2013

Psychological biases affect hedonic ratings

Wannita Jirangrat
Louisiana State University and Agricultural and Mechanical College, wjirangrat@gmail.com

PSYCHOLOGICAL BIASES AFFECT HEDONIC RATINGS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Food Science

by
Wannita Jirangrat
B.S., Kasetsart University, 2004
M.S., Kasetsart University, 2007
May 2013

DEDICATION

I dedicate this dissertation to my parents. Without their knowledge, wisdom, and guidance, I would not have had the goals to strive and be the best to achieve my dream. I love you Mom and Dad, for helping to make me who I am, for teaching me to be proud of who I am, for showing me how to be strong, for giving me the courage to accept my weakness, and for giving me the strength to always strive for better, no matter what. You are my rock and foundation!

ACKNOWLEDGEMENTS

My graduate school experience has proven that academia can be a home for collaboration, collegiality and enduring friendships. My works would not have been possible without an endless support of many mentors, friends and family. First and foremost, I would like to convey my immense gratitude towards my major professor and advisors. An endeavor of this magnitude would not have come true without the diligent, talent, persistent and enlightening criticism of my professor, Dr. Witoon Prinyawiwatkul. I deeply appreciate his guidance, consideration, patience, support and encouragement devoted throughout these four years and six months. He has made my idea and my voice visibly. I am no doubt that his valued supervision has paved my way to be a great researcher in the near future. I have taken the great pleasure of being his student.

I also would like to express my gratitude to all of my committee. I would like to thank Dr. Jing Wang, my statistics committee for an incredible help. My academic accomplishments could not have been achieved without her. I would like to thank Dr. Joan King, Dr. Zhimin Xu and Dr. Vincent Licata for supporting me throughout my research. It has been an honor and privilege for me to have them on my committee, which I could not have asked for a better one.

I would like to thank Dr. Charles Boeneke and everyone in the Dairy Science department for all their help and support. I would like to thank Mrs. Terri Gilmer for her support, encouragement and constant assistance. During four years and a half at LSU, I have gained numerous friends: Karen, Sriwang, Jinjuta, Damir, Anuwat, Wisdom, Kairy, Pamarin, Arranee, Thanida. They have made me laugh, listen and share, and been there for me. Karen and Jinjuta have offered endless time, aid, guidance, support and great company throughout my research and I am thankful to have had an opportunity to meet them. I am really grateful to Ms. Kanya and

Kanitta for all their help and support throughout my time while at LSU. My life would be uneasy without them. Thank you so much for making my graduate experience at LSU a memorable one. I would like to extend my thanks to Thitidej, Kittikoon, Thanaphon, Shettinan, Wichaya and Peeraya for being my closed friends. They have never gave up on me and always provided without me having to ask for it.

Last, but certainly not least, I would like to thank my loving and inspirational parents, my sister, and my brother. Their cherished advice and encouragement helped me strive to be my best. They have sacrificed much for me to pursue my passions. Words cannot express how grateful I am for their unconditional love and support. I am eternally grateful to have them with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Psychological biases in consumer testing may lead to misinterpretation of results and lower experimental power. Reports on various hedonic scales associated with psychological biases induced by sample presentation are limited in the literature. An appropriate experimental protocol could enable sensory scientists to accurately determine if a product is more or less liked.

Overall, in this study some drawbacks of hedonic scales were revealed and some recommendations were made under specific circumstances. A more powerful design (SPRCBD) helped minimize positional and First Serving Order (FSO) biases in consumer tests by extracting more explained variances, resulting in decreased Type-II error in the model. Logistic regression analysis was proven to be an alternative methodology to quantify sensory contrast effects. For sensory testing, a multidimensional attribute tended to be more affected by the contrast effects than a simpler attribute. Several scales have been used for assessing the degree of food liking/disliking. This study provided a good practice protocol, suggesting use of a regular scale length (100 mm.) for assessing a degree of food liking/disliking while Labeled Affective Magnitude (LAM) would be an alternative choice where the scale length effects may be a critical issue. Depending on the type of scale and its polarity, a negative attribute (e.g., bitterness) was more affected than was a positive attribute. When testing extremely liked product, one should be aware of contrast biases that affected more toward positive attributes than negative attributes.

This study demonstrated some psychological biases that affected the hedonic ratings. There are many more factors that could sway sensory responses and prevent experimenters from getting accurate, valid and actionable outcome. Understanding of psychological biases, proper product selection, and proper data analysis should be further studied to minimize misinterpretation of hedonic ratings.

CHAPTER 1 INTRODUCTION

1.1 Significance of research

A reasonable process within an individual's mind that leads a person to perform and select some products refers to a consumer's decision of purchasing (Booth, 1995; Meilgaard *et al.*, 2006; Moskowitz, 2003). In order to gain a chance of success in new product development, sensory scientists are typically part of the R&D team. Meilgaard *et al.* (2006) mentioned "Being consumer-centric is a path to success and resources need to be managed well in a highly value-driven marketplace." A new product development process must proceed carefully with the aim to satisfy people's desires. The information gained from sensory testing can help maximize consumers' satisfaction, which in turn will help minimize the risk of products' failure. In general, approximately within one year after introduction of new products, about 80-90% of these products fail to survive in the market (Morris, 1993). Several factors including marketability, profitability and feasibility (Barabba and Zaltman, 1991; Bradley and Nolan, 1998; Clancy and Krieg, 2000; Pine and Gilmore, 1999; Stanton, 1997; Zaltman, 2003) contribute to a products' failure. The earlier we know and the more we know about the product, the lower the products' failure rate.

Sensory sciences provide a tool to gain product insight. This will help us not only to get to know the product but also to speed up a problem solving if necessary. The top three advantages of using sensory evaluation in product development are to properly design the studies, to properly collect data from both experts and consumers and to properly interpret the results (Moskowitz, 2000). The ideas from marketing, RD teams and consumers are unique and useful, and can be combined to increase productivity of product design and development (Eng and Quaia, 2009; Lu and Yang, 2004). Whether the product development will succeed partially depends upon how well sensory scientists communicate with the R&D team through a consumer language. An appropriate protocol, including proper experimental designs, practical and valid preference tests, and appropriate data analysis, could enable sensory scientists to reliably determine whether a product is more or less liked.

Sensory evaluation has been extensively reviewed by many books, such as Amerine *et al.* (1965); Lawless and Heymann (1999); Meilgaard *et al.* (2006); and Stone and Sidel (2004). Consumer acceptance testing is known as a method to quantify degrees of liking/ disliking of products. This method occupies a unique feature compared to many sensory techniques in term of general applicability (Cordonnier and Delwiche, 2008; Peryam and Pilgrim, 1957), simplicity (Villanueva *et al.*, 2005), and the use of untrained panelists (Daroub *et al.*, 2010). It can be used to determine food choice (Yeomans *et al.*, 2008) and critical information of individuals' likes and dislikes (Jaeger and Cardello, 2009). However, misuse of sensory techniques can induce negative/biased results. To increase a power of an experiment, an experimentor should have a better understanding of sensory foundation of physiological and psychological biases. Psychological effects, induced by sample presentation including positional, halo, central tendency, contrast and convergence effects may influence sensory scores. Failure to detect and take them into consideration may lead to serious misinterpretation and wrong conclusions.

Thus, this dissertation was conducted to provide insight knowledge of physico psychological biases in consumer testing as well as to propose alternative choices of experimental design, statistical analysis and disliking/liking scales for sensory scientists. This may help to decrease a risk of products' failure.

Currently, information related to physico-psychological biases in consumer testing induced by sample presentation is limited in the literature. This dissertation will be the very first research devoted to such an area. The dissertation was divided into 7 chapters. Chapter one provides an introduction and justification of this dissertation research. Chapter two provides a review of relevant literature. Chapters three to six provide results of a series of experiments pertinent to physico-psychological biases including position, contrast, scale types, scale lengths, scale polarities, attributes and product impression effects. Several methods have been conducted to minimize an extraneous error. Theoretically, a proper experimental design (Macfie *et al*., 1989; Williams, 1948), a proper product selection (Lee and Meullenet, 2010; Villanueva, *et al.*, 2005) and a proper data analysis (Hottenstein *et al.,* 2008) could help minimize an irrelevant error. Some of these factors were discussed in this dissertation. Chapter seven provides a summary and significance of this dissertation. All cited references are given at the end of each chapter. The appendices include all supplementary information associated with these studies.

1.2 References

Almeida, S.B., Aparecida, M. and Da Silva, A.P. 2002. Hedonic scale with reference: performance in obtaining predictive models. ***Food Quality and Preference*** 13(1): 57-64.

Amerine, M.A., Pangbnor, R.N. and Roseller, E.B. 1965. ***Principles of Sensory Evaluation of Food.*** New York and London: Academic Press.

Barabba, V. and Zaltman, G. 1991. ***Hearing the voice of the market: Competitive advantage through creative use of market information***. Boston: Harvard Business School Press.

Bradley, S.P. and Nolan, R.L. 1998. ***Sense and Response***. Boston: Harvard Business School Press.

Clancy, K.J. and Krieg, P.C. 2000. ***Counterintuitive Marketing: Achieving Great Results Using Common Sense***. New York: Free Press.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Daroub, H., Olabi, A. and Toufeili, I. 2010. Designing and testing of an Arabic version of the hedonic scale for use in acceptability test. ***Food Quality and Preference,*** 21(1): 33-43

Eng, T.Y. and Quaia, G. 2009. Strategies for improving new product adoption in uncertain environments: A selective review of the literature. ***Industrial Marketing Management*** 38(3): 275-282.

Hottenstein, A.W., Taylor, R. and Carr, T.B. 2008. Preference segment: A deeper understanding of consumer acceptance or a serving order? ***Food Quality and Preference,*** 19: 711-718.

Jaeger, S.R. and Cardello, A.V. 2009. Direct and indirect hedonic scaling methods: A comparison of the labeled affective magnitude (LAM) scale and best-worst scaling. ***Food Quality and Preference*** 20(3): 249-258.

Lawless, H.T., and Heymann, H. 1999. ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher. 827p.

Lee, Y.S. and Meullenet, J.F. 2010. Comparison of eliminating first order samples for minimizing first serving order bias to data correction. ***Food Science Biotechnol*** 19 (3): 703-709.

Lu, L.Y.Y. and Yang, C. 2004. The R&D and marketing cooperation across new product development stages: An empirical study of Taiwan's IT industry. ***Industrial Marketing Management*** 33(7): 593-605.

Macfie, H.J.H., Bratchell, N., Greenhuff, K. and Vallis, L.V. 1989. Designs to balance the effect of order of presentation and first-order carry-over effects in hall test. ***Journal Sensory Studies*** 4: 129-148.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 2006. ***Sensory Evaluation Techniques***. 4th edition. Florida: CRC Press. 448p.

Morris, C.E. 1993. Why new product fail. ***Food Engineering*** 65 (6): 132-136.

Moskowitz, H.R. 1996. Experts versus consumers: A comparison. ***Journal Sensory Studied*** 11: 19-37.

Moskowitz, H.R. 2000. ***"R&D driven product evaluation in the early stage of development."*** in Developing new food products for a changing marketplace. Lancaster, PA: Technology Publishing Co.

Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preference. ***Food Technology*** *(September)*, 9-14

Pine, B.J.II. and Gilmore, J.H. 1999. ***The experience economy***. Boston: Harvard Business School Press.

Stanton, J.L. 1997. Who's no. 1? Consumer! ***Food Processing*** December: 55-57.

Stone, H. and Sidel, J. 2004. ***Sensory Evaluation Practices***. 3rd edition. New York: Academic Press. 338p.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. ***Food Quality and Preference***, 16: 691-703.

William, E.J. 1948. Experimental designs balanced for the estimation of residual effects of treatments. ***Australian Journal of Scientific Research,*** A(2): 149-169.

Yeomans, M.R., Chambers, L., Blumenthal, H. and Blake, A. 2008. The role of expectancy in sensory and hedonic evaluation: The case of smoked salmon ice-cream. ***Food Quality and Preference*** 19(6): 565-573.

Zaltman, G. 2003. ***How customer think***. Boston: Harvard Business School Press.

CHAPTER 2 LITERATURE REVIEW

What is "Sensory Evaluation?" Several definitions have been defined since 1954. One of the definitions provided by the Sensory Evaluation Division of the Institute of Food Technologists is:

"Sensory evaluation is a scientific discipline used to evoke, measure, analyze and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch and hearing" (IFT, 1975).

