking scores for the paper scorecards or the Qualtrics scorecards ($p > 0.05$). A Wilcoxon Signed Rank test indicated that there were no significant differences between the responses when the products were assessed with paper and Qualtrics scorecards ($p > 0.05$). This indicates that, from session 1 to session 2, panelists ranked the peanut butter samples similarly, regardless of data collection method. Armstrong and others (1997) assessed the consistency, defined as, "the number of times when both methods produced the same response by each tester", in panelists responses by calculating the percentage of responses that were duplicated for both data collection methods. Armstrong and others (1997) found a 40% level of consistency existed between responses for paper and computerized acceptance ranking tests. Armstrong and others (1997) attributed the low level of consistency to the subjective nature of consumer acceptability. Additionally, the researchers commented that the type of food sample and the number of samples panelists have to evaluate likely affected the consistency between the paper and computerized ranking scorecard.

Triangle test

98 panelists completed triangle tests with Regular and Oven Baked Crunch Cheetos® (Frito-Lay, Inc., Plano, TX). Panelists first completed the Qualtrics version of the triangle test scorecard, and then returned after one week to repeat the evaluation using paper scorecards. The majority of panelists were female (90.8%) and between the ages of 18-23 (87.8%).

68 (69.5%) panelists correctly identified the odd sample using the paper scorecard, while 66 (67.3%) panelists correctly identified the odd sample when using the Qualtrics scorecard. Using the Critical Values Table for Sensory Differences (Meilgaard and others 2007), it was determined that panelists were able to detect a difference between the two Cheetos® samples with both the paper and Qualtrics scorecards ($p \leq 0.05$). In the study by Armstrong and others (1997), 8 out of 12 panelists (67%) identified the odd sample in a triangle test when using paper triangle test scorecards to evaluate undiluted and diluted tomato juice samples. A much lower percentage, 33% (4 out of 12), of panelists chose the correct odd sample using computerized triangle test scorecards. Without knowing anything about the format and structure of the computerized triangle test scorecard in Armstrong and others' study, it is hard to know if differences in the scorecards caused the discrepancy in accuracy between the two scorecard versions. The difference in electronic triangle test data may be a reflection of the time periods during which these studies were conducted; the study by Armstrong and others was conducted during the mid 1990's when computers were not as widely used as they are today.

Wilcoxon Signed Rank test indicated that there was no significant difference between responses for the two data collection methods ($p > 0.05$). Armstrong and others (1997) found that 50% of responses for the triangle test were consistent across the paper and computerized scorecards.

Tetrad test

108 panelists completed tetrad tests with undiluted and diluted orange juice samples (Kroger, The Kroger Co., Cincinnati, OH). Panelists used the paper version of the tetrad scorecard during the first sensory panel session and the Qualtrics version during the second session. The majority of panelists were female (90.7%) and between the ages of 18-23 (85.2%).

69 (63.8%) panelists correctly grouped the similar samples using the paper scorecard compared to 70 (64.8%) using the Qualtrics scorecard. According to the Critical Values Table for Sensory Differences (Meilgaard and others 2007), panelists were able to detect a difference between the two orange juice samples using both versions of the tetrad scorecard ($p \leq 0.05$). Wilcoxon Signed Rank test indicated no significant differences in the responses between the two data collection methods ($p > 0.05$).

15-point unstructured descriptive line scale

6 trained panelists evaluated the intensity of texture, odor, and flavor attributes in two dry medium-roasted Nonpareil almond samples stored for 18 months under different conditions (The Almond Board of California, Modesto, CA); one sample was held at 25°C at 65% relative humidity in a polypropylene bag and the other was held at 15°C without humidity control in a high barrier bag. Panelists were trained for 25 contact hours and participated in approximately 30 descriptive panels. Panelists used the paper scorecard during the first sensory panel session and the Qualtrics version during the second session. 5 out of 6 panelists were female (83.3%) and 5 out of 6 (83.3%) were between the ages of 21-30.

Scorecard version did not have a significant effect on panelists' responses for any of the sensory attributes (PROC GLM) ($p > 0.05$), except for painty odor ($p \leq 0.05$) (Tables 4.5-4.7). Almond sample did not have a significant effect on panelists' responses for any of the sensory attributes ($p > 0.05$). The difference in the LS-means for intensity ratings between data collection method for the two almond samples were within 0.40 units of each other and fell within the "Not Perceptable" region of the intensity scale. This indicates that, from session 1 to session 2, panelists rated almond samples similarly regardless of data collection method.

Paired samples t-tests (PROC TTEST) showed no significant differences between data collection methods for 16 out of 18 attribute pairs ($p > 0.05$) (Table 4.8). The intensity ratings for sweetness and painty odor in the high barrier bag sample differed significantly ($p \leq 0.05$), however the LS-means differed less than 1 and 2 scalar units, respectively.

Usability of paper and Qualtrics scorecards

9-point structured hedonic scale scorecard

100% of panelists strongly agreed or agreed that the instructions for the paper scorecard were easy to understand and follow compared to 97.3% of panelists for the Qualtrics scorecard. 98.3% of panelists felt very comfortable or comfortable using the paper scorecard compared to 92.9% for the Qualtrics scorecard. Mean ratings for understandability and comfort increased slightly from session one with paper scorecards (understandability: 1.10 ± 0.30; comfort: 1.16 ± 0.42) to session two with Qualtrics scorecards (understandability: 1.26 ± 0.65; comfort: 1.41 ± 0.77). This demonstrates panelists' comfort with this type of affective scorecard and with its administration through Qualtrics on iPads. The high ratings for these two questions could be due to panelists' familiarity with completing 9-point structured hedonic scales in other sensory evaluation studies.

Additionally, the appearance of the hedonic scales, the instructions, and the actions required to complete the scorecard (mark or check the box that represents how you feel toward the sample) were very similar for both data collection versions. The biggest difference was the spacing of the scales. On the paper scorecard, panelists were able to see all four attribute scales together on the same page; on the Qualtrics scorecard, attribute scales were separated so that panelists only saw one scale per screen. The Qualtrics scorecards were designed this way to

prevent panelists from having to scroll up and down the screen, which could increase the occurrence of user errors such as adjusting the wrong scale response.

9-point Unstructured Hedonic Scale Scorecard

99.1% of panelists strongly agreed or agreed that the instructions for the paper scorecard were easy to understand and follow compared to 94.4% for the Qualtrics scorecard. 92.6% of panelists felt very comfortable or comfortable using the paper scorecard compared to 91.7% of panelists for the Qualtrics scorecard. It should be noted that this was the first time the majority of the panelists were exposed to an unstructured line scale scorecard. The paper and Qualtrics versions of the 9-point unstructured scale scorecard were similar in regards to appearance, instructions, and actions required to complete the scale. Mean ratings for understandability and comfort decreased slightly from the first session with Qualtrics scorecards (understandability: 1.43 ± 0.81; comfort: 1.49 ± 0.83) to the second session with paper scorecards (understandability: 1.25 ± 0.46; comfort: 1.37 ± 0.74).

Ranking test

96.3% of panelists strongly agreed or agreed that the instructions for the paper scorecard were easy to understand and follow compared to 89.3% for the Qualtrics scorecard. 95.3% of panelists felt very comfortable or comfortable using the paper scorecard compared to 87.3% for the Qualtrics scorecard. It should be noted that this was the first time the majority of panelists had evaluated the acceptability of food products using a ranking test. Mean ratings for understandability and comfort increased from the first sensory evaluation session with paper scorecards (understandability: 1.26 ± 0.52; comfort: 1.23 ± 0.59) to the second session with Qualtrics scorecards (understandability: 1.43 ± 0.81; comfort: 1.49 ± 0.89).

Triangle test

Out of all of the scorecards assessed in this study, the paper and Qualtrics triangle sensory test scorecards were the most similar in regards to format, instructions, and actions required to complete the scorecard. 100% of panelists strongly agreed or agreed that the instructions on the paper scorecard were easy to understand and follow compared to 98.98% for the Qualtrics scorecard. 100% of panelists felt very comfortable or comfortable using the paper scorecard compared to 94.9% for the Qualtrics scorecard. The majority of panelists had not participated in a triangle sensory test prior to this study. Mean ratings for both usability questions decreased from the first sensory evaluation session with Qualtrics scorecards (understandability: 1.26 ± 0.46; comfort: 1.34 ± 0.64) to the second session with paper scorecards (understandability: 1.06 ± 0.24; comfort: 1.03 ± 0.17). Panelists in the study by Armstrong and others (1997) reported feeling uncomfortable with computerized discrimination tests, however accuracy and consistency in responses generally improved as the tests progressed.

Tetrad test

97.2% of panelists strongly agreed or agreed that the instructions for the paper tetrad scorecard were easy to understand and follow compared to 98.15% for the Qualtrics scorecard. 95.37% of panelists felt very comfortable or comfortable using the paper scorecard versus 91.67% for the Qualtrics scorecard. The mean response for the understandability question increased slightly from the first session with the paper scorecard (1.29 ± 0.51) to the second session with the Qualtrics scorecard (1.34 ± 0.55); comfort scores also icreased slightly from the first session (1.25 ± 0.64) to the second session (1.43 ± 0.79). This is especially interesting due to the difference in instructions and actions required to complete the two scorecard versions. Most panelists had not been exposed to a tetrad sensory test scorecard prior to this study.

15-point unstructured descriptive line scale

100% of panelists strongly agreed or agreed that the instructions for the paper scorecards were easy to understand and follow compared to 83.3% for the Qualtrics scorecard. 100% of panelists felt very comfortable or comfortable using the paper scorecard compared to 66.7% for the Qualtrics scorecards. Because this was the first panel of the study, issues with the Qualtrics scorecard had to be addressed after some of the panelists completed the scorecard, despite pre-testing. Half of the panelists were asked to repeat both sessions of the descriptive test due to formatting issues with the Qualtrics scorecards, which may have affected these panelists' responses to the usability questionnaire. Additionally, although the trained panelists participated in a Qualtrics orientation, they had no practice with the Qualtrics scorecard prior to its use; conversely, they had extensive experience with the paper scorecard. If this data collection method was planned to be used in the future, practice with the scorecard would be included in the descriptive panel training phase and panelists would practice using the Qualtrics scorecard throughout the training phase of the panel preparation.

Overall usability of Qualtrics scorecards

In general, it appears that panelists found the instructions for the Qualtrics scorecards easy to understand and follow, and felt comfortable using the scorecards. The percentage of panelists who answered strongly agree or agree to the question regarding the understandability of the Qualtrics scorecard instructions or very comfortable or comfortable to the question regarding the level of comfort when using the Qualtrics scorecards never fell below 90% for any sensory test scorecard, except the 15-point unstructured descriptive test scorecard. These results are similar to observations and panelist opinions gathered in the studies by Armstrong and others (1997), Findlay and others (1986), and Chuang and others (2008). Although these studies all

involved the use of scorecards administered on computers, the actions required to complete the scorecards and surveys were similar to those used with the Qualtrics scorecards, such as drag-and-drop actions (Chuang and others, 2008) and using a stylus to enter information and responses using an onscreen keyboard (Findlay and others 1986).

Many clinical studies have explored the use of electronic surveys via iPads and other handheld tablets. Participants in the study by Jeannic and others (2014) preferred electronic case report forms to paper forms. Similarly, 82% of panelists in the study by Bliven and others (2001) preferred electronic versions of two health questionnaires to the paper versions. 89% of the panelists found the electronic versions "very easy" or "easy" to use. Although it cannot be assumed from the data gathered in this study that panelists preferred the Qualtrics scorecards to the paper scorecards, it is clear that they were comfortable with the Qualtrics scorecards. This is reinforced by the statistical analyses discussed above.

The reviews by Wilcox and others (2012) and Shapiro and others (2004) both reported that administrators found that electronic data collection tools are easy to modify and required little training and technical support. From the researcher's perspective, Qualtrics is intuitive and has a plethora of options and features in its survey builder that makes it adaptable for use in sensory evaluation. Unlike Compusense, which was developed specifically for sensory evaluation, Qualtrics is a general survey tool, and therefore requires creativity on the part of the researcher when creating scorecards for some sensory tests. The tetrad test, for example, had to be adapted into a drag-and-drop type question because there appeared to be no way to adapt the multiple choice or text entry question types to fit the grouping action required by the tetrad test. Many of the features touted by Compusense, such as flexible question design, force response, question libraries, time delays, branching, in-program statistical analysis, and the ability to

export raw data, are all available features in Qualtrics (Compusense 2015). Qualtrics also has participant feedback options, which may be able to replicate the actions of Compusense *FCM*®, a panelist training and calibration tool. This feature was not evaluated in this study.