Referring to this statement, sensory evaluation is the studies, associated with five human senses that are used to judge/score a product. The best possible way to increase a product's success is to get to know the target consumer in order to determine a key that drives products' satisfaction and purchase intent. The questions such as what consumers like and want need to be answered beforehand. Knowing how a consumer behaves may help predict the product's growth rate and survival rate. Sensory techniques could serve as a useful tool to quantify such estimation. It also helps product developers to transfer consumers' needs into a product description. This is not only used for improving an existing product but also helping to explore a new area of opportunity for further development. The sensory techniques have been extensively applied to various studies, but when did sensory science actually begin? And why it is so popular?

The beginning of sensory science was extended from the psychological research area. Back to 1947, the history of systematic sensory analysis in the United State began during wartime. Sensory techniques were developed in an attempt to improve food acceptability for the American military (Dove, 1947 mentioned in Pangborn, 1964) by Peryam. His colleague and he introduced sensory science to assess a consumer preference. During that time, the growth of trading has made sensory testing more popular. The assessment of food quality based on sensory perception has become more important as a reflection of food grading and prices (Meilgaard *et al.*, 2006). Pfenninger (1979) is the one who conducted the very first study using the word "organoleptic testing" that was referred to as a measurement of sensory testing. At the time, the sensory tests were too subjective and informal. The interpretation was opened to arbitrary unfairness. Sensory evaluation was not well known until 1954. The hedonic method used to assess consumer acceptance, is a well known method developed by Peryam and his colleague. It was adopted very quickly by many companies to assess a degree of products' liking/disliking. However, several questions regarding applications of sensory science and reliability have been raised. Meilgaard *et al.* (2006) said

> "Scientists have only recently developed sensory testing as a formalized, structured, and codified methodology, and they continue to develop new methods and refine existing ones. This is a hard science and much more with sense and feeling"

Regarding this statement, consumer's sense and feeling are somewhat unpredictable and changeable. The equation to predict consumers' need and satisfaction cannot perform perfectly. The sensory results can only be used as a guideline for development and improvement. We cannot create an exact equation to predict true consumer responses. It has to be a case-by-case basis study. This makes sensory science more interesting. It has been almost six decades some sensory science was initiated, yet there is still a need for study associated with fundamental sensory science. This would help to solidify further applications of sensory techniques. Much more evidences are demanded to understand consumer perception and to support various sensory theories and assumptions. Many sensory techniques have been developed to gain more consumers understanding. Three areas of sensory studies including discriminative testing, descriptive testing and consumer preference and acceptance tests will be briefly reviewed in the next topic.

2.1 Sensory evaluation techniques

Each sensory technique requires different elements to maximize its performance and to obtain valid results. Three main elements required for each sensory technique include:

(1) Type of target consumers

(2) Choices of test locations

(3) Objectives of an experiment

Several questions regarding abovementioned elements need to be answered prior to selecting sensory techniques to be applied. What types of panelists will be used: trained or naïve consumers? Where the tests take place in a laboratory, central location or home use test? What is the aim of the study: to determine quality change, to evaluate a products' shelf life, to characterize a product profile or to assess the degree of liking/disliking of new products?

Sensory techniques can be classified as:

(1) Discrimination or difference test

(2) Descriptive analysis

(3) Preference and Acceptance test

Each technique is applied for a different purpose to gain products' insight. Three techniques are described below. However, the first two areas will be brief with more details for preference - acceptance tests.

2.1.1 Discrimination/Different test

The objective of this test is to determine if products are perceived differently. Several tests based on this objective include Paired comparison, Duo-Trio, Triangle, 2-AFC, 3-AFC, same/difference test, A-Not-A and so on (Bayarri *et al.*, 2008; Bi, 2007; Duineveld *et al.*, 2003; Hautus *et al.*, 2009; Kim *et al.*, 2010; Kuesten, 2001; Lee *et al.*, 2007; Lee and Kim, 2008;

McClure and Lawless, 2010; Meyners, 2007; Sauvageot *et al.*, 2012; Wichchukit. and O'Mahony, 2010). There are several different ways to perform different tests but in general this type of test is used to answer the question, for example "Are products different in anyway? Do sensory differences exist between samples? Or "How does attribute X differ between samples?" (Meilgaard *et al*., 2006).

Some researchers apply this type of test to determine the products' similarity. However, this type of sensory technique can not provide a degree of difference or a degree of liking among products. In terms of target participants, discriminative tests require untrained panelists to participate, and in some cases, in house panelists are used. The recruitment, selection and familiarization process needs to be accomplished. Panelists should attend a "warm-up" session to know how to answer the question or how to judge the product. For test location, the test can be executed in a laboratory or central location; a home use test should not normally be performed.

2.1.2 Descriptive Analysis

This sensory technique is considered to be the most complicated method among all three areas. It is time consuming, labor intensive and costly. Several methods including flavor profile, flash profile, texture profile, free-choice profile, time-intensity, Quantitative Descriptive Analysis® (QDA) and Spectrum® (Albert *et al*., 2011; Bleibaum *et al.*, 2002; Campo *et al.*, 2010; Delarue and Sieffermann, 2004; Feria-Morales, 2002; Goto *et al.*, 2009; Lassoued *et al.*, 2008; Moon and Li-Chan, 2007; Nissen *et al.*. 2004; Wang *et al.*, 2007) are recognized as descriptive analysis. The objective of this type of study is to answer the question "Which attributes are being major influences on the product? What can be explained as the sensory characteristic of those attributes? How big is the difference between specific products?" (Meilgaard *et al*., 2006).

Major advantages are to gain insight on product information for more in-depth analysis, to create prototypes, to define sensory properties, and/or to characterize products. In terms of target panelists, this method requires a panel of 8-12 people who are subject to a training period up to 6 months or so. Panelist recruitment is a key element in conducting a successful descriptive test. Panelists should, at least, have proven capability to communicate and express their perception. They must be able to describe and differentiate product attributes and quantify intensities. The longer training period costs more. So it is wise to conduct a thorough screening process rather than to train unqualified panelists. There are a number of steps for selection and training of panelists for descriptive analysis which is not covered in this dissertation. This type of research can only be managed in the laboratory with controlled conditions.

2.1.3 Affective test (Preference - Acceptance test)

In affective testing, we talk about qualitative and quantitative tests separately. Qualitative tests include focus group interviews, in-depth interviews (IDI), focus panels, mini groups, diads and triads and acceptance tests ethnography. Quantitative tests include Paired preference test, and Multi paired preferences ranking tests. For acceptance tests, hedonic scaling and/or the Food action rating scale will be discussed. The objective of this study is to answer the question “Which products are preferred? or “How well are products liked? (Meilgaard *et al*., 2006).

The advantages are to quantify a degree of product liking/disliking with untrained panelists. There is no need for selecting and training of panelists who are product users and potential users. It works well with actual consumers. The choice of test location can be a laboratory, central location (CLT) or home use test (HUT) depending upon a budget, time frame, objective, and product types. The advantages and disadvantages are discussed below (Table 2.1). Thousands of literature articles have been published regarding hedonic scales and their application. However, because consumers are used as a tool to perform a test, several uncontrollable factors including physical and psychological biases are major concerns. Biases will affect consumers' perceptions and hedonic ratings.

Table 2.1 Advantages and disadvantages of test location choices

Test location	Advantages	Disadvantages
Laboratory tests	• Controlling for sample preparation and presentation • Shorter time to recruit in-house employees • Visual appearance can be controlled with lights • High percentages of returning responses • Cost effective (several samples can be tested at a time)	• Lack of normal consumption • Preparation and procedures may not reflect consumer experience at home • Location can influence expectation and product knowledge
Central location tests	• Moderate control for product evaluation • Lower rate of miscommunication or misunderstanding • High percentages of returning responses • Cost effective (several samples can be tested at a time)	• Artificial condition compared with experiencing at home • Limited number of questions • Limited information to be gained • Response based on the first impression
Home use tests	• Natural condition (experience at home). • Wide range of information and number of questions • Response based on a repeated purchase • Easy to apply a statistic sampling plan	• Time consuming • Expensive • Low rate of returning responses • Limited number of samples per household • Large variation due to less control in sample preparation and time, and being used in combination with other materials

Source: Meilgaard *et al.* (2006)

An in-depth discussion on psychological biases will be given in this dissertation. An example of the description, explanation and requirement for a widely used method as well as biases and possible solutions will also be discussed.

2.1.3.1 Focus Group Interviews

The focus group interview is one of the original tools of qualitative techniques. It has been widely used and offers a powerful investigation (Bovell-Benjamine *et al.*, 2009; Cardinal, *et al.* 2003; Huston and Hubson, 2008; Hyde *et al.*, 2005; Letelier, *et al.*, 2000; Meinert *et al.*, 2008; Pigott, 2002; Raz *et al.*, 2008; Rook, 2003; Walsh *et al.*, 2009). This method allows consumers to freely express their opinion toward products, concepts, and services. The benefits of using a focus group discussion are being easy to convene and cheaper than any other market research, getting new ideas from people who are not part of companies, listening and getting a voice from real consumers, and, importantly, helping to "understand human-based phenomena" (Huston and Hubson, 2008; Clancy and Krieg, 2000). This method requires a small group of consumer participants approximately 8-12 selected consumers based on specific criteria (demographic, product usage, available time, etc). The time required per session is about 1-2 hours, operated under the guidance of an experienced moderator.

2.1.3.2 In-Depth interviews

Another qualitative affective test, IDIs (In-depth interviews) is reviewed here. This technique is very similar to a focus group discussion except that it uses individual interviews (one-on-one or face-to-face). The advantages over a focus group interview are that this method can eliminate a group's biases and get information that is more personal and honest without mimicking others' opinion in group setting. This technique is very useful for sensitive issues related to illness, weight, etc. or too personal information such as sexual desires. Applications of in-depth interview can be found in the literature (Baker and Fortune (2008); Burnett *et al.* (2010); Carkhuff and Pierce (1967); Koenigsmann *et al.* (2006); Kort *et al.* (2007); Newman *et al.* (2010); Nicolson and Burr (2003); Stevens and Ahmedzai (2004); and Walter *et al.* (2004)). The disadvantages of in-depth interview are analytical time and cost required to complete the task (Meilgaard e*t al*., 2006). This method requires a larger group of consumers to participate, approximately 12-50 selected consumers. Meilgaard *et al.* (2006) mentioned that this method is unique in term of its test protocol, i.e., to have a person to use/prepare product at an interview site or at the consumer's house. However, later on this group of consumers can be brought in to discuss and compare consumer's and company's expectation. This consumer interview or consumer observation can be used to understand and gain insight information for further prototype creation or innovative development to meet consumer's need.

2.1.3.3 Hedonic

Comparing among scales and protocols, a hedonic scale is unique in terms of general applicability (Cordonnier and Delwiche, 2008; Peryam and Pilgrim, 1957), having a significant ease of use and simplicity (Villanueva *et al.*, 2005), prediction between the target product and the prototypes (Almeida *et al.*, 2002), determination of food choice (Yeomans *et al.,* 2008), and offering critical information about individuals' likes and dislikes (Jaeger and Cardello, 2009). The word "Hedonic" or "Hedonism" means pleasure or the highest good. This word was expressed at first in the 19th century and it was known in the social sense as "The greatest happiness for the greatest number" (Gosling, 1969). In the past, sensory scientists questioned a meaning of preference; although it is referred to choices regardless of the reasons for that choice, and it also implies pleasantness or degree of liking. Psychologist named this value as "Hedonic Value." The hedonic scale was created to relate a degree of liking of an emotive energy human behavior (Figure 2.1) in many psychological applications. The concept of a linear hedonic scale represents an emotion as "a scale ranging from an extremely pleasant or positive pole, to an extremely unpleasant or negative pole" (Johnton, 1999).

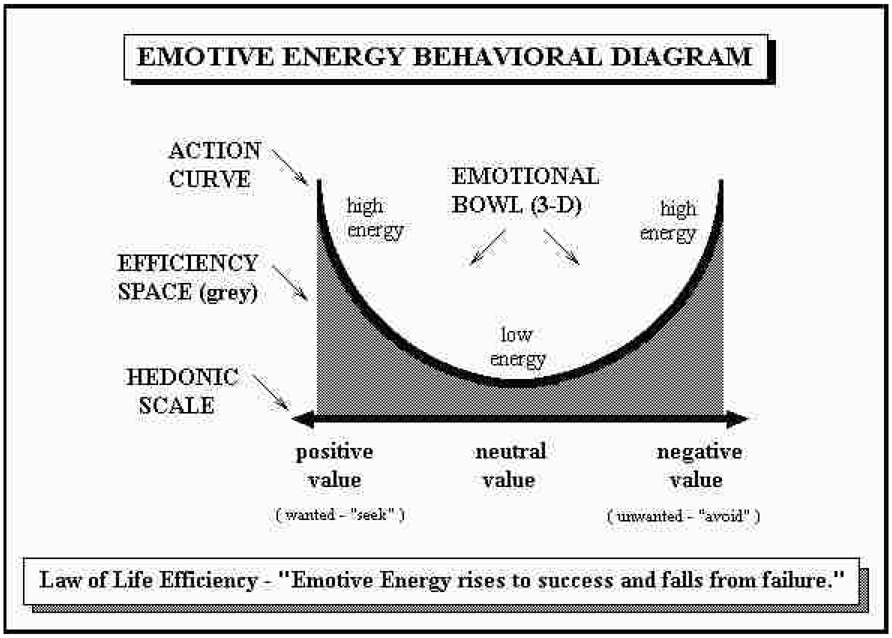

Figure 2.1: The emotive energy behavioral diagram
Source: Keeran (2004).