Time and labor associated with data collection and analysis using paper and Qualtrics scorecards

Time and labor requirements were divided into three steps: scorecard development, panel preparation, and data coding and entry (Table 4.9). The time and labor required for each of these steps varied depending on the type of sensory test.

Scorecard development

Generally, scorecard development for paper scorecards included generating scales or recreating ranking and triangle tests from Meilgaard and others (2007) or Kemp and others (2009) and tetrad tests from Rousseau and Ennis (2013) using Word (Microsoft Office: Mac 2011, Redmond, WA). Scorecard development for Qualtrics scorecards included adapting paper scorecards to fit the types of questions available in the survey builder, taking into account how the data would be coded and organized by Qualtrics. The instructions for some of the scorecards required rewording to better fit the question format chosen in Qualtrics, however this was not included in the time required for scorecard development. The time required to create the usability and demographics questionnaire also was not included in this area, nor was the time required to create the Name and End blocks. Once the Name and End blocks were created, they were saved to the Block Library and simply copied over to each scorecard. This is similar to the paper scorecard; the same usability and demographics questionnaires were used throughout the study. Development of paper and Qualtrics versions of the demographics questionnaire required

very minor changes that took a negligible amount of time when adapted for each specific scorecard.

Both versions of the hedonic and descriptive line scales required the most amount of time to develop. The paper and Qualtrics versions of the descriptive line scale required the most time, mainly due to the nine attribute scales associated with each sample. The paper version took approximately 60 minutes to create compared to 90 minutes for the Qualtrics scorecard. Because of the controls and options available in Qualtrics (Force Response, editing the length of the line scales, editing the number of decimals to which Qualtrics will measure the responses), the time needed to create and check over each scale was increased. Subsequent panels using the same type of scorecard will require less time since blocks and questions can be saved to and copied from the Block Library, however scrolling up and down the website page to edit the sample codes, sample attributes, and instructions will likely make line and category scale scorecards more tedious to edit when compared to their paper counterparts. Despite this, the amount of time saved during panel preparation and data coding and entry when using Qualtrics is well worth the extra time during the scorecard development stage. The paper versions of the 9-point structured and unstructured hedonic line scales required 30 minutes to develop. The Qualtrics version of the 9-point structured hedonic line scale required 60 minutes. The time needed for the unstructured line scale was reduced to 45 minutes by copying and editing the structured line scales from the Block Library, illustrating the usability and time-savings associated with this feature of Qualtrics.

The ranking, triangle, and tetrad tests were straightforward to develop on paper and on Qualtrics, and thus required only 10 minutes each. The shortened time period required for development of these scorecards can be attributed to the single rather than multiple questions

typical of hedonic and descriptive tests. The paper and Qualtrics versions of the ranking and triangle test scorecards used the same instructions and scorecard format, however the Qualtrics version of the tetrad test employed a drag-and-drop type question versus the write-in-the-blank format of the paper scorecard, and thus required a slight modification of the instructions.

Panel preparation

As with scorecard development, the time required for panel preparation varied depending on the type of sensory scorecard. Panel preparation for paper scorecards included making copies of the consent forms, sensory scorecard(s), and questionnaires. Consent forms and questionnaires were coded with a panelist code, while sensory scorecard(s) were coded with a panelist code and/or a sample code. Once all of the components were coded, they were paper clipped together in presentation order into packets for ease of handling during the panels. Preparation for Qualtrics scorecards included making paper copies of the consent forms, loading the Qualtrics scorecards onto the iPad home screens, and locking the iPads into their wall-mounted cases.

Panel preparation for the 9-point structured and unstructured scorecards required the most time due to three sensory scorecards per panelist (one for each sample) as opposed to one sensory scorecard per panelist for the ranking, triangle, and tetrad tests. Preparation for these scorecards also included alternating the order of the scorecards in each packet in order to achieve a semi-balanced order of presentation, which increased the length of time needed for panel preparation. Although a fully balanced order of presentation is ideal in sensory testing, it is unlikely that using a semi-balanced order of presentation with the 9-point structured and unstructured hedonic line scales and ranking test caused any errors because this protocol was used with both scorecard versions. Preparation time for the paper version of the 9-point

structured hedonic line scale scorecard was 138 minutes compared to 30 minutes for the Qualtrics scorecard. Panel preparation time associated with the paper version of the 9-point unstructured hedonic line scale was slightly less at 130 minutes, however the time required for the Qualtrics scorecard did not change. Although the paper version of the 15-point unstructured descriptive line scale had more paperwork per panelist than the two hedonic line scale scorecards, the preparation time was only 30 minutes due to the reduced number of panelists. The preparation time for the Qualtrics version of the descriptive scorecard was also 30 minutes.

Panel preparation for the paper versions of the ranking, triangle, and tetrad tests took less time than the paper versions of the hedonic and descriptive line scale scorecards, mainly due to the reduced amount of paperwork required per panelist (one sensory scorecard per panelist as opposed to three). 100 minutes, 88 minutes, and 85 minutes were needed to prepare for the panels using the paper versions of the ranking, triangle, and tetrad scorecards, respectively. The preparation time needed for the panels using the Qualtrics versions of these sensory tests was 30 minutes for each.

Data coding and entry

The time required and methods used for data coding and entry was slightly different for each sensory test. Data coding and entry with Qualtrics scorecards required vastly less time than that required for the paper scorecards. Data coding and entry for paper scorecards included coding the responses on the scorecards and questionnaires, reviewing coded responses for errors, entering the coded responses into Excel (Microsoft Office: Mac 2011, Microsoft, Redmond, WA), and reviewing entered data for errors. For the Qualtrics scorecards, this step included downloading the data file into Excel, deleting unnecessary information (ISP addresses, for example), and relabeling columns for entry into SAS, the statistical analysis program employed.

As with the other steps, the paper versions of the 9-point structured and unstructured scorecards required the most time for data coding and entry. The paper structured line scale scorecard required 223 minutes for coding and data entry, while the Qualtrics scorecard only required 20 minutes. The paper unstructured line scale scorecard required 610 minutes for coding and data entry because every response on every line scale had to be measured and then reviewed for accuracy. The Qualtrics counterpart only required 20 minutes; in Qualtrics, the number of gridlines (either visible or invisible) on the line scale, as well as the number of decimal points to which the response should be carried, can be set prior to collecting data and thus requires no coding or transformation. The responses on the paper version of the 15-point unstructured descriptive line scale scorecards also required measuring and coding, however the total data coding and entry time was only 77 minutes since there were only six descriptive panelists. Organizing the descriptive data from the Qualtrics version of this scorecard took about 30 minutes, which was slightly more time than the 9-point hedonic line scales because of the greater number of sensory attributes per sample (9 sensory attributes per sample as opposed to 4).

The paper version of the ranking scorecard required 63 minutes for coding and data entry compared to 20 minutes for the Qualtrics scorecards. Data coding and entry for the paper version of the triangle test took 62 minutes compared to 35 minutes for the Qualtrics. Because each sensory booth was assigned its own block to achieve a balanced order of presentation, the raw data output from Qualtrics required more time for organization than the hedonic and ranking test data. However, the Qualtrics data were already coded when downloaded, and therefore still required less time than the paper scorecards, which needed to be coded and entered into Excel, as well as checked for errors.

Data coding and entry for the paper tetrad scorecards took 87 minutes compared to 45 minutes for the Qualtrics scorecards. Unlike the other Qualtrics scorecards, the Qualtrics version of the tetrad scorecard required coding, and therefore data checking. Raw data from the drag-and-drop option used for the tetrad tests codes responses as 0 or 1 depending on the box into which the panelist dropped the sample code. As a result, the Excel spreadsheet downloaded from Qualtrics contained a 4-digit series of 0's and 1's (two 0's and two 1's in each series) for each panelist, which could be checked against a key to determine whether each panelist grouped the correct similar samples. Despite having to code the raw Qualtrics data, it was still more straightforward to code than the paper scorecards, which contained six different sets of four 3-digit codes (each sensory booth had its own unique set of codes in order to achieve a balanced order of presentation).

The total time reduction across all steps (scorecard development, panel preparation, and coding and data entry) was at least 50% for all sensory tests assessed except for the descriptive test. Time requirements associated with the 9-point structured and unstructured hedonic line scale were reduced by 71.8% and 87.7%, respectively, with the use of the Qualtrics scorecard. Because of the small number of panelists and subsequent data, the total time requirements associated with the 15-point unstructured descriptive line scale were reduced by 10.1% with the use of the Qualtrics scorecard. The time requirements for the ranking, triangle, and tetrad test scorecards were reduced by 65.3%, 53.1%, and 53.2%, respectively, with the Qualtrics scorecard.

Few studies assessed the amount of time saved during the data coding and entry step. In the review by Lane and others (2006), 3 out of 9 studies reported that electronic data collection methods reduced the time needed to enter data, however only one study provided data to support

that finding, reporting a reduction of 23% in data entry and transfer time. Chuang and others (2008) found that it took 9.52 minutes per questionnaire for data entry compared to no additional time for the electronic scorecards. Fritz and others (2012) estimated that 1.49 minutes per patient were required to input and process data from the paper questionnaires, totaling 61.5 hours (3,690 minutes) per year for 2020 patients. The amount of time saved during data coding and entry through the use of Qualtrics scorecards varied depending on the sensory test, but the most drastic results were seen with the 9-point unstructured line scale scorecard (87.7% time reduction).

Time and labor requirements associated with running panels

Labor and resource requirements varied between data collection methods. Due to the lack of paper handling and the self-guided nature of the Qualtrics scorecards, it was feasible for one researcher to manage six sensory booths alone. However, managing sensory panels using paper scorecards required at least two researchers. This was an unexpected finding, and it could lead to reduced employee labor costs in sensory panels that use Qualtrics scorecards.

Scorecard and questionnaire completion time by panelists was slightly faster for the paper versions of the hedonic and descriptive line scale scorecards. Unlike the scales on the paper scorecards, which were all on one page, the scales on the Qualtrics scorecard were shown one at a time to keep panelists from having to scroll down the screen to reveal additional content before completing all of the scales. Because of this, it took panelists extra time to complete the Qualtrics scorecard. Panelists rarely took longer than 15 minutes to complete either scorecard version for the 9-point structured and unstructured hedonic line scales. Panelists took between 26-35 minutes to complete the paper version of the 15-point descriptive scale scorecard. Qualtrics has the capability to record the amount of time needed to complete a scorecard,

however this function was incorrectly set up for this particular sensory panel, and therefore only an estimated time of 30-40 minutes can be given for the time needed to complete the Qualtrics version of the scorecard. Although panelists took longer to complete the Qualtrics versions of these scorecards, the time saved during panel preparation and data coding and entry, as well as the reduction in errors due to the options for customization with the use of Qualtrics scorecards negated the extra time required in the sensory booths. Although the Qualtrics scorecards required more time from panelists, the responses from the usability questionnaires do not reflect any negative burden on the panelists. Additionally, the eliminated wait time between scorecards and questionnaires likely kept the panelists focused on the sensory evaluation. During paper scorecard sessions, panelists typically passed the wait time between scorecards and questionnaires on their smart phones, which likely removed their concentration from the sensory evaluation.

Conversely, panelists typically completed the ranking, triangle, and tetrad scorecards more quickly than the paper scorecards due to the self-guided nature of the Qualtrics scorecards; once they received their samples through the hatch, the panelists could move through the remainder of the scorecard at their own pace. Panelists rarely took more than 10 minutes to complete either version of any of these sensory tests.

Overall, there was not a noticeable difference in the amount of time it took panelists to complete the paper and Qualtrics versions of any of the sensory test scorecards, aside from the 15-point unstructured descriptive line scale scorecards. Based on these results, it is unlikely that using Qualtrics sensory scorecards via iPads would cause a panel to be extended beyond the typical timeframe required to administer the panel for the number of panelists required for reliable results. The slight increase in the amount of time needed by some panelists to complete

the Qualtrics scorecard was balanced by the efficiency in which researchers could attend to panelists due to the reduction in paper handling. Participants in similar studies typically completed paper surveys and questionnaires two to three times faster than the electronic versions, which were either administered on computers or handheld tablets (Fritz and others 2012; Chuang and others (2008). In the study by Fritz and others (2012), these results could be due to the older average age of the participants (56.5 years) and the fact that iPads were relatively new at the time this study was conducted. In the study by Chuang and others (2008), the three electronic questionnaires were not similar in instruction or action to the paper scorecard; while the paper scorecard only required the panelists to check a rating along a line scale, the electronic methods required drag-and-drop actions to order the stimuli. Because many of the instructions and scorecards used in sensory evaluations have already been validated, it was important for the Qualtrics scorecards to match their traditional paper counterparts as closely as possible, which is likely why there was no noticeable time difference in completion time between the two scorecard methods, in addition to the young average age of the panelists and the widespread familiarity with iPads.

Errors in data collection and analysis using paper and Qualtrics scorecards

Fritz and others (2012) and Chuang and others (2008) also compared the error rates of paper and electronic data collection. Both studies reported that there were no errors in the form of missing data when the electronic data collection methods were used.

Errors rarely occurred during the scorecard development for either type of scorecard; however, when they did, it was easier to correct the error on the Qualtrics scorecards than it was on the paper scorecards. Qualtrics scorecards can be quickly edited through the Survey Editor

tool. Conversely, paper scorecards had to be recopied and recoded when errors occurred. Not only does this increase the amount of time needed to prepare for the panel, it is also costly and wasteful.

During the panel preparation step, the most common errors associated with the paper scorecards were writing the wrong panelist or sample code, or paper clipping the consent forms, sensory scorecards, and questionnaires in the wrong order. Across all sensory tests, this occurred 11 times. Errors did not occur with the Qualtrics scorecards during this step.

Errors occurred with both types of scorecards during the data coding and entry step. With the paper scorecards, there were 13 errors in the form of missing data across all sensory tests. Additionally, there were two instances of panelists leaving the sensory booth before completing all of the sensory scorecards and questionnaires. Although it is not missing data, there were also 51 instances of missing names on paper scorecards and questionnaires across all paper sensory test sessions (approximately 2065 paper scorecards and questionnaires total). In these instances, panelists' scorecards were matched to their scorecards and questionnaires via the panelist code. Missing data due to panelists skipping over a question or missing names were not an issue with Qualtrics scorecards because of the Force Response feature.

There were 25 errors associated with the use of the Qualtrics scorecards across all tests and responses (approximately 539 scorecards total). 12 of these errors occurred during the first panel with the 9-point unstructured line scales due to connectivity issues that were resolved after the panel. 5 errors were related to panelists incorrectly filling out the name and booth number boxes. When this occurs, the sensory scorecard portion of the Qualtrics scorecard is bypassed and panelists are sent straight to the End Block. The inability to catch mistakes as panelists worked through the various components of the sensory evaluation, which can be done with paper

scorecards as panelists pass them back through the hatch, was the biggest disadvantage of using Qualtrics. Unlike paper scorecards, Qualtrics scorecards are not seen by researchers as panelists complete them, which means that panelist errors that cannot be controlled for through the options in the Survey Builder are not caught until after the data are downloaded. In most cases, panelists alerted the researcher when the Qualtrics scorecard jumped from the name and booth number screen to the usability questionnaire, and therefore the researcher was able to assist the panelist in starting a new session with the Qualtrics scorecard. However, this required the researcher to step into the sensory booth, which is not ideal because it disturbs other panelists and increases interaction between the panelist and researcher albeit not during actual sample evaluation. Further investigation of the multitude of options in Qualtrics' Survey Builder for a potential solution to this issue is needed. There were 8 instances of missing data. Two of these instances were related to panelists not filling in the booth number text entry box appropriately, which resulted in the scorecard skipping directly the usability questionnaire. These particular panelists did not alert the researchers that the sensory evaluation portion was skipped, and simply submitted the usability and demographic portions of the Qualtrics scorecard. The remaining missing data errors could not be attributed to a specific cause. Random errors or glitches resulting in lost data are a risk with any electronic program, but the efficiency and lack of errors during the panel preparation and data coding and entry stages justifies the occasional error that occurred during data collection. This can be compensated for with the number of panelists recruited.

Monetary costs

Qualtrics Research Suite is already used by a number of universities, and because it is a general survey tool, it is more likely to be purchased on a university-wide level when compared to a specialized program developed for sensory research (Qualtrics 2015). This means that programs like Compusense would likely have to be purchased by individual academic departments or individual labs within a department, whereas general survey programs like Qualtrics could be used in individual labs at no additional cost.

Despite high upfront costs, Qualtrics scorecards are cost-effective over time due to reduced labor and paper costs. Fritz and others (2012) found a cost reduction of $149 - $1,243 depending on the type of personnel employed for data entry. The cost to outfit six sensory booths with iPad tablets with screen protectors, wall mounts, surge protectors, and styluses was approximately $3,204. However, after those initial set-up costs, paper supply costs for consent forms were $55.60 (1112 paper copies of consent forms across all six sensory panels at $0.05/copy) for the six sensory panels. In comparison, the paper costs for the six paper scorecard sessions totaled approximately $170.80 (based on 3416 paper copies of consent forms, scorecards, and questionnaires across all six sensory panels at $0.05/copy). In addition, a lab assistant was paid to assist with the last three sensory panels, which lasted 8 hours each; this totaled to $198 in labor costs. Although the lab assistant was paid to assist with the Qualtrics scorecard sessions as well, it would be feasible for one researcher to manage a panel using Qualtrics scorecards instead of paper, which would cut down on labor costs. In a university setting, the amount of time saved and errors reduced rather than monetary savings may be the deciding factors when deciding if using Qualtrics scorecards is justified.

Table 4.1 Paper and Qualtrics[a] scorecards: LS-Means and standard error of acceptability for appearance, texture, flavor, and overall acceptability[b]; oatmeal cookie brand effects[c], data collection method effects[c], and interaction effects[c] (n=113)

		Appearance	Texture	Flavor	Overall Acceptability
		LS-Means ± SE			
Oatmeal Cookie Brand	Kroger[f]	4.63a ± 0.12	5.55a ± 0.12	5.98a ± 0.12	5.71a ± 0.12
	Homekist[g]	5.59b ± 0.12	3.97b ± 0.12	5.11b ± 0.12	4.77b ± 0.12
	Voortman's[h]	4.99c ± 0.12	4.42c ± 0.12	4.72c ± 0.12	4.65a ± 0.12
Data Collection Method	Paper	5.19d ± 0.10	4.63d ± 0.10	5.38d ± 0.11	5.08d ± 0.10
	Qualtrics	4.95d ± 0.10	4.67d ± 0.10	5.16d ± 0.11	5.01d ± 0.10

		Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Brand x Method	Kroger	4.49e ± 0.17	4.77e ± 0.17	5.62g ± 0.18	5.48g ± 0.18	6.04e ± 0.18	5.92e ± 0.18	5.71e ± 0.17	5.72e ± 0.17
	Homekist	5.69f ± 0.17	5.50f ± 0.17	4.19e ± 0.18	3.75e ± 0.18	5.20e ± 0.18	5.02e ± 0.18	4.92e ± 0.17	4.61e ± 0.17
	Voortman's	5.39f ± 0.17	4.59e ± 0.17	4.08e ± 0.18	4.77f ± 0.18	4.91e ± 0.18	4.53e ± 0.18	4.60e ± 0.17	4.70e ± 0.17

[a] Qualtrics Research Suite, Provo UT

[b] Using a repeated measures study design, panelists first assessed oatmeal cookie samples using paper scorecards, and then returned one week later to repeat the assessment using Qualtrics scorecards on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[c] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the main effect

[e] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the brand and data collection method interaction

[f] Psst!, The Kroger Co., Cincinnati, OH

[g] Homekist, South Street Bakery Co., Massillon, OH

[h] Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario

Table 4.2 Assessment of appearance, texture, flavor, and overall acceptability in three commercial oatmeal cookies using paper and Qualtrics[a] versions of 9-point structured hedonic scale scorecard[b] (n=113) [c]

	LS-means ± standard error[d]					
	Kroger[e]		Homekist[f]		Voortman[g]	
Attribute	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Appearance	4.49a ± 0.17	4.77a ± 0.17	5.69a ± 0.17	5.50a ± 0.17	5.39a ± 0.17	4.59b ± 0.17
Texture	5.62a ± 0.18	5.48a ± 0.18	4.19a ± 0.18	3.75a ± 0.18	4.08a ± 0.18	4.77b ± 0.18
Flavor	6.04a ± 0.18	5.92a ± 0.18	5.20a ± 0.18	5.02a ± 0.18	4.91a ± 0.18	4.53a ± 0.18
Overall Acceptability	5.71a ± 0.17	5.72a ± 0.17	4.92a ± 0.17	4.61a ± 0.17	4.60a ± 0.17	4.70a ± 0.17

[a] Qualtrics Research Suite, Provo UT

[b] Sensory attributes were evaluated on a 9-point scale where 1=extremely dislike and 9=extremely like

[c] Using a repeated measures study design, panelists first assessed oatmeal cookie samples using paper scorecards, and then returned one week later to repeat the assessment using Qualtrics scorecards on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[d] LS-means ± SE followed by the same letter within oatmeal cookie brand and between data collection methods are not significantly different according to a paired samples t-test ($p > 0.05$) (SAS University Edition, Cary, NC)

[e] Psst!, The Kroger Co., Cincinnati, OH

[f] Homekist, South Street Bakery Co., Massillon, OH

[g] Voortman Sugar Free, Voortman Cookies Ltd., Burlington, Ontario

Table 4.3 Paper and Qualtrics[a] scorecards: LS-Means and standard error for acceptability of appearance, texture, flavor, and overall acceptability[b]; vanilla yogurt variety effects[c], data collection method effects[c], and interaction effects[c] (n=108)

		Appearance	Texture	Flavor	Overall Acceptability
		LS-Means ± SE			
Vanilla yogurt variety[f]	Carbmaster	6.19a ± 0.12	6.35a ± 0.13	5.02a ± 0.15	5.54a ± 0.13
	Greek	4.69b ± 0.12	5.39b ± 0.13	4.53b ± 0.15	4.77a ± 0.13
	Low-fat	5.33c ± 0.12	5.62b ± 0.13	5.63c ± 0.15	5.68b ± 0.13
Data Collection Method	Paper	5.58d ± 0.10	5.82d ± 0.11	5.08d ± 0.12	5.42d ± 0.11
	Qualtrics	5.23e ± 0.10	5.75d ± 0.11	5.04d ± 0.12	5.23d ± 0.11

		Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Variety x Method	Carbmaster	6.31f ± 0.17	6.07f ± 0.17	6.31f ± 0.18	6.38f ± 0.18	5.00f ± 0.21	5.04f ± 0.21	5.61f ± 0.19	5.48f ± 0.19
	Greek	4.88f ± 0.17	4.51f ± 0.17	5.46f ± 0.18	5.31f ± 0.18	4.71f ± 0.21	4.35f ± 0.21	5.04f ± 0.19	4.50f ± 0.19
	Low-fat	5.54f ± 0.17	5.13f ± 0.17	5.68f ± 0.18	5.56f ± 0.18	5.53f ± 0.21	5.74f ± 0.21	5.63f ± 0.19	5.72f ± 0.19

[a] Qualtrics Research Suite, Provo UT

[b] Using a repeated measures study design, panelists first assessed vanilla yogurt samples using Qualtrics scorecards on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA), and then returned one week later to repeat the assessment using paper scorecards

[c] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the main effect

[e] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the variety and data collection method interaction

[f] Kroger, The Kroger Co., Cincinnati, OH

Table 4.4 Assessment of appearance, texture, flavor, and overall acceptability in three commercial vanilla yogurts using paper and Qualtrics[a] versions of 9-point unstructured hedonic scale scorecard[b] (n=108) [c]

	LS-means ± standard error[d]					
	Carbmaster[e]		Greek[e]		Low-Fat[e]	
Attribute	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Appearance	6.31a ± 0.17	6.07a ± 0.17	4.88a ± 0.17	4.51a ± 0.17	5.54a ± 0.17	5.13b ± 0.17
Texture	6.31a ± 0.18	6.38a ± 0.18	5.46a ± 0.18	5.31a ± 0.18	5.68a ± 0.18	5.56a ± 0.18
Flavor	5.00a ± 0.21	5.04a ± 0.21	4.71a ± 0.21	4.35a ± 0.21	5.53a ± 0.21	5.74a ± 0.21
Overall Acceptability	5.61a ± 0.19	5.48a ± 0.19	5.04a ± 0.19	4.50b ± 0.19	5.63a ± 0.19	5.72a ± 0.19