The hedonic methodology in the United State was first developed by David Peryam and his colleagues in a sensory laboratory at the Quartermaster Food and Container Institute of the U.S. Armed Forces (Peryam, 1954). The hedonic scale has been used to assess the degree of liking with untrained panelists who frequently use or interest in products (Cordonnier and Delwiche, 2008). In general, the hedonic scale has been used to (1) determine an overall acceptance or product's liking by a target consumer and/or product users, (2) determine a factor affecting overall acceptance or product's liking and (3) establish a relationship between consumer responses and descriptive data. The data generated from this method are spontaneous without requiring prior experience and it is appropriate for use with a wide range of populations. In term of target population, this method requires a large sample size to have a valid inference. A group of consumers approximately 50 to several hundreds selected target consumers is required.

The 9-point hedonic is most popular among other scales. The scale has nine points with given word description at both anchors ranging from dislike extremely to like extremely with a neutral category (neither like nor dislike) in order to make the scale even. All categories are described below.

1 = Dislike Extremely	2 = Dislike Very Much	3 = Dislike Moderately
4 = Dislike Slightly	5 = Neither Like nor Dislike	6 = Like Slightly
7 = Like Moderately	8 = Like Very Much	9 = Like Extremely

The label or word description along with numerical values was used to aid consumers' interpretation (Cordonnier and Delwiche, 2008). The utilization of hedonic scale has been applied to global populations for all ages. Several scales have been developed to overcome some weakness of the 9 point hedonic categorical scale; however, none of these has proven to have superior performance than the original version. Three scale types that are typically found are listed (Figure 2.2 (a-c)):

(a) A nine point categorical hedonic scale (CAT)

(b) A nine point line scale (LIN)

(c) A labeled affective magnitude scale (LAM)

In the late 1900s, some publications reported the use of a magnitude estimation (ME) scaling or the ratio scaling intended to replace and/or minimize the use of hedonic. With ME, panelists freely assign a chosen number or the number may be given from the experimenter as a reference to the first sample to describe a sensation. Panelists are then asked to assign the subsequent samples in proportion to the first sample score. In this case, if the score of the subsequent sample is two times greater than the first sample score, it implies the second sample is twice as strong as the first sample. However, it is time consuming, less effective and complicated to consumers.

In 2001, Schutz and Cardello developed a LAM scale (Figure 2.2 (c)), a modification of ME, to assess a degree of liking/disliking score. It was found to be successful as an alternative choice for quantifying a degree of liking. Still, these three scales have been debated regarding its sensitivity and application and which scale should be utilized. These three below mentioned scales were used in this dissertation research to identify the scale that best suits different objectives. The application, advantages and disadvantages of each scale will be discussed in chapters (5-6).

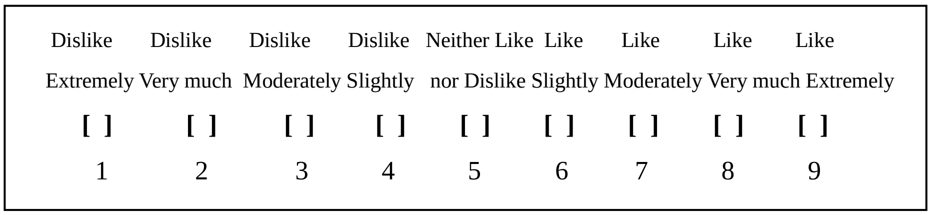

Figure 2.2 (a) A 9-point categorical hedonic scale (100 mm.)

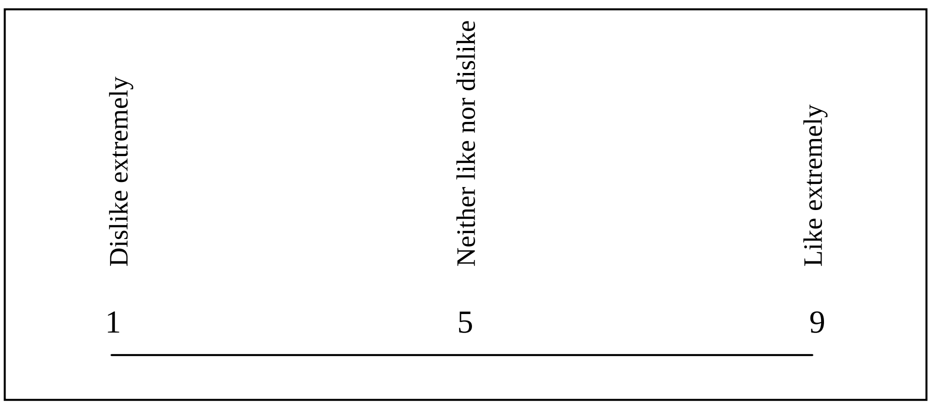

Figure 2.2 (b) A 9-point line hedonic scale (100 mm.)

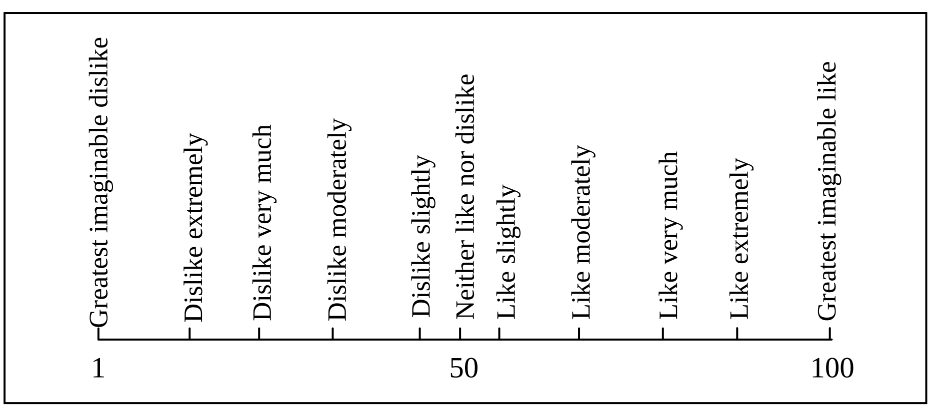

Figure 2.2 (c) Labeled Affective Magnitude scales (100 mm.)
Source: Schutz and Cardello (2001)

Back to 1950s, when a nine-point hedonic scale was developed, several questions such as how does this scale work? Is it possible to use shorter or longer scale? Are the data generated from such scale reliable? Can statistical analysis be applied?" were raised. Jones *et al.* (1955) proved that longer scales up to nine intervals tended to be more discriminating than shorter scales; however, the longer line up to eleven intervals was user unfriendly. Peryam and Pilgrim (1957) also supported that the responses from a nine interval scale were repeated more consistently within a similar consumer group. The question about an effect of different scale positions: vertical vs. horizontal was answered by Peryam and Pilgrim (1957) who reported such variations appeared minimally on the outcomes. The questions about a violation of parametric statistics assumption, ANOVA, including the lack of equivalence of the interval scale, the excessive use of neutral space (mid-scale), avoiding the use of an extreme choice at the end of both anchors and etc. are remain unclear (Dine and Olabi, 2009; McDaniel and Sawyer, 1981; Schutz and Cardello, 2001; Warnock *et al.*, 2006). Despite these concerns, parametric statistical analysis rather than non-parametric statistical analysis has been continuously used (Cardello *et al.*, 2005). Regarding hedonic scale issues can be found in more detail (Lawless and Heymann, (1999) and Meilgaard *et al.* (2006.)

2.2 Evolution of hedonic scaling method

There are thousands of literatures published related with hedonic scales. In this dissertation, an author would like to provide details of hedonic scale chronological order to facilitate a discussion (Table 2.2).

Table 2.2 Evolution of hedonic scaling method

Year	Authors	Milestone
1952	Peryam, D.R. and Girardot, N.F.	• Developed a hedonic scale to use with naïve consumer • Laboratory tests found to be more reproducible than field tests.

Table 2.2 Continued

Year	Authors	Milestone
1952	Peryam, D.R. and Girardot, N.F. (cont.)	• Standard deviation about 0.68-2.04 considered to be typical, reflecting that there are normally wide differences among people in their feeling. • Scores below 5 considered as "poor quality", over 7.5 considered "good quality."
1955	Jones, L.V., *et al.*	• As the longer the hedonic scale, the higher the power of discriminating. • The values increase when the number of intervals increases and when the midpoint is omitted.
1957	Peryam, D.R. And Pilgrim, F.J.	• Conducted an experiment to investigate scale types. Major effects of vertical or horizontal scale. • Spontaneously obtained the hedonic data can be without prior experience and can be handled by the statistics of variable. • Suitable for a wide range of consumers. • Useful for indicating general levels of acceptance. • Experimental designs can be applied to increases the sensitivity of the tests and reduce the sample size.
1957	Steven, S.S. and Galanter, E.H.	• There are prothetic (apparent length, duration, area, etc.) and metathetic (visual position, inclination, pitch, etc.) factors affecting hedonic scores. • The result showed an equal power over range (prothetic continua) between both scales (ME and category scale).
1971	Moskowitz, H.R. and Sidel, J.L.	• Compared a magnitude estimation (ME) ratio scale with a hedonic 9-point category scale of food acceptance. • Result indicated an equal sensitivity for food differentiation between these two scales. The ME scale helped to quantify the ratio of food acceptability while the hedonic provided a numerical and verbal interpretation.
1996	Moskowitz, H.R., *et al.*	• Conducted a study to quantify the odor intensity and pleasantness using ME and hedonic scales. • The result showed higher variation in hedonic judgments than in intensity judgments.
1981	McDaniel, M.R. and Sawyer, F.M.	• Conducted a preference testing of whiskey sour formulation to compare ME and 9-pt category scale between laboratory and home panel environment. • The ME resulted in more statistically significant difference for both group of panel than a hedonic scale. • Home panel resulted in more significant results than lab.

Table 2.2 Continued

Year	Authors	Milestone
1983	Giovanni, M.E. and Pangborn, R.M.	• Conducted a test to measure taste intensity of beverages and degree of liking using Graphic scaling (GS) and ME. • GS was a structured 10-cm horizontal line anchored with “Dislike extremely” and “Like extremely.” • GS was simpler and less affected by numerical and contextual effects. The data were reproducibly except for, a lemonade testing.
1990	Kroll, B.J.	• The nine (child friendly) verbal scale with “Super good” to “Super bad” performed better than either traditional 9-pt or smiley facial scale with children 5-10 years.
1996	Green, B.G., *et al.*	• Developed a Labeled Magnitude Scale (LMS), a vertical, semantic scale with quasi-logarithmic verbal labels. • LMS is as easy to use as a 9-point hedonic scale and has a greater significant ease of use than Magnitude estimation (ME). • It can be used with broadly defined sensation taste.
1998	Yeh, L.L., *et al.*	• The 9-point hedonic scale was first translated to determine a cross-cultural (Americans, Korean, Chinese and Thais) effect on 9-point hedonic scale usage. • These ethnic groups use a smaller range of scale than Americans regardless of residency or length of stay.
2000	Preston, C.C. and Colman, A.M.	• Conducted an experiment to determine an optimum number of hedonic categories to be used. • 2, 3, and 4 point scales performed poorly but the hedonic score was found significantly higher for scales with up to 7 intervals. • The test-retested reliability is likely, to decrease with more than 10-point category. • Recommended to use 7, 9 or 10 point interval.
2001	Cox, D.N., *et al.*	• Both a labeled 9-pt category scale and an unstructured-anchored line scale found no systematic cultural bias (Malaysians and Australians). However, an unstructured line scale encouraged greater use of a range of possible responses.
2001	Schutz, H. G. and Cardello, A. V.	• The Labeled Affective Magnitude (LAM) was proposed as a specialized type of modified LMS. • LAM is a line scale anchored at its end points with the phrases “greatest imaginable like” and “greatest imaginable dislike.”

Table 2.2 Continued

Year	Authors	Milestone
2001	Schutz, H. G. and Cardello, A. V. (cont.)	• It was found to have equal reliability, but greater sensitivity than a 9-point hedonic scale and more user friendly than ME.
2001	Curia, A.V., *et al.*	• A 9- point hedonic scale was translated into Spanish to compare with an English version for testing with Argentina population. • It was found that approximately 30% of the subjects rated the translated phrases differently in relation to the English version • Translated version needs to be used with caution.
2002	Bergara-Almeida, S. and Da Silva, A.A.P.	• The study was conducted to determine a performance of a hedonic scale with a reference to generate predictive models. • The models generated by the two scales were similar with respect to the adjusted R^2.
2004	Jeon, S.Y. *et al.*	• Conducted an experiment to compare between a category and line scale under various experimental protocol • It was found that neither scale has advantages over each others. Category and line can be used interchangeably.
2004	Cadello, A.V. and Schutz, H.G.	• Offered a precise numerical value corresponding to a verbal term in the scale intended for investigators' utilization with either paper or computer-based ballots.
2005	Cardello, A.V., *et al.*	• Developed Satiety Labeled Intensity Magnitude (SLIM) scale in compared with Visual Analogous Scales (VAS) to indicate perceived hungry/fullness. • SLIM has higher sensitivity, reliability and ease of use compared with VAS.
2005	Villanueva, N.D.M. *et al.*	• The results evidenced the superiority of the hybrid hedonic scale as compared to the structured and self-adjusting scales. Both with respect to the discriminating power and the ANOVA assumptions. • Both the structured and hybrid hedonic scales had greater significant ease of use than the self-adjusting scale.
2006	Greene, J.L., *et al.*	• Conducted a research to test an off flavor of fermented fruity using category and line scales. • Line scale was applied and it was more effective than category scale in terms of sensitivity even with low intensity testing.

Table 2.2 Continued

Year	Authors	Milestone
2007	Epler, S., *et al.*	• Based on paired preference testing, the hedonic scale resulted in better prediction of optimal sweetness than the JAR scale. • JAR gave a significantly lower score than hedonic scale.
2007	Munoz, A.M. and King, S.C.	• The nine points was translated into several foreign languages and test for validity across many countries including Argentina, Brazil, Canada, China, French, Republic of India, Japan, Republic of Korea, Mexico, The Netherlands, New Zealand, Australia, Philippine, Poland, Spain, Thailand and United Kingdom to see if countries and cultures affect consumer products.
2008	Cardello, A.V., *et al.*	• Conducted an experiment to test the scale end anchors between "Greatest imaginable like/dislike for any experience" and "Greatest imaginable like/dislike." • It was found that no apparent advantage of using different anchor in term of discriminating power. • Using "any experience" restricted the range of scale or created a compression effects.
2008	Hein, K.A., *et al.*	• Comparing 9-point hedonic, labeled affective magnitude and unstructured line scales, they found an equal ease of use and accurate information among three scales. • However, they suggested sample size, product type and type of data produced should be taken into account when selecting a test.
2009	Lim, J., *et al.*	• Developed a Labeled Hedonic Scale (LHS) • LHS yielded identical ratings to those obtained from ME. • LHS obtained a similar result with the 9-point scale
2009	Cook, D.A. and Beckman, T.J.	• Conducted an experiment comparing 5 and 9 point hedonic scale for the mini clinical evaluation exercise. • The nine point scale was found to provide more accurate score (54%) than 5-point scale (44%) while both yielded the same reliability (0.40-0.43).
2009	Villanueva, N.D.M., Maria, A. and Da Silva, A.P.	• The results indicated superiority of the hybrid scale over the traditional hedonic and self-adjusting scales based on MDPREF values of significantly fitted consumers: 79.5 (hybrid scale), 54.5% (self-adjusting) and 51.8% (9-point scale).
2010	Hein, K.A., *et al.*	• The contrast between a natural consumption context and accurate hedonic ratings were observed.

Table 2.2 Continued

Year	Authors	Milestone
2010	Hein, K.A., *et al.* (cont.)	• When removing a product from its natural consumption context, accurate hedonic ratings may not be obtained.
2010	Daroub, H., *et al.*	• Results showed that the 9-point scales either Arabic or English version was equal in terms of reliability, sensitivity, skewness, kurtosis and percent of neutral value.
2010	Nicolus, L., *et al.*	• Foods that were placed in the same verbal category might be given different numerical scores on the second scale. The proportion of those responding differently to the two scales ranged from 100% to 79%. • To check polarity effects: verbal categories (bipolar) and the numbers (unipolar), the experiment was conducted using a bipolar number scale (–4 through 0 to +4). The relative strategy was confirmed for the unstructured numerical scale but the absolute strategy was not confirmed for the scale using only verbal categories
2010	Lawless, H.T., *et al.*	• Three scales: 9-point scale, LAM scale, and an 11-point category scale being compared, it was found that LAM was more preferred to evaluate the acceptability of highly liked foods. • All three scales performed equally well without showing a consistent superiority over another. All three scales were able to differentiate acceptability.
2010	Lim, J. and Fujimaru, T.	• Comparing a 9-point scale and Labeled Hedonic Scale (LHS), both of which have an equal discriminative power. • LHS has more resistance to ceiling effects. • Data obtained from LHS satisfied the normality assumption for statistical analysis. • The misuse of LHS was observed with consumers who had a prior experience with a 9-point scale.

2.3 Advantages and disadvantages of hedonic method

The hedonic scale applied in sensory area was first developed by Peryam and his colleagues in 1954. Afterward, Jones and Thurstone (1955) developed a balanced 9-point hedonic categorical scale; however, the unused highest/lowest categories and the frequent use of a midpoint were questioned. Thirty two years later Peryam developed an 8-point unbalanced scale with more like than dislike categories; it was found somehow better than a 9-point scale but only when use with a well-liked sample. Several scales have been continually developed in the last 5 decades including the labeled affective magnitude (LAM), labeled magnitude scale (LMS), labeled hedonic scale (LHS), magnitude estimation (ME), unstructured line, self-adjusting, ranking scale, hybrid hedonic, oral pleasantness and unpleasantness (OPUS), and positional relative rating (PRR) (Cardello *et al.*, 2008; Cordonnier and Delwiche, 2008; EL Dine and Olabi, 2009; Giovanni and Pangborn, 1983; Green *et al.*, 1993, 1996; Guest, *et al.*, 2007; Lim *et al.*, 2009; McPhearson and Randall, 1985; Peryam and Pilgrim, 1957; Schutz and Cardello, 2001; Warnock *et al.*, 2006) as an alternative choice for assessing food liking. The hedonic scale is unique in terms of general applicability (Cordonnier and Delwiche, 2008; Peryam and Pilgrim, 1957), having a significant ease of use and simplicity (Villanueva *et al.*, 2005), prediction between the target product and the prototypes (Almeida *et al.*, 2002), determination of food choice (Yeomans *et al.*, 2008), offering critical information about individuals' likes and dislikes (Jaeger and Cardello, 2009) and the use of untrained panelists (Daroub *et al.*, 2010).However, Meilgaard *et al.* (2006) mentioned "an exploration into alternative approaches is an ongoing endeavor." Lawless *et al.*, (2010b) and Stone and Sidel, (2004) confirmed that taking efforts to replace a nine-point hedonic scale was not successful.

However, several drawbacks of such scale are several human biases such as error of habituation, contextual and central tendency effects, restricted consumers' freedom, lacking of residual normality, and not reflecting equal difference in perception, (Curia *et al*., 2001; Gay and Mead, 1992; Giovanni and Pangborn, 1983; Lim and Fujimaru, 2010; Marchisano *et al.*, 2003; McPherson and Randall, 1985; Schutz and Cardello, 2001; Villanueva *et al.*, 2000 and Villegas-

Ruiz *et al*,. 2008). The verbal phases of a nine point hedonic scale are hard to examine on the basis of quantitative data regarding a psychological magnitude (McDaniel and Sawyer, 1981; Stevens, 1975). This non-equivalence issue may reduce a mathematical power when statistical analysis is used parametric analysis; however, most researchers still apply a parametric analysis instead of non-parametric analysis (Cardello *et al.*, 2005).

Briefly, a hedonic method is a tool to gain insight information from an actual choice made by consumers in normal consumption environments. There are three main advantages: simple (easy to conduct and understand), cost effective in term of budget and time (the use of untrained panelists), and the data can be analyzed by various statistics (Peryam and Pilgrim, 1957; Daroub *et al.*, 2010). Sensory attribute of the food and its product usage can help consumers determine food acceptance (Booth, 1995). The hedonic scale reflects the attitudes and/or acceptance of consumers toward certain foods under a given condition. However, there are several factors that can influence an experiment not to get a true response. The proper test protocol: the testing experimental plan, sampling procedure, sample preparation and environment (Amerine *et al.*, 1965) could impact consumers' attitudes. Meilgaard *et al.* (2006) also suggested that researchers should be responsible for proper tests with selected target consumers, representative products and cost effectiveness. An improper testing due to testing protocol, experimental design, questionnaire, target consumer, and data analysis could decrease discriminative power of an experiment. A proper protocol is needed to minimize possible extraneous errors. The biases created by consumers during testing will be discussed below.

2.4 Biases of sample presentation

The ultimate goal of a sensory testing is to properly use human subjects as measuring instruments. Sensory scientists do realize that consumers are prone to biases. To properly conduct an experiment, an experimenter should understand a foundation of physiological and psychological biases that can sway sensory perception and scoring. Failure to detect such effects may lead to serious misinterpretation (Macfie *et al*., 1989). The biases found frequently in sensory testing are sample order presentation biases, halo effects, and contrast and convergence effects (Dine and Olabi, 2009) will be discussed below.

2.4.1 Expectation error

Expectation error refers to an error when consumers may intentionally or accidently know about products, research or company. Consumers will consequently use their autosuggestion to judge the product and may disregard perceived product characteristics. For example, during the test period, if a panelist knows that an aged product is being tested or a storage test is being tested, she/he tends to report/focus more on an off flavor whether the product contains such compound or not. The appropriate way to avoid this error is to keep the product detail secret by using the blind coded sample in conjunction with a random presentation (Meilgaard *et al.* 2006).

2.4.2 Error of habituation

Error of habituation refers to an irrelevant error from panelists who tend to give the same response continuously even when a series of samples are served. As a result, a researcher may miss a developing trend and possibly accept a false sample when a sample with small difference is tested. Such biases can be found in quality control process or during the storage test. For example, panelists are asked to evaluate samples daily; the acceptable level will be unintentionally developed so they tend to disregard the subtle difference. One way to avoid this error is to provide proper task instruction, and use balanced and randomized presentation (Meilgaard *et al.*, 2006).

2.4.3 Stimulus error

The word stimulus can be explained as a factor or event that could evoke a specific reaction. Stimulus error can be induced from several unrelated cues (confounding factors) such as light, container color or styles, product un-uniformity, etc. Some consumers tend to use any cues, given or not given to help judge the product instead of rating based on actual perception. The unrelated criteria will be used and it would influence the panelist if the experiment was not well planned. For example, with non-randomized presentation, consumers may expect/rate a sample that will be served at last to be more flavorful. The remedy to this case is to provide proper task instruction, and use balanced and randomized presentation, and avoid leaving irrelevant cues (Meilgaard *et al.*, 2006).

2.4.4 Logical error

The meaning of logic is a reasonable assessment based on prior experience. Prior knowledge or experience can influence on how consumers rated the product if the sample characteristics are related to personal experiences (Meilgaard *et al.*, 2006). This type of error can be induced in conjunction with stimulus error. Consumers are likely to use relevant cues, logically related them to the question of interest, and then score a sample. For example, the more yellowness of mangoes, which indicates more ripening, tends to taste sweeter; the lighter the toasted bread, the less crispiness; the darker the roasted coffee, the stronger the coffee flavor and etc. Such biases can be avoided by keeping the sample uniform, masking any unintentional cues and using balanced and randomized presentation.

2.4.5 Halo effect

The word "halo" means an association of something to ideal or a circle of something resembling. Based on this meaning, the prior attribute evaluated may affect scores of the succeeding attributes. In situation in which consumers simultaneously score several attributes along with overall acceptability; they tend to adjust the scores to correlate with the overall acceptability score. However, the score might be different if those attributes are rated separately. This error has been observed regularly in consumer testing. For example, in a consumer test of grape juice, panelists are requested to rate the sample for 3 different attributes: color, taste and overall acceptability. If subjects evaluate taste and color of the product as like moderately, they are likely to rate the overall acceptability as like moderately as well. In this case, the process of judgment does not involve a direct interest but the initial response sets the range for the subsequent response. Also with many questions and many samples, re-tasting results in physiological fatigue in addition to this halo effect.