[a] Qualtrics Research Suite, Provo UT

[b] Sensory attributes were evaluated on a 9-point scale where 1=extremely dislike and 9=extremely like

[c] Using a repeated measures study design, panelists first assessed vanilla yogurt samples using Qualtrics scorecards on iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA), and then returned one week later to repeat the assessment using paper scorecards

[d] LS-means ± SE followed by the same letter within yogurt brand and between data collection methods are not significantly different according to a paired samples t-test ($p > 0.05$) (SAS University Edition, Cary, NC)

[e] The Kroger Co., Cincinnati, OH

Table 4.5 Paper and Qualtrics[a] scorecards: LS-Means and standard error of intensity of texture and taste attributes[b] in two dry medium-roasted Nonpareil almond sample[c]; storage bag effects[d], data collection method effects[d], and interaction effects[f] (n=6)

		Hardness		Sweetness		Crunchiness	
		LS-Means ± SE					
Storage Bag	HBB[g]	10.58a ± 0.44		0.73a ± 0.27		6.14a ± 0.37	
	Polypropylene[h]	10.55a ± 0.46		0.83a ± 0.28		5.66a ± 0.39	
Data Collection Method	Paper	11.18b ± 0.44		0.75b ± 0.27		6.22b ± 0.39	
	Qualtrics	9.95b ± 0.46		0.80b ± 0.28		5.58b ± 0.37	
		Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Storage Conditions x Method	HBB	11.33c ± 0.62	9.82c ± 0.62	1.28c ± 0.38	0.17c ± 0.38	6.45c ± 0.53	5.83c ± 0.53
	Polypropylene	11.02c ± 0.62	10.07c ± 0.68	0.32c ± 0.42	1.33c ± 0.38	5.98c ± 0.58	5.33c ± 0.53

[a] Qualtrics Research Suite, Provo UT

[b] Using a repeated measures study design, panelists first assessed almond samples using Qualtrics scorecards on paper scorecards, and then returned 1 – 24 hours later to repeat the assessment using iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[c] Intensity of sensory attributes were evaluated on a 15-point scale where 0=Not perceptible and 15=High intensity

[d] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the main effect

[f] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the storage conditions and data collection method interaction

[g] Almond samples (The Almond Board of California, Modesto, CA) were held at 25°C at 65% relative humidity in a polypropylene bag

[h] Almond samples (The Almond Board of California, Modesto, CA) were held at 15°C without humidity control in a high barrier bag

Table 4.6 Paper and Qualtrics[a] scorecards: LS-Means and standard error of intensity of odor rancidity attributes[b] in two dry medium-roasted Nonpareil almond sample[c]; storage bag effects[d], data collection method effects[d], and interaction effects[f] (n=6)

		Cardboard		Rancid		Painty	
		LS-Means ± SE					
Storage Bag	HBB[g]	0.78a ± 0.66		2.57a ± 0.86		0.37a ± 0.22	
	Polypropylene[h]	1.68a ± 0.66		2.92a ± 0.86		0.59a ± 0.21	
Data Collection Method	Paper	2.00b ± 0.63		4.04b ± 0.82		0.68b ± 0.21	
	Qualtrics	0.46b ± 0.69		1.45b ± 0.90		0.28c ± 0.22	
		Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Storage Conditions x Method	HBB	1.08c ± 0.89	0.48c ± 0.98	3.28c ± 1.16	1.86c ± 1.27	0.47e ± 0.30	0.26e ± 0.33
	Polypropylene	2.92c ± 0.89	0.44c ± 0.98	4.80c ± 1.16	1.04c ± 1.27	0.88de ± 0.30	0.30d ± 0.30

[a] Qualtrics Research Suite, Provo UT

[b] Using a repeated measures study design, panelists first assessed almond samples using Qualtrics scorecards on paper scorecards, and then returned 1 – 24 hours later to repeat the assessment using iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[c] Intensity of sensory attributes were evaluated on a 15-point scale where 0=Not perceptible and 15=High intensity

[d] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the main effect

[f] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the storage conditions and data collection method interaction

[g] Almond samples (The Almond Board of California, Modesto, CA) were held at 25°C at 65% relative humidity in a polypropylene bag

[h] Almond samples (The Almond Board of California, Modesto, CA) were held at 15°C without humidity control in a high barrier bag

Table 4.7 Paper and Qualtrics[a] scorecards: LS-Means and standard error of intensity of flavor rancidity attributes[b] in two dry medium-roasted Nonpareil almond sample[c]; storage bag effects[d], data collection method effects[d], and interaction effects[f] (n=6)

		Cardboard		Rancid		Painty	
		LS-Means ± SE					
Storage Bag	HBB[g]	1.32a ± 0.37		4.86a ± 0.99		1.37a ± 0.41	
	Polypropylene[h]	0.99a ± 0.39		3.15a ± 1.04		0.66a ± 0.43	
Data Collection Method	Paper	1.58b ± 0.37		4.61b ± 0.99		1.31b ± 0.43	
	Qualtrics	0.74b ± 0.39		3.40b ± 1.04		0.72b ± 0.43	
		Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Storage Conditions x Method	HBB	1.65c ± 0.50	0.99c ± 0.55	4.03c ± 1.41	5.68c ± 1.41	1.33c ± 0.58	1.33c ± 0.58
	Polypropylene	1.50c ± 0.55	0.48c ± 0.55	5.18c ± 1.41	1.12c ± 1.54	1.28c ± 0.58	0.04c ± 0.63

[a] Qualtrics Research Suite, Provo UT

[b] Using a repeated measures study design, panelists first assessed almond samples using Qualtrics scorecards on paper scorecards, and then returned 1 – 24 hours later to repeat the assessment using iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[c] Intensity of sensory attributes were evaluated on a 15-point scale where 0=Not perceptible and 15=High intensity

[d] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the main effect

[f] LS-Means within sensory attribute followed by the same letter are not significantly different ($p > 0.05$) according to two-way ANOVA and LS-Means separation with PDIFF (SAS University Edition, Cary, NC) for the storage conditions and data collection method interaction

[g] Almond samples (The Almond Board of California, Modesto, CA) were held at 25°C at 65% relative humidity in a polypropylene bag

[h] Almond samples (The Almond Board of California, Modesto, CA) were held at 15°C without humidity control in a high barrier bag

Table 4.8 Assessment of the intensity of texture, odor, and flavor rancidity attributes in two dry medium-roasted Nonpareil almond samples using paper and Qualtrics[a] versions of 15-point unstructured descriptive scale scorecard[b] (n=6) [c]

	LS-means ± standard error[d]			
	High Barrier Bag[e]		Polypropylene Bag[f]	
Attribute	Paper	Qualtrics	Paper	Qualtrics
Hardness	11.33a ± 0.62	9.82a ± 0.62	11.02a ± 0.62	10.07a ± 0.68
Sweetness	1.28a ± 0.38	0.17b ± 0.38	0.32a± 0.42	1.33a ± 0.38
Crunchiness	6.45a ± 0.53	5.83a ± 0.53	5.98a± 0.58	5.33a ± 0.53
Cardboard Odor	1.08a ± 0.89	0.48a ± 0.98	2.92a ± 0.89	0.44a ± 0.98
Rancid Odor	3.28a ± 1.16	1.86a ± 1.27	4.80a ± 1.16	1.04a ± 1.27
Painty Odor	0.47a ± 0.30	0.26b ± 0.33	0.88a ± 0.30	0.30a ± 0.30
Cardboard Taste	1.65a ± 0.50	0.99a ± 0.55	1.50a ± 0.55	0.48a ± 0.55
Rancid Taste	4.03a ± 1.41	5.68a ± 1.41	5.18a ± 1.41	1.12a ± 1.54
Painty Taste	1.33a ± 0.58	1.41a ± 0.58	1.28a ± 0.58	0.04a ± 0.63

[a] Qualtrics Research Suite, Provo UT

[b] Intensity of sensory attributes were evaluated on a 15-point scale where 0=Not perceptible and 15=High intensity

[c] Using a repeated measures study design, panelists first assessed almond samples using Qualtrics scorecards on paper scorecards, and then returned 1 – 24 hours later to repeat the assessment using iPads (Apple iPad with Retina Display MD510LL/A 4th Generation, Apple Inc., Cupertino, CA)

[d] LS-means ± SE followed by the same letter within almond sample and between data collection methods are not significantly different according to a paired samples t-test ($p > 0.05$) (SAS University Edition, Cary, NC)

[e] Almond samples (The Almond Board of California, Modesto, CA) were held at 25°C at 65% relative humidity in a polypropylene bag

[f] Almond samples (The Almond Board of California, Modesto, CA) were held at 15°C without humidity control in a high barrier bag

Table 4.9 Time (in minutes) requirements for data collection and analysis using paper and Qualtrics[a] versions of sensory tests representative of each sensory test class

	9-point structured hedonic line scale		9-point unstructured hedonic line scale		15-point unstructured descriptive line scale		Ranking Test		Triangle Test		Tetrad Test	
Step	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics	Paper	Qualtrics
Scorecard Development	30	60	30	45	60	90	10	10	10	10	10	10
Panel Preparation[b]	138	30	130	30	30	30	100	30	88	30	85	30
Coding and Data Entry[c]	223	20	610	20	77	30	63	20	62	35	87	45
Total time reduction across all steps with use of Qualtrics scorecards	71.8%		87.7%		10.1%		65.3%		53.1%		53.2%	

[a] Qualtrics Research Suite, Provo, UT

[b] Panel preparation for paper scorecards included copying consent forms, scorecards, and questionnaires; writing sample and panelist codes on scorecards and questionnaires; paper clipping consent forms, scorecards, and questionnaires together. Preparation for Qualtrics scorecards included copying consent forms; loading Qualtrics scorecards on iPad home screen.

[c] Coding and data entry for paper scorecards included coding responses, entering responses into Excel (Microsoft Office: Mac 2011, Microsoft, Redmond, WA), and reviewing both steps for errors. Coding and data entry for Qualtrics scorecards included downloading data file into Excel; cleaning up data file.

References

Armstrong G, McIlveen H, McDowell D, Blair I. 1997. Sensory analysis and assessor motivation: can computers make a difference? Food Quality and Preference 8(1):1-7.

Bliven BD, Kaufman SE, Spertus JA. 2001. Electronic collection of health-related quality of life data: Validity, time benefits and patient preference. Quality of Life Research 10(1):15-21.

Chuang Y, Chen L, Chuang M. 2008. Computer-based rating method for evaluating multiple stimuli on multiple scales. Computers in Human Behavior 24(5):1929-1946.

Compusense. 2015. Compusense Five Software Features. Retrieved May 7, 2015 from Compusense on the World Wide Web: https://www.compusense.com/files/7614/0237/7890/c5_features.pdf

Findlay C. 1986. Computers and the Internet in sensory quality control. Food Quality and Preference 13(6):423-428.

Fritz F, Balhourn S, Riek M, Breil B, Dugas M. 2012. Qualitative and quantitative evaluation of HER-integrated mobile patient questionnaires regarding usability and cost-efficiency. International Journal of Healthcare Informatics 81(5):303-13.

Harrison AA, Tutone RM, McFagden DG. 1971. Effects of frequency of exposure of changing and unchanging stimulus pairs on affective ratings. Journal of Personality and Social Psychology 20(1): 102-111.

Jeannic A, Quelen C, Alberti C, Durand-Zaleski I. 2014. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Medical Research Methodology 14(1):1-24.

Lane SJ, Heddle NM, Arnold E, Walker I. 2006. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods of data collection. BMC Medical Informatics and Decision Making 6(23):23-33.

Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell.

Meilgaard MC, Civille GV, Carr BT. 2007. Sensory Evaluation Techniques. Boca Raton, FL: CRC Press. p. 255-309.

Plemmons LE and Resurreccion AVA. 1999. Note on computerized data collection in consumer sensory evaluation. Journal of Sensory Studies14(1):19-34.

Qualtrics. 2015. Qualtrics Research Suite. Retrieved May 7, 2015 from Qualtrics on the World Wide Web: http://www.qualtrics.com/research-suite/

Rousseau B and Ennis JM. 2013. Importance of correct instructions in the tetrad test. Journal of Sensory Studies 28(4): 264-269.

Shapiro JS, Bessette MJ, Baumlin KM, Ragin DF, Richardson LD. 2004. Automating research data collection. Academic Emergency Medicine 11(11):1223-28.

Wilcox AB, Gallagher KD, Boden-Albala B, Bakken SR. 2012. Research data collection methods from paper to tablet computers. Medical Care 50(7):S68-73.