2.4.6 Mutual suggestion

The facial expression, the posture and/or the vocalizing opinion can affect others' opinion in either positive or negative way. The response of one observer can influence others. This type of biases generally occurs in consumer testing. The most effective way to solve the problem is to use a separated booth for each panelist while performing the sensory test. Otherwise consumers tend to distract, interfere or interact with each other. Also proper task instructions should be given. The researcher should clearly state that the interactions among subjects are discouraged.

2.4.7 Sample coded biases

People tend to use all intentional or unintentional cues around the product to help with decide their preference scores. Some may try to get some cues from even the blinded code that was originally intended to use to minimize cues. The number of digit can be used ranging from to four digits (several random number tables or the random number generator from the internet has been utilized). Miller (1956) reported that memory span is limited in terms of the largest meaningful unit in the presented material that the person recognizes. Three was an ideal size and meaningful for grouping the letters and numbers. It has been found that some digit should be avoided such as 1, 2, 3 or 4 which may be related to the order or product quality (Lawless and Heymann, 1999). Caul and Raymond (1965) mentioned that numbers or letters e.g., the letter X, the letter-number code A-l, the butter-score numbers 88 or 93 for margarine, or G-11 for soap would also introduce biases.

2.4.8 Order of presentation

The order of sample presentation can influence consumer's perception which consequently causes acceptance score to become inflated or deflated. Several biases can be classified under sub-category "presentation order" biases.

2.4.8.1 Contrast effects

The meaning of contrast is "the state of being different from something else." The contrast effect in sensory field typically means the evaluation bias affected by earlier or previous samples (Amerine *et al.*, 1965; Ferris *et al*., 2003; Lawless and Heyman, 1999). Meilgaard, *et al.* (1999) defined the contrast effect as "The presentation of good quality just before the poor one may cause the subsequence sample to receive a lower score than if it had been rated monadically." Clark and Lawless (1994) referred this bias to the positive correlation of unrelated attributes with the negative correlation being called "horns effect". Even though we know that the contrast effect has pronounced in many cases, no logical relationship or an exactly ratio of correlation has been reported in the literature. Several studied has been conducted to investigate and/or to minimize this bias. Elss *et al.* (2007) studied the potential effect of carry-over in odor and taste off-flavor compounds in orange and apple juice. From the sensory threshold data, it became evidence that carryover effects on several orange juice samples were more obvious for odor than taste. An important carryover flavor component in orange and apple juice is γ-decalactone and limonene. Brockhoff and Sommer (2008) found that the closer the products are the lower the contrast standard error, and, in contrast, the farther the products are the greater the contrast standard error. Cordonnier and Delwiche (2008) observed that the ability to differentiate samples of both traditional 9-point hedonic scale and Positional Relative Rating (PRR) was similar; however, the mean values from PRR were consistently lower than those of the traditional 9-point hedonic. Moreover, they suggested a simultaneous sample presentation rather than a serial monadic presentation with the former showing reduced consistency errors.

2.4.8.2 Centering biases

The word "center" means a middle point of a circle or any circumference. Stevens and Galanter (1957) referred "centering biases" to as "central tendency" or "regression effect." It is likely to happen when panelists match the midpoint of stimulus with the midpoint of the response scale to spare adjustment for further samples having more intense sensation. Poulton (1989) tested identical samples with different scales anchored with "weak" to "strong" and "none" to "moderate," and reported that participants tended to match the midpoint of the stimuli to the midpoint of the response. This resulted in suppression of end category scale usage and limited a discriminative ability (Cardello, *et al.*, 2005). Also Meilgaard *et al.* (1999) found that samples, scales and categories placed near the center tend to be preferred over those placed at the ends. This, consequently, would induce misunderstanding when products or processes need some value to interpolate on a psychophysical function or equation.

2.4.8.3 Positional bias

Change in a sequence of samples tested affect sensory results. Consumers may feel very hungry for the first sample, and very fatigue for the last sample. Often the first sample is rated with bias which results in abnormally preferred or rejected. Meilgaard *et al.* (2006) mentioned that the first sample often encounters a bias for a short period test (sip and evaluate); in contrast, for a long term test, the last sample will expereince the bias. This bias is recognized as one of often observed biases in sensory testing; however, the research devoted for this type of biases is scarce. The popular method used for minimizing this bias is by applying a randomized design. One of the most popular designs is a randomized complete block design (RCBD); however, Macfie *et al.* (1989) found the RCBD was ineffective in reducing the serving-order bias. Another widely used design for minimizing positional biases called “William design” or “carry-over” design, a modification of Latin Square (LS) design was purposed by William in 1948. It is more complicated than RCBD and user unfriendly. Therefore, an attempt to develop an effective design continues.

2.5 Sensory Analysis

The immediate effects on one person leading to performing and picking some products by reasonable processes within that individual’s mind refer to a consumer’s purchasing decision (Booth, 1995; Meilgaard *et al.*, 2006; Moskowitz, 2003). To develop a successful new product, Meilgaard *et al.* (2006) mentioned “Being consumer-centric is a path to success and resources needed to be managed well in a highly value-driven marketplace.” The new product development process must proceed carefully with what can be used to satisfy people’s desires. To use consumer data, a proper sensory data analysis is necessary.

A source of variances generated by human cannot be completely controlled in sensory test. Sensory techniques are used to draw data via the behavioral research and to quantify human responses with fluctuating data inevitably. The non-equivalence scale interval and lacking of residual normality in hedonic testing remain an important issue. However, many researchers prefer to apply the most popular parametric mathematics including an analysis of variance (ANOVA) and post-hoc comparisons, instead of non-parametric analysis for hedonic responses (Cardello *et al.,* 2005). As a result, misinterpretation of data may occur. Subsequently, consumers lose their chance to purchase a great product that they want while companies also cannot sell a truly compelling product (Meilgaard *et al.*, 2006). This confirms why sensory data and analysis contributes to products' success or failure.

2.6 Experimental design

Several experimental designs have been applied for different purposes. For sensory study, in particular, consumer research, consumer responses have been treated as an outcome variable. RCBD has been extensively utilized in order to minimize extraneous effects. However, Macfie *et al*. (1989) and Kunert and Sailer (2007) mentioned that a simple randomize design might not be sufficient for preventing the position and carryover effects. Lee and Meullenet (2010) proposed a method to minimize such effect by removing the first sample score from the experiment; this is not known to be a proper idea. Their method may help minimize a bias in term of serving a doctored sample; however, it created unintentional carry-over, and if unaccounted for, the first presentation, would inflate the error term. For a Latin Square (LS) design, there are concerns about a restriction of LS design that prevents crossing all factor levels with other factor levels, thus limitation of interaction. A better alternative would be a "*Split plot or nested design"*. By applying this design, one could reduce the error term in the model by including a variable into the model for the same reason that adding blocks. This is a simply way to reduce the error term and increasing a power for an experiment. Each panelist is not crossed with all possible combination but is rather put on one set of the sample; this favors a limitation of panelists (number of panelist and consumer's fatigue). "*Split-plot with repeated randomized complete* *block design",* allows blocking the confounding effect, which is generally referred to variation of consumer assessment. The SP design may approach the problem better and could minimize the error term better than regular RCBD design. This dissertation research will explore the possibility.

$$Y_{ijk} = \mu + \rho_k + \alpha_i + \delta_{ik} + \beta_j + (\alpha\beta_{ij}) + \varepsilon_{ijk} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Y_{ijk} is the observed value for the k^{th} replication of the i^{th} level of factor A and the j^{th} level of factor B; where i = 1 to a, j = 1 to b and k = 1 to r.
μ is the grand mean.
ρ_k is the block effect for the k^{th} block; the block effect may be either fixed or random.
α_i is the effect of the i^{th} level of factor A; the effect may be either fixed or random.
δ_{ik} is the whole plot random error effect, for the i^{th}, k^{th} combination of block and factorA
β_j is the effect for the j^{th} level of factor B; the effect may be either fixed or random.
$\alpha\beta_{ij}$ is the interaction effect of the i^{th} level of factor A with j^{th} level of factor B; the interaction effect may be either fixed or random.
ε_{ijk} is the subplot random error effect associated with the Y_{ijk} subplot unit.

This design is useful when there is a random effect, i.e., a panelist effect. The levels of a factor that is chosen at random rather than being fixed are called a random effects model. The response (Y_{ijk}) was computed from the sum of a common value (grand mean). The definitions of each effect for Factor A, Factor B, the interaction effect of Factor A and B, the whole plot random error effect, block effect, and the residual are shown in the equation (1).

2.7 Multicategorical Logit Models

Both quantitative and qualitative data can be obtained from consumer response. If the result can be quantified as continuous values, a regression will be applied otherwise a logistic regression will be performed. Multicategory logistic regression is used to model categorical response variables with more than two categories. The models can be classified into two different versions based on an outcome variable: nominal and ordinal response. The analyses are different but both use the maximum likelihood method. When there are more than two categories, a multinomial distribution will be assumed as the count in the categories of response (Y).

2.7.1 Baseline-Category Logits

This model uses a nominal variable as a response variable. Each model will compare each category with a baseline category. Agresti (2007) mentioned "Multicategory logit models simultaneously use all pairs of categories by specifying the odds of outcome in one category instead of another." Let Z denote the number of categories for Y. Let π denote the response probability of each category (π_1, π_2, …, π_A) and $\sum_A \pi_a = 1$. Because this category treats the response variable as nominal, the order of category is ignored. The baseline is arbitrary but by default the SAS program usually sets the last category (Z) as a baseline-category logit; the log odds model is shown in equation (2). The model shown below when a referred to 1, 2, …, A-1

$$\log (\pi_a/\pi_A) = \alpha_a + \beta_a X \quad \text{(2)}$$

This model will have A-1 equations with separate α and β for each. The model has one less equation because the last category set as a baseline. When there are 2 categories (A=2), the model will be: $\log (\pi_1/\pi_2)$, or it is equal to $\log (\pi_1)$ in ordinary logistic regression for dichotomous responses. Denote $\log (\pi_1) = \log (\pi_1/ 1-\pi_1) = \log (\pi_1/\pi_2)$. However, when there are more than 2 categories, for example, A=3, the model will paire each probability with the baseline category. For instance, A = 1, 2 and 3, the SAS program will provide two possible outcomes with two different α and β for both models: $\log (\pi_1/\pi_3)$ and $\log (\pi_2/\pi_3)$. Then we can calculate the model of $\log (\pi_1/\pi_2)$ as shown in equation (3).

$$\begin{aligned} \log \frac{\pi_1}{\pi_2} &= \frac{\log (\pi_1/\pi_3)}{\log (\pi_2/\pi_3)} = \log \frac{\pi_1}{\pi_3} - \log \frac{\pi_2}{\pi_3} \\ &= (\alpha_1 + \beta_1 X) - (\alpha_2 + \beta_2 X) \\ &= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) X \quad \text{(3)} \end{aligned}$$

The model of $\log (\pi_1/\pi_2)$ will be presented in the form of $\alpha + \beta x$. The intercept (α) is equal to ($\alpha_1 - \alpha_2$) and the slope (β) is ($\beta_1 - \beta_2$). The choice of the baseline is arbitrary and it could be any category decided by the experimenter.

To estimate the response probability, the log odds model can be converted into probability of each category as follows:

$$\pi_a = \frac{e^{\alpha a + \beta a X}}{\sum M \, e^{\alpha M + \beta M X}} \quad \text{……......................... (4)}$$

when a = 1, 2, …, A

Regarding equation (4), the denomination in the equation is always the same for each probability to satisfy $\sum_A \pi_a = 1$. The numerators will change over a summing of denominator ($\sum_M$). For example, from previous sample, A=3, the three probabilities would be:

$$\pi_1 = \frac{e^{\alpha 1 + \beta 1 X}}{1 + e^{\alpha 1 + \beta 1 X} + e^{\alpha 2 + \beta 2 X}}$$

$$\pi_2 = \frac{e^{\alpha 2 + \beta 2 X}}{1 + e^{\alpha 1 + \beta 1 X} + e^{\alpha 2 + \beta 2 X}}$$

$$\pi_3 = \frac{1}{1 + e^{\alpha 1 + \beta 1 X} + e^{\alpha 2 + \beta 2 X}}$$

With SAS program, the last term will set to be "0"; then term "1" in the equation represents term $e^{\alpha 3 + \beta 3 X} = 0$, which denoted $\alpha_3 = \beta_3 = 0$.

2.7.2 Cumulative Logits

When the outcome responses are classified in order, the cumulative logistic regression can be performed. This model uses an ordinal variable as a response variable. The advantages of this model are being simpler and having greater power than the baseline-category logits. The cumulative probability for Y refers to the probability that Y will fall in or below a specific category. The probability can be calculated as following:

$$P\,(Y \leq a) = \pi_1, \pi_2, \ldots, \pi_a$$

when a = 1, 2, …, A

Again to satisfy $\sum_A \pi_a = 1$, the model has one less category, i.e., the last probability $P\,(Y \leq A)$ is set to be redundant. The nature of the scale is ordered so each probability will be ordered and corresponds with $P (Y \leq 1) \leq P (Y \leq 2) \leq P (Y \leq 3) \leq \ldots\ldots \leq P (Y \leq A) = 1$. The logits for cumulative probabilities would be as model 5.

$$\text{Logit}\ \frac{P (Y \leq a)}{1 - (P \geq a)} = \log \frac{\pi_1+\pi_2+\ldots+\pi_a}{\pi_a+1+\pi_a+2+\ldots+\pi_A} = \frac{\log P (Y \leq a)}{\log P (Y \geq a)} \quad \ldots\ldots\ldots\ldots\ldots (5)$$

For example, if there are 3 categories, A=3, the first and second probability would be

$$\text{Logit}\ P (Y \leq 1) = \log \frac{P (Y \leq 1)}{P (Y \geq 1)} = \log \frac{\pi_1}{\pi_2+\pi_3}$$

$$\text{Logit}\ P (Y \leq 2) = \log \frac{P (Y \leq 2)}{P (Y \geq 2)} = \log \frac{\pi_1+\pi_2}{\pi_3}$$

$$P (Y \leq 3) = 1 - P (Y \leq 1) - P (Y \leq 2)$$

The proportional odds can be calculated from a cumulative logit model. Because the model is ordered, the proportional odds represents the odd of response changing when an explanatory factor changes by a single unit. We can compare this model with a binary logistic regression in which the sum of categories 1 to a forms a single category and the sum of categories a+1 to A forms a second category. The logit model will be as following

$$\text{Logit}\ P (Y \leq a) = \alpha_a + \beta X \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

when $a = 1, 2, \ldots, A$

Regarding equation (6), the term β, the effect of explanatory factor on the log odds of response, does not contain any letter subscription because this formula assumes that the effect of X is the same for all A-1 cumulative logits. If this model fits well, it requires only a single parameter, α, to interpret the result rather than A-1 parameters as in a baseline category logistic model. The value of the $|\beta|$ indicates how steeply the slopes increase or decrease. To calculate a probability in each category, the equation will be

$$P (Y = a) = P (Y \leq a) - P (Y \leq a-1) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

For interpretation, the odds ratios can be used to explain the difference between the cumulative logits at two specific values of X. Comparing the cumulative probabilities use, the equation below.

$$\frac{P\,(Y \leq a) \mid X = x2) \,/\, P\,(Y > a) \mid X = x2)}{P\,(Y \leq a) \mid X = x1) \,/\, P\,(Y > a) \mid X = x1)}$$

Because the distance is proportional between two X values, the formula will equal β (x_2-x_1). The odds of response under a specific category multiply by e^{β} implies a change in odds with each unit increase in X. However, when the categories are reversed, the same fit still applies but with opposite sign due to the method of Maximum likelihood that fits the process by a simultaneous iterative algorithm for all a.

2.8 References

Agresti, Alan. 2007. **An introduction to categorical data analysis**. 2nd Ed. Wiley: New York.

Albert, A., Varela, P., Salvador, A., Hough, G. and Fiszman, S. 2011. Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. ***Food Quality and preference,*** 22(5), 463-473.

Almeida, S.B., Aparecida, M. and Da Silva, A.P. 2002. Hedonic scale with reference: performance in obtaining predictive models. ***Food Quality and Preference*** 13(1): 57-64.

Amerine, M.A., Pangbnor, R.N. and Roseller, E.B. 1965. Principles of Sensory Evaluation of Food. New York and London: Academic Press.

Baker, D. and Fortune, S. 2008. Understanding Self-Harm and Suicide Websites: A Qualitative Interview Study of Young Adult Website Users. ***Crisis: The Journal of Crisis Intervention and Suicide Prevention*** 29 (3): 118-122.

Bayarri, S., Carbonell, I. and Tarrega, I.A. 2008. Replicated triangle and duo–trio tests: Discrimination capacity of assessors evaluated by Bayes' rule. ***Food Quality and preference,*** 19(5), 519-523.

Bergara-Almeida, S. and Da Silva, A.A.P. 2002. Hedonic scale with reference: performance in obtaining predictive models. ***Food Quality and preference,*** 13(1): 57-64.

Bi, J. 2007. Similarity testing using paired comparison method. ***Food Quality and preference,*** 18(3), 500-507.

Bleibaum, R.N., Stone, H., Tan, T., Labreche, S., Saint-Martin, E. and Isz, S. 2002. Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices. ***Food Quality and preference,*** 13(6), 409-422.

Booth, D.A. 1995. The cognitive basis of quality. ***Food Quality and preference,*** 6, 201-207.

Bovell-Benjamine, A.C., Dawkin, N., Pace, R.D. and Shikany, J.M. 2009. Use of focus groups to understand African-Americans' dietary practices: Implications for modifying a food frequency questionnaire. ***Preventive Medicine*** 48 (6): 549-554.

Brockhoff, P. B. and Sommer, N.A. 2008. ***Accounting for scaling differences in sensory profile data.*** Sensometrics Meeting: Discover a New World of Data. Technical University, Denmark.

Burnett, E., Lee, K., Rushmer, R., Ellis, M., Noble, M. and Davey, P. 2010. Healthcare-associated infection and the patient experience: a qualitative study using patient interviews. ***Journal of Hospital Infection*** 74 (1): 42-47.

Campo, E., Ballester, J., Langlois, J., Dacremont, C. and Valentin D. 2010. Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: An application to Burgundy Pinot noir wines. ***Food Quality and preference,*** 21(1), 44-55.

Cardinal, P., Flores, A., Contarini, A. and Hough, G. 2003. Focus group research on sensory language used by consumers to described mayonnaise. J***ournal Sensory Studies*** 18: 47-59.

Carkhuff, R.R. and Pierce, R. (1967). Differentiate effect of therapist race and social class upon patient depth of self-exploration in the initial clinical interview. ***Journal of consulting psychology*** 31 (6): 632-634.

Cardello, A.V., Lawless, H.T. and Schutz, H.G. 2008. Effects of extreme anchors and interior label spacing on labeled affective magnitude scales. ***Food Quality and Preference,*** 19: 473-480.

Cardello, A.V., Schutz, H.G. Lesher, L.L. and Merrill, E. 2005. Development and testing of a labeled magnitude scale of perceived satiety. ***Appetite,*** 44: 1-13.

Caul, J.F. and Raymond S.A. 1965. Principle of Consumer Product Testing. J. Soc. Cosmetic Chemists 16: 763-776.

Clancy, K.J. and Krieg, P.C. 2000. ***Counterintuitive Marketing: Achieving Great Results Using Common Sense***. New York: Free Press.

Clark, C.C. and Lawless, H.T. 1994. Limiting response alternatives in time-intensity scaling: an examination of the halo dumping effect. ***Chem. Sense*** 19: 583-594.

Cook, D.A. and Beckman, T.J. 2009. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. ***Adv in Health Sci Educ,*** 14: 655-664.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Cox, D.N., Clark, M.R. and Mialon, V.S. 2001. A cross-cultural methodological study of the uses of two common hedonic response scales. ***Food Quality and Preference,*** 12(2):119-131.

Curia, A , M.C. and Margalef, M.L. 2001. How Argentina consumers understand the Spanish translation of the 9-point hedonic scale. ***Food Quality and Preference,*** 13(2): 217-221

Daroub, H., Olabi, A. and Toufeili, I. 2010. Designing and testing of an Arabic version of the hedonic scale for use in acceptability test. ***Food Quality and Preference,*** 21(1): 33-43

Delarue, J. and Sieffermann, J. M. 2004. Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. ***Food Quality and Preference,*** 15 (4): 383-392.

Dine, A.N. and Olabi, A. 2009. Effect of reference foods in repeated acceptability tests: testing familiar and novel foods using 2 acceptability scales. ***Journal of Food Science*** 74 (2): S97-S106.

Duineveld, C.A.A., Arents, P. and King, B.M. 2003. Determination of equivalence between intensity scales by means of paired comparison. ***Food Quality and Preference,*** 14 (5-6): 419-424.

Elss, S., Kleinhenz, S. and Schreier, P. 2007. Odor and taste thresholds of potential carry-over/off-flavor compounds in orange and apple juice. ***LWT*** 40: 1826-1831.

Eng, T.Y. and Quaia, G. 2009. Strategies for improving new product adoption in uncertain environments: A selective review of the literature. ***Industrial Marketing Management*** 38(3): 275-282.

Epler, S., Chamber, E. and Kemp, K. 2007. Hedonic scales are a better predictor than Just-about-like scales of optimum sweetness in lemonade. ***Journal Sensory Studies*** 13(2):191-197.

Feria-Morales, A.M. 2002. Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. ***Food Quality and Preference,*** 13(6): 355-367.

Ferris, S.J., Kempton, R.A. and Muir, D.D. 2003. Carryover in sensory trials. ***Food Quality and Preference,*** 14: 299-304.

Giovanni, M.E. and Pangborn, R.M. 1983. Measurement of taste intensity and degree of liking of beverages by graphic scales and magnitude estimation. ***Journal of Food Science,*** 48: 1175-1182.

Goldman, A. 1994. Predicting product performance in the marketplace by immediately and extended-use sensory testing. ***Food Technology,*** 48(10): 103-106.

Gosling, J.C.B. 1969. ***Pleasure and desire: the case for hedonism reviewed***, Oxford press.

Goto, K., Bianco-Simeral,S. and Welden, V. 2009. Sensory Evaluation of Locally Grown Walnuts among Consumers and Trained Panelists. ***Journal of Nutrition Education and Behavior*** 41(4): S43.

Green, B.G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K. and Higgins, J. 1996. Evaluating the "labeled magnitude scale" for measuring sensations of taste and smell. ***Chemical Senses,*** 21: 323-335.

Green, B.G., Shaffer, G.S. and Gilmore, M.M. 1993. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. ***Chemical Sense.*** 18: 683-702.

Greene, J.L., Bratka, K.J., Drake, M.A. and Sanders, T.H. 2006. Effective of category and line scales to characterize consumer perception of fruity fermented flavors in peanuts. ***Journal Sensory Studies*** 21:146-154.

Guest, S., Essick, G., Patel, A., Prajapati, R. and MaGlone, F. 2007. Labeled magnitude scales for oral sensations of sweetness, dryness, pleasantness and unpleasantness. ***Food Quality and Preference*** 18: 342-352.

Hautus, M.J., Hout, D.V. and Lee, H.S. 2009. Variants of A Not-A and 2AFC tests: Signal Detection Theory models. ***Food Quality and Preference,*** 20(3): 222-229.

Hein, K.A., Jaeger, S.R., Carr, B.T. and Delahunty, C.M. 2008. Comparison of five common acceptance and preference methods. ***Food Quality and Preference*** 19: 651-661.

Hein, K.A., Hamid, N., Jaeger, S.R. and Delahunty, C.M. 2010. Application of a written scenario to evoke a consumption context in a laboratory setting: Effects on hedonic rating. ***Food Quality and Preference,*** 21(4):410-416

Hetherington, M.M., Pirie, L.M. and Nabb, S. 2002. Stimulus satiation: effect of repeated exposure to foods on pleasantness and intake. ***Appetite*** 38: 19-28.

Huston and Hubson, 2008. Using focus group to inform pharmacy research. ***Research in Social and Administrative Pharmacy*** 4(3): 186-205.

Hyde, A., Howlett, E., Brady, D. and Drennan J. 2005. The focus group method: Insights from focus group interviews on sexual health with adolescents. ***Social Science and Medicine*** 61 (12): 2588-2599.

IFT. 1975. ***Minutes of sensory evaluation division business meeting.*** June 10, Chicago, IL.

Jaeger, S.R. and Cardello, A.V. 2009. Direct and indirect hedonic scaling methods: A comparison of the labeled affective magnitude (LAM) scale and best-worst scaling. ***Food Quality and Preference*** 20(3): 249-258.

Jeon, S.Y., O'Mahony, M. and Kim, K.O. 2004. A comparison of category and line scales under various experimental protocols. ***Jornal Sensory Studies,*** 19: 49-66.

Jones, L.V., Peryam, D.R., and Thurstone, L.L. 1955. Development of a scale for measuring soldiers' food preferences. ***Food Research,*** 20, 512-520.