CHAPTER 5: SUMMARY AND CONCLUSIONS

Despite its simplicity and low initial costs in equipment, training, and support (Wilcox and others 2012), paper-based data collection requires manual data entry that is time-consuming and subject to human error (Shapiro and others 2004; Fritz and others 2012). Potential alternative data collection methods include using tablets and Internet-based forms. Studies in clinical settings revealed that these methods are user-friendly, more time-efficient, and more accurate than paper-based methods (Wilcox and others 2012; Shapiro and others 2004). Research concerning the use of iPad and Internet-based data collection in food sensory analysis is limited, but computers have produced consistent panelist responses using an intuitive evaluation system (Plemmons and Resurreccion 1999; Armstrong and others 1997).

Electronic data collection methods have been shown to generate responses that are consistent with their paper-based counterparts. Plemmons and Resurreccion (1999) evaluated three food products using a 9-point hedonic scale on computer scorecards administered via Compusense *five* computer software (Compusense Inc.) and paper scorecards. A paired samples t-test showed that there were no significant differences ($p > 0.05$) between computer and paper scorecards for all three attribute ratings (overall acceptance, appearance, and flavor) in two out of the three food products. Armstrong and others (1997) compared paper and computerized scorecards for various types of sensory tests from all three classes of sensory evaluation tests, including triangle, hedonic descriptive, and ranking acceptance tests.

Panelists in a number of sensory studies reported feeling comfortable using electronic data collection methods to evaluate food products and other stimuli (Armstrong and others 1997; Findlay and others 1986; Chuang and others 2008). In many clinical studies, similar results were found (Jeannic and others 2014; Bliven and others 2001). Patients in the study by Jeannic and others (2014) preferred electronic case report forms to their paper counterparts. The majority of participants in the study by Bliven and others (2001) preferred the electronic versions of two health questionnaires to the paper versions. Additionally, reviews by Shapiro and others (2004) and Wilcox and others (2012) reported the ease with which administrators can format and modify Internet-based forms and other electronic data collection methods to prevent incomplete and incorrect answers with little training and technical support.

Manual data entry, which is a significant source of human error in paper-based data collection, is eliminated in electronic data collection (Shapiro and others 2004). Several studies compared the error rates of electronic and paper data collection, which included both incorrect or incomplete entries and errors made during data transfer, and found that electronic methods have a lower error frequency rate (Fritz and others 2012; Chuang and others 2008). The elimination of manual data entry also substantially reduces the time requirements associated with paper-based data collection. Although participants in studies by Fritz and others (2012) and Chuang and others (2008) completed paper-based questionnaires faster than iPad and computer-based questionnaires, the time required to process the paper-based data ultimately made the electronic methods more efficient. The reduction in processing time also translated to greater cost-efficiency with the electronic data collection methods (Fritz and others 2012).

Although tablets are not as widely owned as computers, consumers generally find iPads easy to use due to widespread familiarity with Apple iOS on iPods and iPhones (Mintel 2013).

iPads are portable, relatively inexpensive, and can mimic the action of marking on paper through the use of a stylus. Qualtrics Research Suite is an online survey platform that allows users to design and administer online surveys, and then analyze collected data through the Qualtrics program or through other statistical software such as SPSS or SAS. Additionally, Qualtrics is used by a number of universities (Qualtrics 2015), and therefore may be more accessible to university-level sensory labs than specialized programs such as Compusense.

The overall objective of this study was to determine the validity of Qualtrics sensory scorecards administered via iPads as a data collection and organization method in food sensory analysis. The concurrent validity of Qualtrics scorecards was evaluated, using paper scorecards as the criterion measure. Additionally, the usability, error rate, and time requirements associated with both paper and Qualtrics sensory scorecards were assessed. A counterbalanced, repeated measures design was used to compare paper and Qualtrics sensory scorecards in six tests representative of each class of sensory test. 9-point structured (n=113) and unstructured hedonic line scales (n = 108) and a ranking test (n = 107) represented affective sensory tests. Triangle (n = 98) and tetrad tests (n = 108) represented discrimination tests. 15-point unstructured line scales using predefined and standardized lexicons, as part of the Spectrum-like Descriptive Analysis Method™ (Meilgaard and others 2007), represented descriptive tests. Panelists also completed questionnaires regarding the usability of both data collection methods.

Paired samples t-tests showed panelists were consistent in their responses using the two data collection methods. 10 out of 12 pairs of attributes in the 9-point structured and unstructured line scale scorecards were not significantly different ($p > 0.05$). The 15-point unstructured descriptive line scale showed similar results with 16 out of 18 attribute pairs not significantly different ($p > 0.05$). Additionally, the difference in the LS-means for sensory

attribute ratings between data collection methods for all sensory tests evaluated in this study were within 1.5 units of each other and fell within the same range of their respective line scales. Friedman's Test showed no significant difference in the distribution of ranking scores between the two ranking test scorecard versions ($p > 0.05$). Wilcoxon Signed Rank test indicated that there was no significant difference in the responses between the two scorecard versions for both the triangle and tetrad tests ($p > 0.05$). These results indicate that panelists' responses were similar across both sessions for each sensory test evaluated, regardless of scorecard version. Further, these results support the concurrent validity of Qualtrics scorecards for use with the sensory tests evaluated in this study.

The usability questionnaires revealed that panelists found the instructions for both scorecard versions for every sensory test easy to understand and follow and comfortable to use. The percentage of panelists who answered strongly agree or agree to the question regarding the understandability of the Qualtrics scorecard instructions or very comfortable or comfortable to the question regarding the level of comfort when using the Qualtrics scorecards never fell below 90% for any sensory test scorecard, except the 15-point unstructured descriptive test scorecard. However, the descriptive test scorecard was the first Qualtrics scorecard evaluated in this study, and experienced formatting errors that were rectified for subsequent sensory tests, but these errors nevertheless may have negatively affected the trained panelists' opinion of the Qualtrics scorecard's usability. In addition, these trained panelists employed the Qualtrics scorecard at the conclusion of a 14-month study throughout which paper scorecards had been used.

Reductions in time and labor associated with data collection and analysis through the use of Qualtrics scorecards were most pronounced for the panel preparation and data coding and entry steps. In general, panelists completed both scorecard versions within the same time frame,

however the number of researchers needed to manage six sensory booths was reduced with the Qualtrics scorecards due to reduced paper handling. Across all sensory tests evaluated in this study, there was a 68.8% reduction in time required for scorecard development, panel preparation, and data coding and entry with the use of the Qualtrics scorecards.

Qualtrics scorecards did not have a clear advantage over paper scorecards in terms of error rate and monetary costs. Although Qualtrics could prevent panelists from accidentally skipping questions, it could not prevent panelists from incorrectly filling out text entry boxes or leaving the sensory booths before finishing the scorecard. Additionally, random errors resulting in lost data are a risk with any electronic program. However, Qualtrics had a lower error rate during the panel preparation step compared to the paper scorecards and contributed no errors during the data entry steps. Although Qualtrics did not have an overall lower error rate compared to paper scorecards, the efficiency and lack of errors during the panel preparation and data coding and entry stages justifies the use of Qualtrics scorecards in place of their paper counterparts during future sensory evaluations.

This study was one of the few that has explored the use of electronic data collection methods in food sensory analysis. Previous sensory studies have focused on the use of electronic sensory scorecards on computers; however, most of these studies are outdated. This study assessed several factors that have not been considered together in one study: concurrent validity, usability, time requirements, error rate, and monetary costs of electronic data sensory scorecards in comparison to traditional paper sensory scorecards. The results provide valuable information for other researchers to consider, particularly those in the university setting, when deciding whether or not to use generic survey software such as Qualtrics Research Suite to collect food sensory analysis data.

References

Armstrong G, McIlveen H, McDowell D, Blair I. 1997. Sensory analysis and assessor motivation: can computers make a difference? Food Quality and Preference 8(1):1-7.

Bliven BD, Kaufman SE, Spertus JA. 2001. Electronic collection of health-related quality of life data: Validity, time benefits and patient preference. Quality of Life Research 10(1):15-21.

Chuang Y, Chen L, Chuang M. 2008. Computer-based rating method for evaluating multiple stimuli on multiple scales. Computers in Human Behavior 24(5):1929-1946.

Findlay C. 1986. Computers and the Internet in sensory quality control. Food Quality and Preference 13(6):423-428.

Fritz F, Balhourn S, Riek M, Breil B, Dugas M. 2012. Qualitative and quantitative evaluation of HER-integrated mobile patient questionnaires regarding usability and cost-efficiency. International Journal of Healthcare Informatics 81(5):303-13.

Jeannic A, Quelen C, Alberti C, Durand-Zaleski I. 2014. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Medical Research Methodology 14(1):1-24.

Meilgaard MC, Civille GV, Carr BT. 2007. The Spectrum™ Descriptive Analysis Method. 4th. ed. In: Sensory Evaluation Techniques. Boca Raton, FL: CRC Press. p. 189-192.

Mintel. 2013. Tablets – Executive summary. Retrieved January 28, 2014 from Mintel Academic: http://academic.mintel.com.proxy-remote.galib.uga.edu/display/637690/.

Plemmons LE and Resurreccion AVA. 1999. Note on computerized data collection in consumer sensory evaluation. Journal of Sensory Studies14(1):19-34.

Qualtrics. 2015. Qualtrics Research Suite. Retrieved May 7, 2015 from Qualtrics on the World Wide Web: http://www.qualtrics.com/research-suite/

Shapiro JS, Bessette MJ, Baumlin KM, Ragin DF, Richardson LD. 2004. Automating research data collection. Academic Emergency Medicine 11(11):1223-28.

Wilcox AB, Gallagher KD, Boden-Albala B, Bakken SR. 2012. Research data collection methods from paper to tablet computers. Medical Care 50(7):S68-73.

REFERENCES

Amerine MA, Pangborn RM, Roessler EB. 1965. Principles of Sensory Evaluation of Food, New York: Academic Press. p. 549).

Anonymous. 1975. Minutes of division business meeting. Institute of Food Technologists – Sensory Evaluation Division, IFT, Chicago, IL. Cited in: Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell. p. 1.

Armstrong G, McIlveen H, McDowell D, Blair I. 1997. Sensory analysis and assessor motivation: can computers make a difference? Food Quality and Preference 8(1):1-7.

Bliven BD, Kaufman SE, Spertus JA. 2001. Electronic collection of health-related quality of life data: Validity, time benefits and patient preference. Quality of Life Research 10(1):15-21.

Cheely AC. 2014. Consumer Acceptability of Dry Medium-Roasted Almonds Stored Under Varying Storage Conditions (Master's Thesis). University of Georiga.

Compusense. 2015. Compusense Five Software Features. Retrieved May 7, 2015 from Compusense on the World Wide Web: https://www.compusense.com/files/7614/0237/7890/c5_features.pdf

Chuang Y, Chen L, Chuang M. 2008. Computer-based rating method for evaluating multiple stimuli on multiple scales. Computers in Human Behavior 24(5):1929-1946.

Dove WE. 1946. Developing food acceptance research. Science 103: 187.

Ennis JM. 2012. 2012. Guiding the switch from triangle testing to tetrad testing. Journal of Sensory Studies 27(4):223-231.

Ennis JM and Christensen RHB. 2014. Precision of measurement in Tetrad testing. Food Quality and Preference 32(SI):98-106.

Findlay C. 1986. Computers and the Internet in sensory quality control. Food Quality and Preference 13(6):423-428.

Fritz F, Balhourn S, Riek M, Breil B, Dugas M. 2012. Qualitative and quantitative evaluation of HER-integrated mobile patient questionnaires regarding usability and cost-efficiency. International Journal of Healthcare Informatics 81(5):303-13.

Harrison AA, Tutone RM, McFagden DG. 1971. Effects of frequency of exposure of changing and unchanging stimulus pairs on affective ratings. Journal of Personality and Social Psychology 20(1): 102-111.

Ishii R, O'Mahoney M, Rousseau B. Triangle and tetrad protocols: Small sensory differences, resampling and consumer relevance. Food Quality and Preference 31:49-55.

Jeannic A, Quelen C, Alberti C, Durand-Zaleski I. 2014. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Medical Research Methodology 14(1):1-24.

Kemp SE, Hollowood T, Hort J. 2009. Sensory Evaluation: A Practical Handbook. Ames, IA: Wiley-Blackwell.

Lane SJ, Heddle NM, Arnold E, Walker I. 2006. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods of data collection. BMC Medical Informatics and Decision Making 6(23):23-33.