Jones, L.V. and Thurstone, L.L. 1955. The psychophysics of semantics: An experimental investigation. ***Journal Applied Psychophysic*** 39: 31-36.

Johnton, N.V.S. 1999. ***Why We Feel: The Science of Human Emotions***. Perseus Press: Reading, Mass.

Keeran, J.A. 2004. ***Hedonic scale of emotion***. Emotive energy: A theory of life, mind and emotion. Source: http://www.theoryofmind.org/unified/emotive/hedonic.html. 8th February, 2012.

Kim, M.A., Lee, Y.M., and Lee, H.S. 2010. Comparison of d′ estimates produced by three versions of a duo-trio test for discriminating tomato juices with varying salt concentrations: The effects of the number and position of the reference stimulus. ***Food Quality and Preference,*** 21(5): 504-511.

Koenigsmann, M., Koehler, K., Regner, A., Franke, A. and Frommer, J. 2006. Facing mortality: A qualitative in-depth interview study on illness perception, lay theories and coping strategies of adult patients with acute leukemia 1 week after diagnosis. ***Leukemia Research*** 30 (9): 1127-1134.

Kort, S.J., Kenny, N., Dijk, P., Gevers, S., Richel, D.J. and Willems, D.L. 2007. Cost issues in new disease-modifying treatments for advanced cancer: In-depth interviews with physicians. ***European Journal of Cancer*** 43 (13): 1983-1989.

Kroll, B.J. 1990. Evaluating rating scales for sensory testing w/children. ***Food Technology*** 44(11): 78-80,82, 84, 86.

Kuesten, C.L. 2001. Sequential use of the triangle, 2-AC, 2-AFC, and same-different methods applied to a cost-reduction effort: consumer learning acquired throughout testing and influence on preference judgements. ***Food Quality and Preference,*** 12 (5-7): 447-455.

Kunert, J. and Sailer, O. 2007. Randomization of neighbour balanced generalized Youden designs. ***Journal of Statistical planning and inference***. 137: 2045-2055.

Lassoued, N., Delarue, J., Launay, B. and Michon C. 2008. Baked product texture: Correlations between instrumental and sensory characterization using Flash Profile. ***Journal of Cereal Science.*** 48(1): 133-143.

Lawless, H.T., and Heymann, H. 1999. ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher. 827p.

Lawless, H.T., Sinopoli, D. and Chapman, K.W. 2010. A comparison of the labeled affective magnitude scale and the 9-point hedonic scale and examination of categorical behavior. ***Journal Sensory Studies*** 25: 54-66.

Lee, H.S., Hout, D.V. and Hautus, M.J. 2007. Comparison of performance in the A–Not A, 2-AFC, and same–different tests for the flavor discrimination of margarines: The effect of cognitive decision strategies. ***Food Quality and Preference,*** 18 (6): 920-928.

Lee, Y.S. and Meullenet, J.F. 2010. Comparison of eliminating first order samples for minimizing first serving order bias to data correction. ***Food Science Biotechnol*** 19 (3): 703-709.

Lee, H.J. and Kim, K.O. 2008. Difference test sensitivity: Comparison of three versions of the duo–trio method requiring different memory schemes and taste sequences. ***Food Quality and Preference,*** 7 (1): 97-102.

Letelier, M.F., Spinosa, C. and Calder, B.J. 2000. Taking an expanded view of customer's needs: Qualitative research for aiding innovation. ***Marketing Research*** 12(4): 4-11.

Lim, J. and Fujimaru, T. 2010. Evaluation of the Labeled Hedonic Scale under different experimental conditions. ***Food Quality and Preference,*** 21 (5): 521-530.

Lim, J. Wood, A. and Green, B.G. 2009. Derivation and evaluation of a labeled hedonic scale. ***Chemical Sense*** 34(9): 739-751.

Macfie, H.J.H., Bratchell, N., Greenhuff, K. and Vallis, L.V. 1989. Designs to balance the effect of order of presentation and first-order carry-over effects in hall test. ***Journal Sensory Studies*** 4: 129-148.

McClure, S. and Lawless, H.T. 2010. Comparison of the triangle and a self-defined two alternative forced choice test. ***Food Quality and Preference,*** 21 (5), 547-552.

McDaniel, M.R. and Sawyer, F.M. 1981. Preference testing of whiskey sour formulations: Magnitude estimation versus the 9-point hedonic. ***Journal of Food Science*** 46(1): 182-185.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 1999. ***Sensory Evaluation Techniques***. 3rd edition. Florida: CRC Press. 416p.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 2006. ***Sensory Evaluation Techniques***. 4th edition. Florida: CRC Press. 448p.

Meinert, L., Christiansen, S.C., Kristensen, L., Bjergegaard, C., Aaslyng, M.D. 2008. Eating quality of pork from pure breeds and DLY studied by focus group research and meat quality analyses. ***Meat Science*** 80 (2): 304-314.

Meyners, M. 2007. Easy and powerful analysis of replicated paired preference tests using the $\chi 2$ test. ***Food Quality and Preference*** 18(7): 938-948.

Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". ***Psychological Review*** 63 (2): 81–97.

Moon, S.Y., and Li-Chan, E.C.Y. 2007. Assessment of added ingredient effect on interaction of simulated beef flavour and soy protein isolate by gas chromatography, spectroscopy and descriptive sensory analysis. ***Food Research International*** 40(10): 1227-1238.

Moskowitz, H.R. 1996. Experts versus consumers: A comparison. ***Journal Sensory Studied*** 11: 19-37.

Moskowitz, H.R. 2003. When bad data happen to good researchers: A contrarian's point of view regarding measuring consumer response to food products. ***Food Quality and Preference*** 14: 33-36.

Moskowitz, H.R. and Sidel, J.L. 1971. Magnitude and hedonic scales of food acceptability. ***Journal of Food Science*** 36(4): 677-680.

Munoz, A.M. and King, S.C. 2007. International consumer product testing across cultures and countries. ***ASTM International*** MNL 55

Newman, C. Mao, L., Canavan, P.G., Kidd, M.R., Saltman, D.C., and Kippax, S.C. 2010. HIV generations? Generational discourse in interviews with Australian general practitioners and their HIV positive gay male patients. ***Social science and medicine*** xx: xxx-xxx. (in press).

Nicolson, P. and Burr, J. 2003. What is 'normal' about women's (hetero)sexual desire and orgasm?: a report of an in-depth interview study. ***Social Science and Medicine*** 57 (9): 1735-1745.

Nicolus, L., Marquilly, C. and O'Mahony, M. 2010. The 9-point hedonic scale: Are words and numbers compatible?."***Food Quality and Preference,*** 21(8): 1008-1015.

Nissen, L.R., Byrne, D.V., Bertelsen, G. and Skibsted, L.H. 2004. The antioxidative activity of plant extracts in cooked pork patties as evaluated by descriptive sensory profiling and chemical analysis. ***Meat Science*** 68 (3): 485-495.

Pangborn, R.M. 1964. Sensory evaluation of food: A look backward and forward. ***Food Technology*** 18: 1309.

Peryam, D.R. 1954. Field testing of armed force rations in Peryam, D.R., Pilgrim, F.L. and Peterson, M.S. ***Food Acceptance rating methodology: A symposium***. Advisory board on Quartermaster research and development 75-78. National Research Council, Washington D.C.

Peryam, D.R. and Girardot, N.F. 1952. Advanced taste test method. ***Food Engineering***, 24, 58-61, 194.

Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preference. ***Food Technology*** *(September)*, 9-14

Pfenninger, H.B. 1979. "Methods of quality control in brewing," ***Schweizer Brauerei-Rundschau*** 90: 121.

Pigott, M. 2002. Looking beyond traditional methods. ***Marketing Research*** 14(3): 8-11.

Preston, C.C. and Colman, A.M. 2000. Optimum number of response categories in rating scales: Reliability, Validity, Discrimination power, and Respondent preferences. ***Acta Psychologica,*** 104: 1-15.

Raz, C., Piper, D., Haller, R., Nicod, H., Dusart, N., and Giboreau, A. 2008. From sensory marketing to sensory design: How to drive formulation using consumers' input? ***Food Quality and Preference*** 19 (8): 719-726.

Rook, D.W. 2003. Out-of-focus groups. ***Marketing Research*** 15(2): 10-15.

Sauvageot, F., Herbreteau, V., Berger, M. and Dacremont, C. 2012. A comparison between nine laboratories performing triangle tests. ***Food Quality and Preference,*** 24(1): 1-7

Schutz, H.G. and Cardello, A.V. 2001. A labeled affective magnitude scale for assessing food liking/disliking. ***Journal Sensory Studied*** 16: 117-59.

Stone, H. and Sidel, J. 2004. ***Sensory Evaluation Practices***. 3rd edition. New York: Academic Press. 338p.

Stevens, T. and Ahmedzai, S.H. 2004. Why do breast cancer patients decline entry into randomised trials and how do they feel about their decision later: a prospective, longitudinal, in-depth interview study. ***Patient Education and Counseling*** 53 (3): 341-348.

Thorndike, E.L. 1920. A constant error on psychological rating. ***Journal of Applied Psychology,*** 4: 25-29

Villanueva, N.D.M., Peternate, G.J. & Da Silva, M.A.A.P. 2000. Performance of three affective methods and diagnosis of the ANOVA model. ***Food Quality and Preference*** 11: 363-370.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. ***Food Quality and Preference***, 16: 691-703.

Villanueva, N.D.M. and Da Silva, M.A.A.P. 2009. Performance of the nine-point hedonic, hybrid and self-adjusting scales in the generation of internal preference maps. ***Food Quality and Preference***, 20: 1-12.

Walsh, J.R., White, A.A., and Greaney, M.L. 2009. Using focus groups to identify factors affecting healthy weight maintenance in college men. ***Nutrition Research*** 29 (6): 371-378.

Walter, M., Papachristou C., Danzer G., Klapp, BF. and Frommer J. 2004. 289- Communication with live organ donors before liver transplantation: an in-depth interview study. ***Journal of Psychosomatic Research*** 56 (6): 625.

Wang, R., Zhou, W. and Isabelle, M. 2007. Comparison study of the effect of green tea extract (GTE) on the quality of bread by instrumental analysis and sensory evaluation. ***Food Research International*** 40(4): 470-479.

Warnock, A.R., Shumaker, A.N. and Delwiche, J.F. 2006. Considering of Thurstonian scaling of rating data. ***Food Quality and Preference*** 17: 556-561.

Wichchukit, S. and O'Mahony, M. 2010. Paired preference tests: 'Liking', 'Buying' and 'Take Away' preferences. ***Food Quality and Preference,*** 21(8): 925-929.

William, E.J. 1948. Experimental designs balanced for the estimation of residual effects of treatments. ***Australian Journal of Scientific Research,*** A(2): 149-169.

Yeomans, M.R., Chambers, L., Blumenthal, H. and Blake, A. 2008. The role of expectancy in sensory and hedonic evaluation: The case of smoked salmon ice-cream. ***Food Quality and Preference*** 19(6): 565-573.

CHAPTER 3 AN APPLICATION OF THE SPLIT-PLOT DESIGN TO CONSUMER TESTS

3.1 Justification

Collaboration of a product development team and sensory scientists plays a significant role in generating and commercializing new products under the overall strategic process of product life cycle management. Companies must plan ahead on to what information they want and how to collect and evaluate consumer responses. Several methods including a proper experimental design (Macfie *et al*., 1989; Williams, 1948), a proper product selection (Lee and Meullenet, 2010; Villanueva, *et al.*, 2005) and a proper data analysis (Hottenstein *et al.* 2008) could be used to minimize an extraneous error in an experiment to gain more reliable consumer data.

However, research devoted to the first sample biases is scarce. A simple randomized complete block design (RCBD), which has been extensively used, may not be sufficient to minimize or prevent positional and carryover effects (Macfie et al, 1989; Kunert and Sailer, 2007). Some researchers applied Latin Square (LS) design (Dine and Olabi, 2009; Ferris *et al.*, 2003; Hottenstei *et al.*, 2008; Kermadec and Pages, 2005; Macfie *et al*., 1989; Schlich, 1993; Villanueva, *et al.*, 2005; Wakeling and Macfie, 1995; Williams, 1948) in their sensory work to overcome carryover and positional effects. Nevertheless, this limitation of the LS design makes it unsuitable for consumer testing. LS design ignores the nature of the interaction term which is a major concern in the sensory trial. Also to execute an experiment with the same level for all variables is impossible. In some studies, the first sample score (a dummy sample) was removed during data analysis (Lee and Meullenet, 2010); however, this practice may not be appropriate since there is still a source of variation (first positional biases) that, if unaccounted for, would inflate the error terms. The "*Split-plot with repeated randomized complete block design (SP),"* may be used to solve this problem. By using the SP design, it might help to reduce the error terms by including a known variance into the model, theoretically for the same reason for adding blocks.