Larmond E. 1977. Methods for Sensory Testing: Preference Tests. In: Laboratory Methods for Sensory Evaluation of Food. Winnipeg, Canada: Kromar Printing Ltd. p. 55-59.

Lawless HT, Heymann H. 2010. Sensory Evaluation of Food. New York, NY: Springer.

Lim J. 2011. Hedonic scaling: A review of methods and theory. Food Quality and Preference 22(8):733-747.

Meilgaard M, Civille GV, Carr BT. 1991. Sensory Evaluation Techniques. 2nd ed. Florida: CRC Press. 354 p.

Meilgaard MC, Civille GV, Carr BT. 2007. Sensory Evaluation Techniques, 4th ed. Boca Raton, FL: CRC Press.

Mintel. 2013. Tablets – Executive summary. Retrieved January 28, 2014 from Mintel Academic: http://academic.mintel.com.proxy-remote.galib.uga.edu/display/637690/.

Muñoz AM and Civille GV. 1998. Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. Journal of Sensory Studies 13(1):57-75.

Murray JM, Delahunty CM, and Baxter IA. 2001. Descriptive sensory analysis: Past, present, and future. Food Research International 34(6):461-471.

Pangborn RM. 1964. Sensory evaluation of food: A look backward and forward. Food Technology 18:1309. Cited in Meilgaard MC, Civille GV, Carr BT. 2006. Sensory Evaluation Techniques. p. 1.

Plemmons LE and Resurreccion AVA. 1999. Note on computerized data collection in consumer sensory evaluation. Journal of Sensory Studies 14(1):19-34.

Rousseau B and Ennis JM. 2013. Importance of correct instructions in the tetrad test. Journal of Sensory Studies 28(4): 264-269.

Qualtrics. 2015. Qualtrics Research Suite. Retrieved June 13, 2014 from Qualtrics on the World Wide Web: http://www.qualtrics.com/research-suite/.

Qualtrics. 2015. Qualtrics University Support and Training. Retrieved June 13, 2014 from Qualtrics on the World Wide Web: http://www.qualtrics.com/university/researchsuite/?utm_campaign=Help%20Button&utm_medium=Product&utm_source=MySurveysSection.

Szczesniak AS. 2002. Texture is a sensory property. Food Quality and Preference 13(4): 215-225.

Shapiro JS, Bessette MJ, Baumlin KM, Ragin DF, Richardson LD. 2004. Automating research data collection. Academic Emergency Medicine 11(11):1223-28.

Swaney-Stueve M and Heymann H. 2002. A comparison between paper and computerized ballots and a study of simulated substitution between the two ballots used in descriptive analysis. Journal of Sensory Studies 17(6): 527-537.

Talmage H and Rasher SP. 1981. Validity and reliability issues in measurement instrumentation. Journal of Nutrition Education 13(3):83-84.

Wilcox AB, Gallagher KD, Boden-Albala B, Bakken SR. 2012. Research data collection methods from paper to tablet computers. Medical Care 50(7):S68-73.

APPENDIX A: 9-POINT STRUCTURED HEDONIC LINE SCALE PAPER SCORECARD

Product code #: _____ Panelist #: _____

Please print your First and Last Name: _____

Oatmeal Cookie Sensory Scorecard

You will evaluate 3 oatmeal cookie samples today. Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by marking the box that best reflects how much you like this product. Please drink water and take a bite of carrot to to cleanse your mouth after each sample.

Appearance

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Extremely Dislike — Extremely Like

Texture

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Extremely Dislike — Extremely Like

Flavor

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Extremely Dislike — Extremely Like

Overall Acceptability

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Extremely Dislike — Extremely Like

Thank you for your participation.

APPENDIX B: 9-POINT UNSTRUCTURED HEDONIC LINE SCALE PAPER SCORECARD

Product code #: _____ Panelist #: _____

First and Last Name: _____

Vanilla Yogurt Sensory Scorecard

You will evaluate 3 vanilla yogurt samples today. Eat the sample and evaluate appearance, texture, flavor, and overall acceptability by using a pen to mark (____) the spot on the line scale that best represents how much you like this product. Please drink water and eat some crackers to cleanse your mouth after each sample.

Appearance

_____

Dislike Extremely — Like Extremely

Texture

_____

Dislike Extremely — Like Extremely

Flavor

_____

Dislike Extremely — Like Extremely

Overall acceptability

_____

Dislike Extremely — Like Extremely

Thank you for your participation!

APPENDIX C: RANKING PAPER SCORECARD

Panelist #: _____

First and Last Name: _____

Peanut Butter Ranking Test

Drink seltzer water before starting, and between each sample. Spread the peanut butter on the crackers provided, or taste the peanut butter samples with the sample spoons provided.

You are provided with three samples of peanut butter, each labeled with a three-digit code. **Please assess the samples in the order presented, from left to right.** Rank the samples from most preferred to least preferred using the following numbers:

1 = least preferred, 3 = most preferred
(Ties are NOT allowed)

Sample	Rank (1 to 3)
Code#	_____
Code#	_____
Code#	_____

Thank you for your participation!

APPENDIX D: TRIANGLE TEST PAPER SCORECARD

Panelist #: _____

First and Last Name: _____

Cheetos Triangle Test Scorecard

Take a sip of seltzer water before starting, and between each sample.

You will receive 1 set of three coded samples. Two of the samples are the same and one is different. Please taste the samples in the order presented, from left to right. **Circle the number of the sample that is different.**

975 **201** **558**

Please wait for additional questionnaires.

APPENDIX E: TETRAD TEST PAPER SCORECARD

Panelist #: _____

First and Last Name: _____

Orange Juice Tetrad Test Scorecard

Take a sip of water and a bite of cracker before starting and after each sample.

Please taste the four orange juice samples in the order shown below, from left to right. Group the samples into two groups of two based on similarity. Write the sample codes for the two groups in the spaces below.

890 787 260 597

Group A _____ _____

Group B _____ _____

Please wait for additional questionnaires.

APPENDIX F: 15-POINT UNSTRUCTURED DESCRIPTIVE LINE SCALE PAPER SCORECARD

Almond Scorecard

Panelist #: _____ Date: _____

Directions:
1. Place one almond at a time between the molars; bite through once, evaluate for hardness. Continue to chew sample and evaluate for sweetness.

2. Mark an X on scale along with product number.

Hardness – use reference hardness card

0_____15
Not perceptible High Intensity

Sweetness - no reference needed

0_____15
Not perceptible High Intensity

Directions:
1. Place one almond at a time between the molars; bite through once; evaluate for crunchiness. Continue to chew sample and evaluate for overall quality.

2. Mark an X on scale along with product number.

Crunchiness – use crunchiness reference card

0_____15
Not perceptible High Intensity

Directions:
1. Quickly open lid and take 3 short sniffs of the sample, through the nose; close lid and evaluate for presence of a cardboardy, rancidity, and/or painty note. Then rate its intensity. Between each sample wait 30 seconds and continue.

2. Mark an X on scale.

Cardboardy – use Universal Intensity scale

0_____15
Not perceptible High Intensity

Rancid – use Universal Intensity scale

0_____15
Not perceptible High Intensity

Painty – Use Universal Intensity scale

0_____15
Not perceptible High Intensity

Directions:
1. Taste one almond and evaluate for presence of a cardboardy, rancidity, and/or painty note. Then rate its intensity. Wait 30 seconds between almonds.

2. Mark an X on the scale.

Cardboardy – use Universal Intensity scale

0_____15
Not perceptible High Intensity

Rancid – use Universal Intensity scale

0_____15
Not perceptible High Intensity

Painty – use Universal Intensity scale

0_____15
Not perceptible High Intensity

APPENDIX G: USABILITY PAPER QUESTIONNAIRE

Product code #: _____ Panelist #: _____

Please print your First and Last Name: _____

Usability Questionnaire

Please answer the following questions regarding your experience using the paper scorecard to assess today's food product.

1. The instructions on this scorecard are easy to understand and follow.

_____ Strongly agree
_____ Agree
_____ Neutral
_____ Disagree
_____ Strongly disagree

2. How comfortable did you feel when completing this scorecard?

_____ Very comfortable
_____ Comfortable
_____ Neutral
_____ Uncomfortable
_____ Very uncomfortable

APPENDIX H: DEMOGRAPHIC PAPER QUESTIONNAIRE

Product code #: _____ Panelist #: _____

Please print your First and Last Name: _____

Please answer the following information about yourself. Responses will allow analysis of the participants as a whole. All information is and will remain confidential.

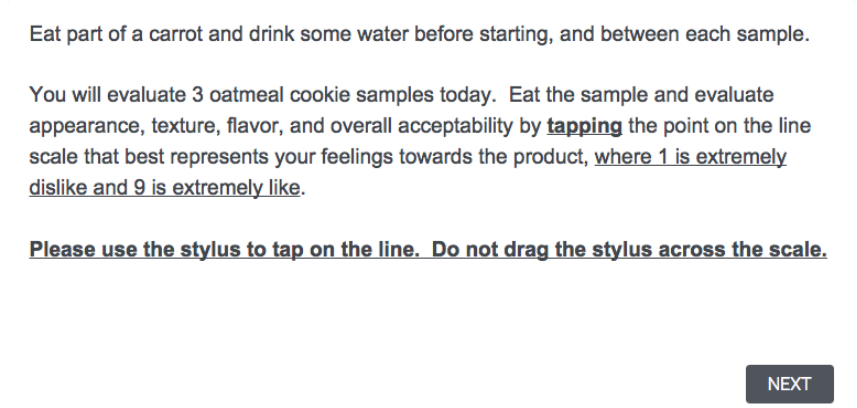
1. Age: ___ 18-20 ___ 21-23 ___ 24-26 ___ 27-29 ___ 30-39 ___40-49 ___ 50-Over

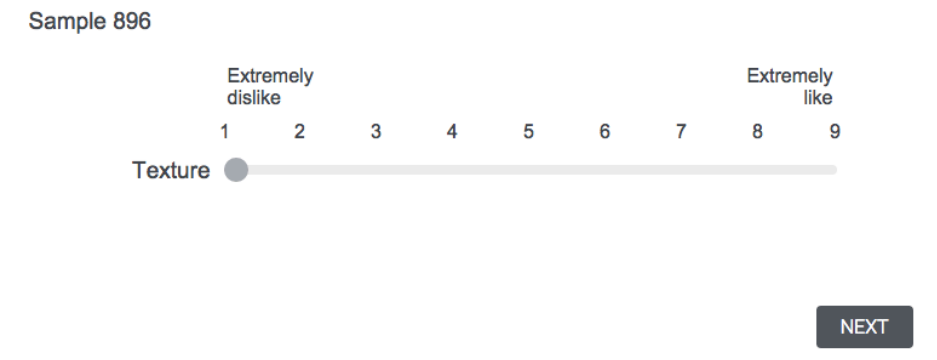
2. Sex: ___ Male ___ Female

3. How often do you consume XXX? (Check 1)

_____ Daily
_____ Several times a week
_____ Several times a month
_____ Once a month
_____ Several times a year
_____ Never

APPENDIX I: 9-POINT STRUCTURED HEDONIC LINE SCALE QUALTRICS SCORECARD

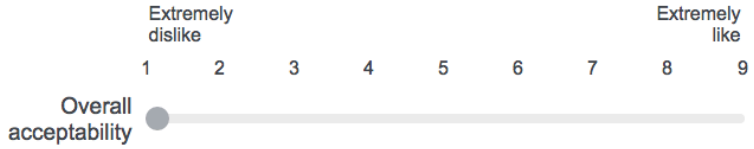




Sample 896

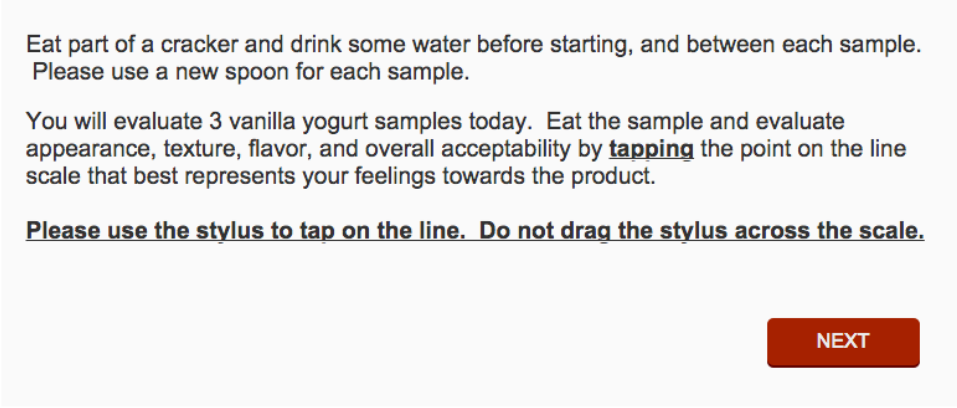
NEXT

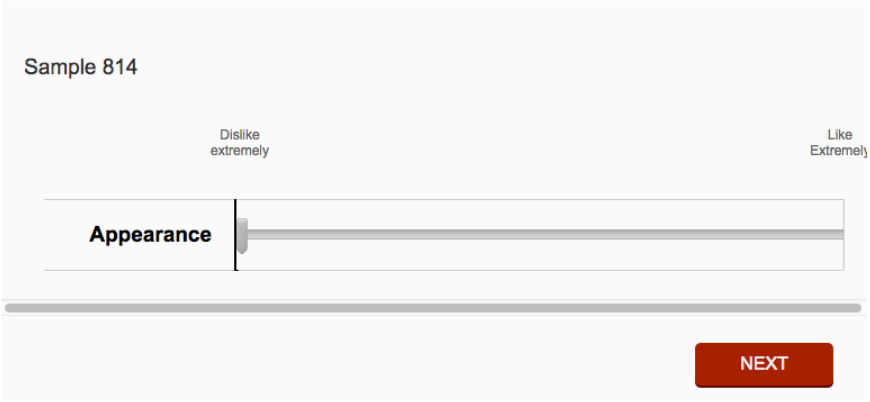
Sample 896

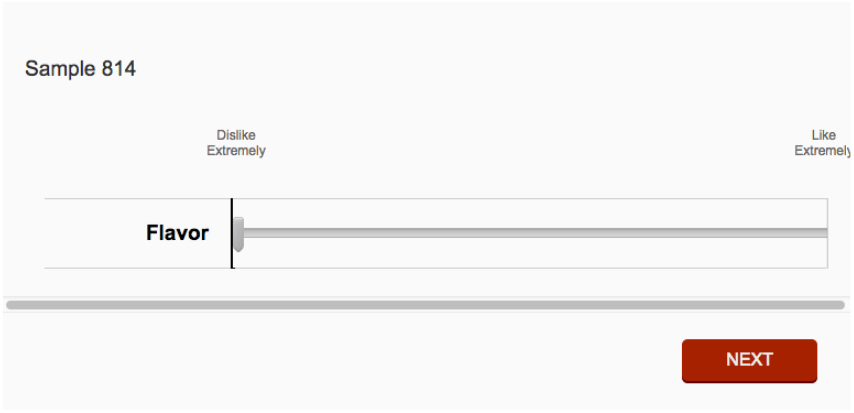


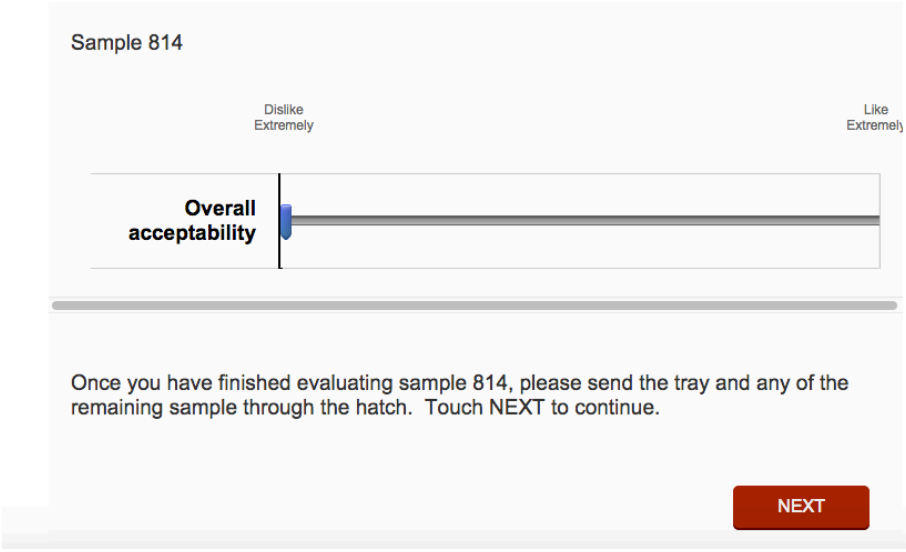
NEXT

APPENDIX J: 9-POINT UNSTRUCTURED HEDONIC LINE SCALE QUALTRICS SCORECARD






Sample 814
Dislike Extremely
Like Extremely
Flavor
NEXT


Sample 814
Dislike Extremely
Like Extremely
Overall acceptability
Once you have finished evaluating sample 814, please send the tray and any of the remaining sample through the hatch. Touch NEXT to continue.
NEXT

APPENDIX K: RANKING TEST QUALTRICS SCORECARD

Drink seltzer water before starting, and between each sample. Spread the peanut butter on the crackers provided, or taste the peanut butter samples with the sample spoons provided.

You are provided with three samples of peanut butter, each labeled with a three-digit code. **Please assess the samples in the order presented, from left to right.** Rank the samples from most preferred to least preferred using the following numbers:

1 = least preferred, 3 = most preferred

(Ties are NOT allowed)

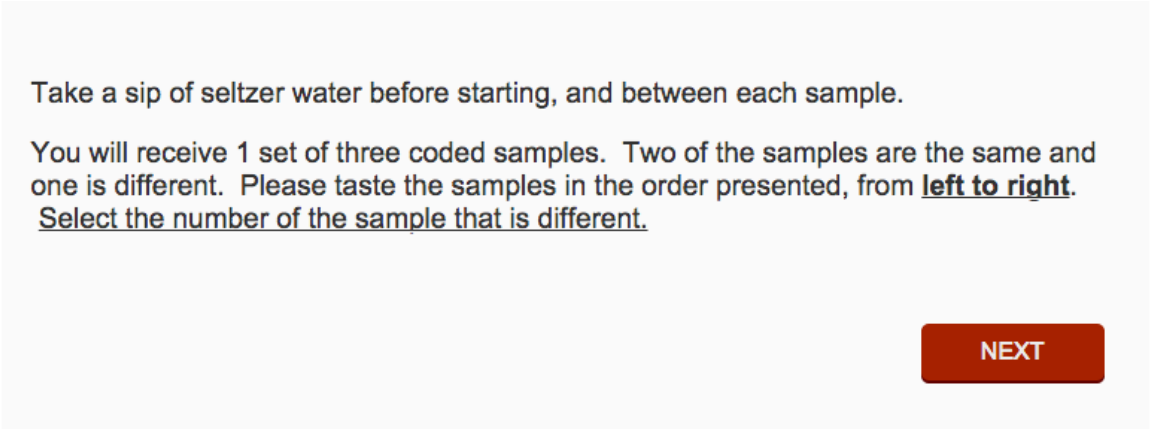
NEXT

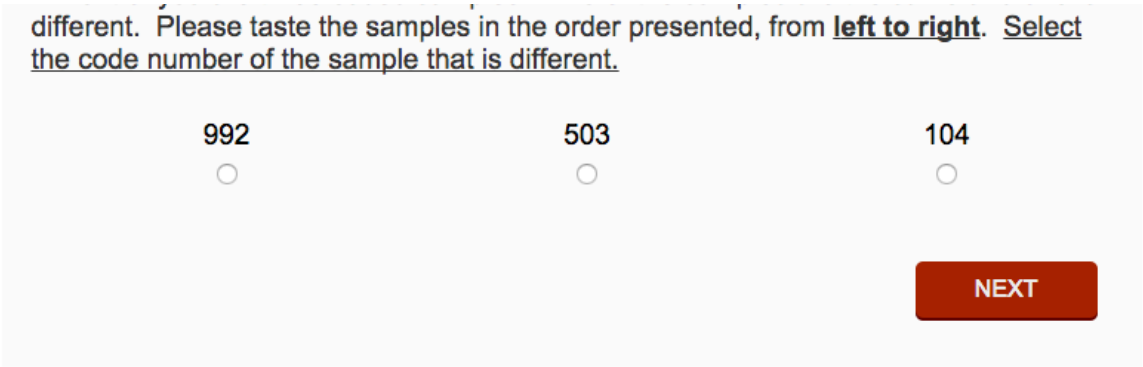
Please assess the samples in the order presented, from left to right.

Rank the samples on a scale of 1 (least preferred) to 3 (most preferred) by typing the number into the box beside the corresponding sample code.

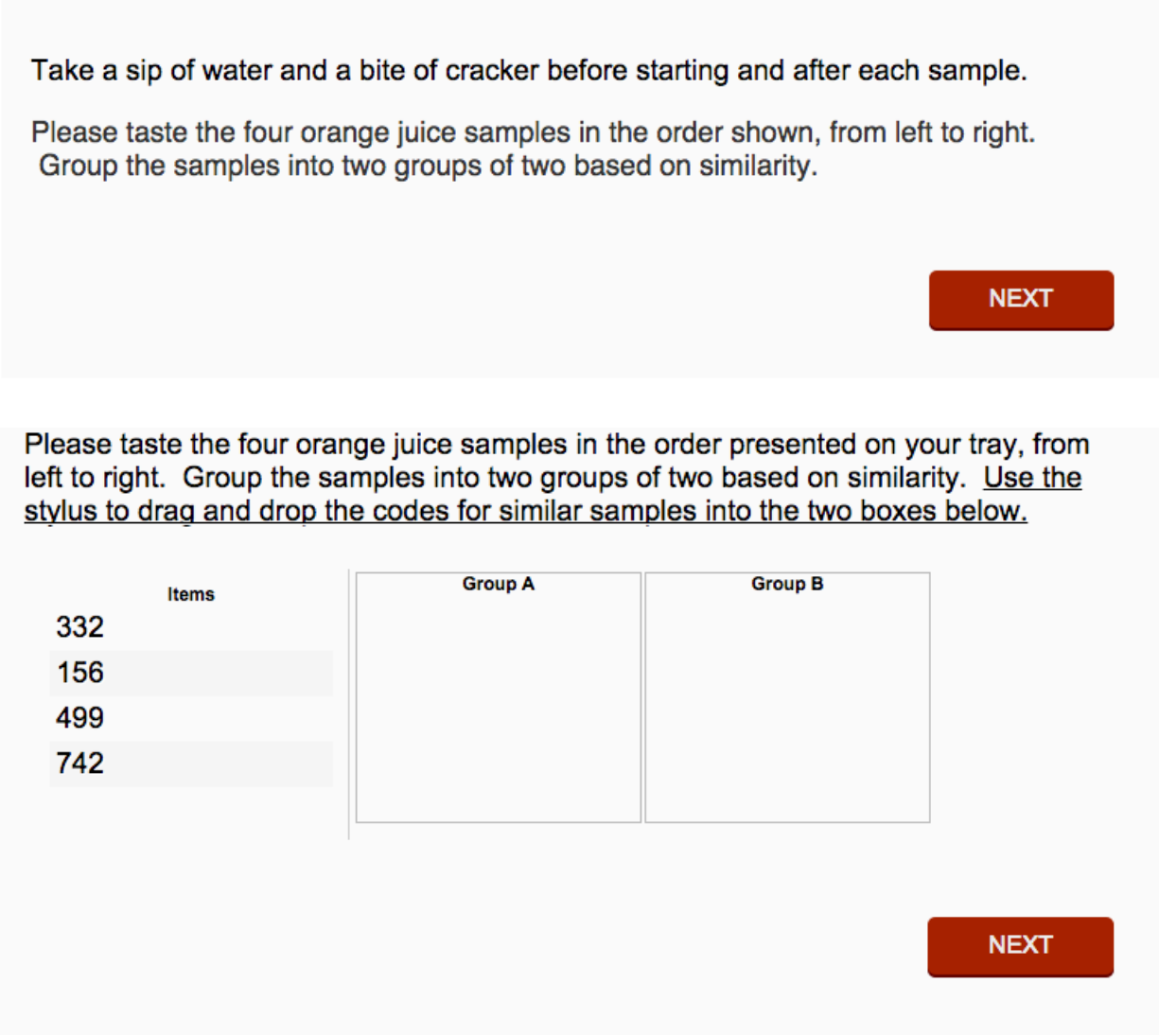
- [ ] 351
- [ ] 602
- [ ] 440

APPENDIX L: TRIANGLE TEST QUALTRICS SCORECARD





APPENDIX M: TETRAD TEST QUALTRICS SCORECARD



APPENDIX N: 15-POINT UNSTRUCTURED DESCRIPTIVE LINE SCALE QUALTRICS SCORECARD

Place one almond at a time between the molars; bite through once, evaluate for hardness. Continue to chew sample and evaluate the sweetness.

Use the hardness reference card. Drag the slider to indicate the intensity of the hardness in the sample. The full length of the line is shown.

No reference card needed for sweetness. Drag the slider to indicate the intensity of the sweetness in the sample. The full length of the line is shown.

Place one almond at a time between the molars; bite through once; evaluate for crunchiness. Continue to chew sample and evaluate for overall quality.

Use crunchiness reference card. Drag the slider to indicate the intensity of the crunchiness in the sample. The full length of the line is shown.

Quickly open lid and take 3 short sniffs of the sample, through the nose; close lid and evaluate for presence of a cardboardy, rancidity, and/or painty note. Then rate its intensity. Between each sample wait 30 seconds and continue.

Cardboardy - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

Rancid - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

Painty - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

Taste one almond and evaluate for presence of a cardboard, rancidity, and/or painty note. Then rate its intensity. Wait 30 seconds between almonds.

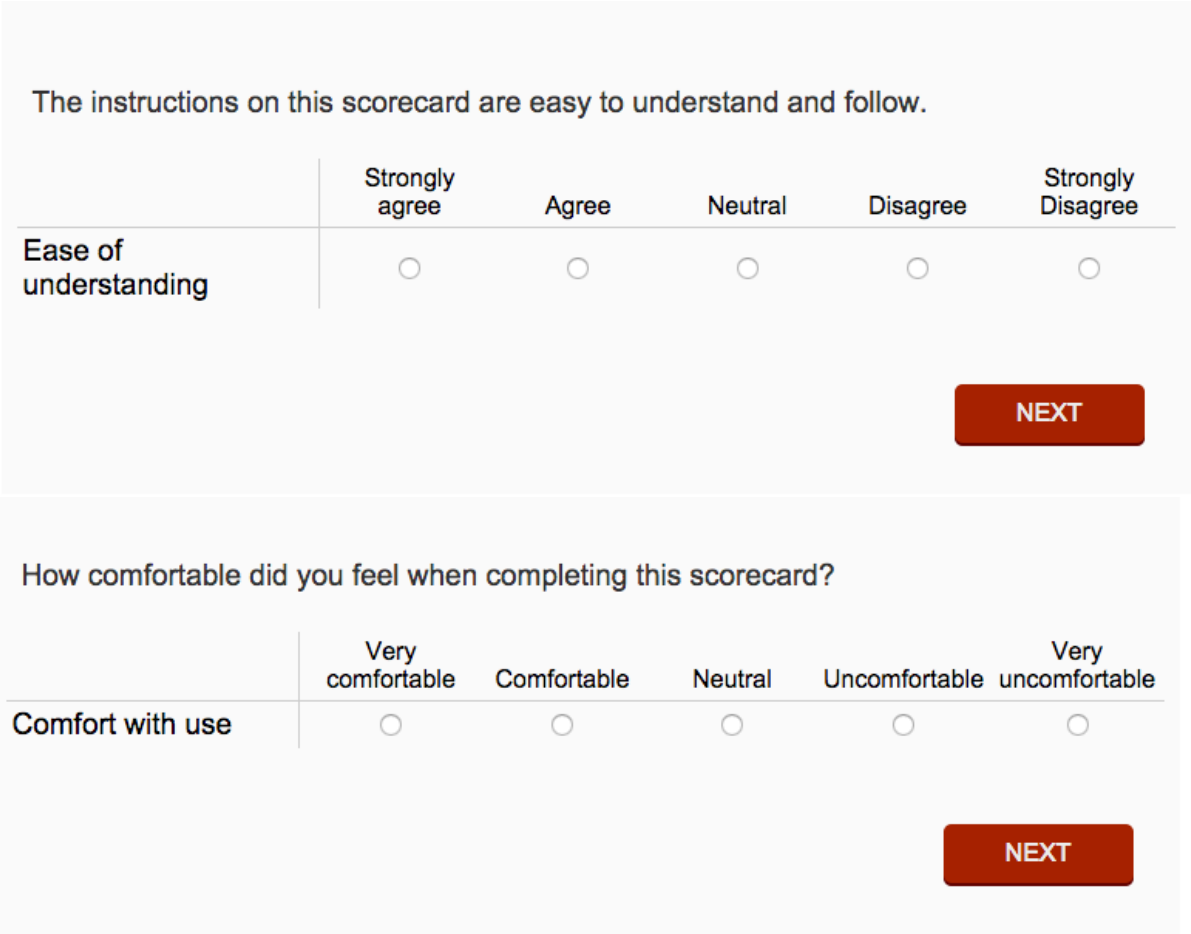
Cardboardy - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

Rancid - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

Painty - use Universal Intensity scale. Drag the slider to indicate the intensity of the characteristic in the sample. The full length of the line is shown.

>>

APPENDIX O: USABILITY QUALTRICS QUESTIONNAIRE



APPENDIX P: DEMOGRAPHIC QUALTRICS QUESTIONNAIRE

Please choose your age category.

- ○ 18-20
- ○ 21-23
- ○ 24-26
- ○ 27-29
- ○ 30-39
- ○ 40-49
- ○ 50 and over

NEXT

Please choose your sex.

- ○ Male
- ○ Female

NEXT

How often do you eat cheetos? (check one)

- ○ Daily
- ○ Several times a week
- ○ Several times a month
- ○ Once a month
- ○ Several times a year
- ○ Never

NEXT

APPENDIX Q: CONSUMER SENSORY PANEL CONSENT FORM

UNIVERSITY OF GEORGIA
CONSENT FORM
METHODOLOGY FOR FOOD SENSORY ANALYSIS

Researcher's Statement
I am/We are asking you to take part in a research study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. This form is designed to give you the information about the study so you can decide whether to be in the study or not. Please take the time to read the following information carefully. Please ask the researcher if there is anything that is not clear or if you need more information. When all your questions have been answered, you can decide if you want to be in the study or not. This process is called "informed consent." A copy of this form will be given to you.

Principal Investigator: Dr. Ruthann Swanson
Dept. of Foods and Nutrition
706-542-4834

Purpose of the Study
Traditional sensory data collection methods require manual data collection and entry, which is time-consuming and prone to human error. The purpose of this study is to learn whether alternative data collection methods can provide reliable information while decreasing the need for manual data handling. You are being asked to participate because information collected from you will help us better understand the advantages and disadvantages of various sensory data collection methods.

Study Procedures
If you agree to participate, you will be asked to …

- Evaluate commercially prepared food products on two separate occasions (5-10 minutes each) and complete a short demographic and food choices questionnaire and assess the sensory evaluation process (5 minutes each).

Risks and discomforts

- There are no risks associated with participation for any person who does not have food allergies to any of the ingredients in the commercially prepared food products being used in this study.

- Food allergies that you have include:

_____
_____
_____

Benefits

- Participation in this research may enhance discussions in the classroom, facilitating a better understanding of the process and limitations associated with development/success of products available to consumers.
- This research is also expected to update research regarding the use of data collection methods in food sensory analysis. These results may facilitate the use of more efficient data collection methods in the food industry and in academic sensory laboratories that train sensory professionals.

Incentives for participation

Students who have selected participation on this sensory panel as an extra credit option will receive class credit. In classes where extra credit is offered, other options are available, however these options vary with class and are determined by the instructor in the class. Partial credit will be given for participation in the first session even if the student elects not to participate in the follow-up session. Optional commercially prepared snacks will be made available following the participant's involvement in the research study.

Privacy/Confidentiality

Due to the need to accurately assess responses and the data analysis technique that will be used, participants in this study will need to be identifiable until all data are collected and the data have been summarized. Names will not be released with results nor made public in any way. All data will be accessible only by researchers involved in this study. Storage of data will be in a locked file cabinet only accessible by researchers. Electronic data will be password protected and only accessible by the researchers. Researchers will not release identifiable results of the study to anyone other than individuals working on the project without your written consent unless required by law.

Taking part is voluntary

Your involvement in the study is voluntary, and you may choose not to participate or to stop at any time without penalty or loss of benefits to which you are otherwise entitled. Your decision about participation will have no bearing on your grades earned via routine class assignments/quizzes/activities or your class standing.

If you have questions

The main researchers conducting this study are Dr. Ruthann Swanson and Sara Sheridan at the University of Georgia. Please ask any questions you have now. If you have questions later, you may contact Dr. Ruthann Swanson at rswanson@fcs.uga.edu or at 706-542-4834. If you have any questions or concerns regarding your rights as a research participant in this study, you may contact the Institutional Review Board (IRB) Chairperson at 706.542.3199 or irb@uga.edu.

Research Subject's Consent to Participate in Research:

To voluntarily agree to take part in this study, you must sign on the line below. Your signature below indicates that you have read or had read to you this entire consent form, and have had all of your questions answered.

_____	_____	_____
Name of Researcher	Signature	Date
_____	_____	_____
Name of Researcher	Signature	Date
_____	_____	_____
Name of Participant	Signature	Date

Please sign both copies, keep one and return one to the researcher.

APPENDIX R: DESCRIPTIVE SENSORY PANEL CONSENT FORM

UNIVERSITY OF GEORGIA
CONSENT FORM
DESCRIPTIVE METHODOLOGY FOR FOOD SENSORY ANALYSIS

Researcher's Statement
I am/We are asking you to take part in a research study. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. This form is designed to give you the information about the study so you can decide whether to be in the study or not. Please take the time to read the following information carefully. Please ask the researcher if there is anything that is not clear or if you need more information. When all your questions have been answered, you can decide if you want to be in the study or not. This process is called "informed consent." A copy of this form will be given to you.

Principal Investigator: Dr. Ruthann Swanson
Dept. of Foods and Nutrition
706-542-4834

Purpose of the Study
Traditional sensory data collection methods require manual data collection and entry, which is time-consuming and prone to human error. The purpose of this study is to learn whether alternative data collection methods can provide reliable information while decreasing the need for manual data handling. As a trained panelist who has participated in the Evaluation of Almonds study, we would like your input regarding the feasibility of using electronic data collection methods in future studies in which products are assessed by trained panelists.

Study Procedures
If you agree to participate, you will be asked to …

- Evaluate roasted almonds using an electronic version of the sensory scorecard previously employed, complete a short demographic and food choices questionnaire and assess the sensory evaluation process (5 minutes).

Risks and discomforts

- There are no risks for those who do not have an allergy to almonds.

Benefits

- Participation in this research may enhance discussions in the classroom, facilitating a better understanding of the process and limitations associated with development/success of products available to consumers.
- This research is also expected to update research regarding the use of data collection methods in food sensory analysis. These results may facilitate the use of more efficient data collection methods in the food industry and in academic sensory laboratories that train sensory professionals.

Incentives for participation
You will be paid $25 for participation in this additional sensory session.

Privacy/Confidentiality
Due to the need to accurately assess responses and the data analysis technique that will be used, participants in this study will need to be identifiable until all data are collected and the data have been summarized. Names will not be released with results nor made public in any way. All data will be accessible only by researchers involved in this study. Storage of data will be in a locked file cabinet only accessible by researchers. Electronic data will be password protected and only accessible by the researchers. Researchers will not release identifiable results of the study to anyone other than individuals working on the project without your written consent unless required by law.

Taking part is voluntary
Your involvement in the study is voluntary, and you may choose not to participate or to stop at any time without penalty or loss of benefits to which you are otherwise entitled. Your decision about participation will have no bearing on your continued participation in the Evaluation of Almond Study.

If you have questions
The main researchers conducting this study are Dr. Ruthann Swanson and Sara Sheridan at the University of Georgia. Please ask any questions you have now. If you have questions later, you may contact Dr. Ruthann Swanson at rswanson@fcs.uga.edu or at 706-542-4834. If you have any questions or concerns regarding your rights as a research participant in this study, you may contact the Institutional Review Board (IRB) Chairperson at 706.542.3199 or irb@uga.edu.

Research Subject's Consent to Participate in Research:
To voluntarily agree to take part in this study, you must sign on the line below. Your signature below indicates that you have read or had read to you this entire consent form, and have had all of your questions answered.

_____	_____	_____
Name of Researcher	Signature	Date
_____	_____	_____
Name of Researcher	Signature	Date
_____	_____	_____
Name of Participant	Signature	Date

Please sign both copies, keep one and return one to the researcher.