3.2 Introduction

Thousands of new products are launched to the food market every year; however, approximately 80-90% of the new products failed to survive (Morris, 1993) due to lack of marketability, profitability and feasibility (Barabba and Zaltman, 1991; Bradley and Nolan, 1998; Clancy and Krieg, 2000; Pine and Gilmore, 1999; Stanton, 1997 and Zaltman, 2003). The critical factors contributing to product success is integration between the inclusive voice of target consumers and a perspective of marketing, sensory scientists and research and development (R&D) teams (Karalaya and Kobu, 1994).

Sensory procedures have been used to determine a product success rate. Appropriated procedures include experimental design (Macfie *et al*., 1989; Williams, 1948), product selection (Lee and Meullenet, 2010; Villanueva, *et al.*, 2005) and data analysis (Hottenstein, *et al*., 2008). However, a major concern of these procedures is that the sensory results are subject to carry-over and serving-order biases, which may inflate the experiment error variance and consequently lead to improper interpretations of the results (Dine and Olabi, 2009; Lawless and Heymann, 1999; Lee and Meullenet, 2010; Meilgaard *et al.*, 2006 and Stone and Sidel, 1993). Meilgaard *et al.* (2006) found that the first-served sample causes biases for a short-term test whereas the last-served sample causes biases for a long-term test induced by carryover and convergence effects. Several experimental designs have been considered to control such biases. These include RCBD and LS designs (Dine and Olabi, 2009; Ferris *et al.*, 2003; Hottenstei *et al.*, 2008; Kermadec and Pages, 2005; Lawless *et al.* 2010; Macfie *et al.*, 1989; Schlich, 1993; Villanueva, *et al.*, 2005;

Wakeling and Macfie, 1995; Williams, 1948). However, a major concern of the LS design is that it requires no interactions between the blocking and treatment factors, which are often an unrealistic assumption in sensory trials. Macfie *et al.* (1989) found that RCBD was ineffective in reducing the serving-order bias.

The purpose of this study was to reduce/minimize psychological biases caused by sample presentation order by utilizing a SP design. Three commercial brands of grape juice were tested. Each participant repeatedly tasted three/four juice samples at a time and rated three attributes of each sample. To reduce the serving-order bias, different random serving orders of juice samples were used. The pairwise tests were carried out to test for the carry-over bias.

3.3 Research specifics

Product: 3 commercial grape juices

Target population: Adults (student, staffs and faculty) age $\geq$ 18 years.

Sampling: a convenience sampling method without specifying genders, races or ages.

Sample size: 540 consumers

Test locations: LSU campus

Risk involved: minimal risks, except for allergy, were involved in the test. Commercially available food products were used for the research. The identity of the individuals was not revealed. It was impossible to connect the results presented to the subjects who participated in the test.

3.4 Material and Method

3.4.1 Material

In this study, three commercial brands of grape juice were pre-selected based on five critical sensory attributes: color, transparency, grape flavor, sweetness, and sourness. Welch's

100% (A); good quality, Welch's light (B); moderate quality, and Juicy Juice DHA (C); bad quality were used as test samples. Five hundred and forty participants (N=540) were randomly recruited at Louisiana State University, including 269 women and 271 men. Each juice sample (about 30 ml.) was poured in a 60 mL (2.0 oz.) clear lidded plastic cup (Propak™ Soufflé clear plastic, Comercializado Por Independent Marketing Alliance, Houston, TX). The 3-digit blinding codes were applied.

3.4.2 Statistical experimental design and analysis

The experiment was divided into two stages. At stage 1, one juice sample was randomly served in a counter-balanced design to each participant. Subsequently, at stage 2 three different juice samples (3) in one of the following (6) random orders: ABC, ACB, BAC, BCA, CAB, and CBA, were repeatedly served to each participant, so that each random order was assigned to (30) participants (3*6*30). All samples were blind tasted; brand names were revealed only after a completion of the tasting. To reduce the presentation protocol errors (Cordonnier and Delwiche, 2008), each participant was exposed to all 3 samples at the same time. Water and unsalted crackers were served as palate neutralizers during the experiment. Re-tasting of products was allowed to refresh memory when needed (Lee and Kim, 2001). After tasting, each participant rated 3 attributes of each sample, including the overall color (OC), overall taste (OT) and overall liking (OL) using a 9-point hedonic categorical scale, where 1 = dislike extremely and 9 = like extremely. Note that our experiment has 3 unique features: 1) random assignment was conducted in both stages, 2) the juice factor was used in both stages, and 3) the factor was repeated in the second stage but not in the first stage. As such, the experiment can be thought of an ad-hoc split-plot experiment with whole-plot and split-plot factors being the same (a juice factor). Table 3.1 presented a layout of this experiment.

Two experiments were executed separately.

3.4.2.1 Randomized Completely Block Design (RCBD) Note that stage 2 alone was a repeated-measures experiment (I) where 3 treatments (juices) were assigned at random. This ad-hoc repeated-measures experiment was analogous to a RCBD with 540 blocks and 3 treatments.

3.4.2.2 Split Plot (SP) A split-plot experiment with 2 stages was treated as if the experiment were a repeated measures experiment with a single stage and 4 treatments, in which 18 possible random assignments were utilized: AABC, AACB, ABAC, ABCA, ACAB, ACBA, BABC, BACB, BBAC, BBCA, BCAB, BCBA, CABC, CACB, CBAC, CBCA, CCAB, and CCBA. This ad-hoc 4-treatment experiment was analogous to the RCBD w/o (RCBD without first served accounted) experiment proposed by Lee and Meullenet (2010). This SP design was used to assess the serving-order and carry-over biases.

Table 3.1 The split-plot experimental layout

Stage 1	Subject ID	Stage 2		
A_1	1-30	A_{21}	B	C
	91-120	A_{22}	C	B
	181-210	B	A	C
	271-300	B	C	A
	361-390	C	A	B
	451-480	C	B	A
B_1	31-60	A_{21}	B	C
	121-150	A_{22}	C	B
	211-240	B	A	C
	301-330	B	C	A
	391-420	C	A	B
	481-510	C	B	A
C_1	61-90	A_{21}	B	C
	151-180	A_{22}	C	B
	241-270	B	A	C
	331-360	B	C	A
	421-450	C	A	B
	511-540	C	B	A

The analysis of variance (ANOVA) was conducted with the MIXED procedure (SAS, 2003) on (I) the ad-hoc 3-treatment experiment, (II) the ad-hoc 4-treatment experiment, and (III) a split-plot experiment. The ANOVA under (I)-(III) in terms of the mean square error (MSE), the estimate of the error variance were compared. A smaller MSE indicates higher accuracy of parameter estimation.

3.5 Results and Discussion

The ANOVA F value was calculated to test if there was a significant difference in the hedonic mean score among 3 brands of grape juice for OC, OT and OL. Table 3.2 presents the ANOVA results. It can be seen the split-plot experiment yielded the smallest mean square error (MSE) for OC, OT and OL (1.79, 2.66, and 2.28, respectively) as compared to the other two designs.

Table 3.2 ANOVA mean squares for the (I) ad-hoc repeated-measures experiment with 3 treatments (RCBD), (II) the ad-hoc repeated-measures experiment with 4 treatments (RCBD w/o), and (III) split-plot experiment (SP).

Attribute	(I)		(II)		(III)	
OC	Juice (J)	2652.16	Juice (J)	2472.70	Whole-plot factor (W)	46.66
	Participant(P)	1.98	Participant(P)	2.04	Split-plot factor (S)	2427.24
	Error (E)	2.24	Error (E)	1.99	W*S	12.64
					Whole-plot error	2.16
					Error	**1.79**
OT	Juice (J)	1711.48	Juice (J)	1547.23	Whole-plot factor (W)	22.43
	Participant(P)	3.46	Participant(P)	3.43	Split-plot factor (S)	1547.08
	Error (E)	3.073	Error (E)	2.98	W*S	6.19
					Whole-plot error	3.67
					Error	**2.66**
OL	Juice (J)	1925.41	Juice (J)	1749.13	Whole-plot factor (W)	31.09
	Participant(P)	2.92	Participant(P)	3.09	Split-plot factor (S)	1748.88
	Error (E)	2.72	Error (E)	2.60	W*S	6.96
					Whole-plot error	3.25
					Error	**2.28**

For OL, the F values for treatment (juice) effects were high for all three designs (1925.41, 1749.13 and 1779.97 (W+S), respectively). To test for the serving-order bias for juice samples in experiments (I) and (II), the Student-t test was performed to see if each brand of juice has the same hedonic means regardless of the order it was served at stage 1 (first position) (Table 3.3).

Table 3.3 Comparisons between hedonic mean scores for juices A, B, C in different serving orders for experiments (I) and (II).

Attribute	Juice	Position 1		Position 2		Position 3	
		I	II	I	II	I	II
OC	A	***7.83***±1.27	***7.54***^{y}±1.13	7.61±1.27	7.63^{xy}±1.17	7.79±1.37	7.79^{x}±1.22
	B	6.62±1.42	6.67^{x}±1.32	6.83±1.18	6.72^{x}±1.36	6.68±1.34	6.84^{x}±1.32
	C	3.74±1.77	3.93^{x}±1.76	3.38±1.78	3.47^{y}±1.59	3.34±1.68	3.36^{y}±1.64
OT	A	7.26±1.63	7.29^{x}±1.53	7.48±1.30	7.30^{x}±1.39	7.33±1.53	7.33^{x}±1.57
	B	6.34±1.76	6.56^{x}±1.66	6.27±1.75	6.17^{y}±1.80	6.49±1.75	6.57^{x}±1.67
	C	***4.24***±2.05	***4.46***^{x}±1.89	4.17±2.11	4.09^{y}±1.98	3.27±1.96	3.37^{z}±1.85
OL	A	7.31±1.56	7.39^{x}±1.35	7.47±1.23	7.34^{x}±1.32	7.42±1.55	7.47^{x}±1.42
	B	6.34±1.60	6.55^{x}±1.59	6.25±1.63	6.22^{y}±1.68	6.53±1.58	6.58^{x}±1.62
	C	***4.01***±1.86	***4.48***^{x}±1.82	3.94±1.92	3.94^{y}±1.84	3.24±1.91	3.32^{z}±1.71

1) Sample calculation template $7.54 = \frac{A_1{:}A_{21} + A_1{:}A_{22} + B_1{:}A_{21} + B_1{:}A_{22} + C_1{:}A_{21} + C_1{:}A_{22}}{180}$ (Table 3.3)

2) Underlined numbers in columns I indicate that at alpha=0.05 the corresponding juices differ significantly in their mean scores (compared among 3 positions) .

3) Numbers in the column II were superscripted with letters x, y, or z. If two numbers do not share a common letter in their superscripts, then at alpha=0.05 their corresponding juices differ significantly in their mean scores (compared among 3 positions) .

4) Bold and Italic numbers in columns of I and II indicate that at alpha=0.05 the corresponding juices differ significantly in their mean scores (compared between design (I&II)).

Table 3.3 presents the results when three serving orders (position 1, 2, and 3) were compared for each brand of juice (A, B, and C) for each attribute OC, OT, and OL. Results are summarized as follows:

3.5.1 Randomized Complete Block Design (RCBD) It can be seen that the serving-order bias was pronounced in experiment (I) for C (bad quality juice) only. In particular, when C was served first, the mean scores for OC, OT and OL as 3.74, 4.24, 4.01 (underlined), respectively, was higher than when C was served last, mean scores 3.34, 3.27 and 3.24, respectively. This confirmed the findings of Meilgaard *et al.* (2006), Popper *et al.* (2004), Vickers *et al.* (1993), Villanueva, *et al.* (2005) and Kermadec and Pages (2005).

3.5.2 Split-plot with repeated randomized complete block design (SP) In experiment (II), each mean hedonic score was superscripted with letters x, y, and z. Two scores did not differ if their superscripts share a common letter. For example, for OC the mean score for C was estimated by 3.93 when C was served first, which was significantly higher than the mean score 3.47 when C was served second or 3.36 when C was served last. It can be seen that the serving-order bias was presented for all brands of juice (A, B and C). In particular, when C was first-served it yielded the highest mean scores of 3.93, 4.46, and 4.48 for OC, OT and OL, respectively. This evidence also supported the idea that presentation biases can be pronounced strongly particularly when testing with a poor quality product.

Significance for the comparison test (experiments (I) and (II)) was denoted by the bold and Italic numbers in columns I and II in Table 3.3. For example, for OC the means for sample A when served first are significantly different (7.83 in (I) vs. 7.54 in (II)). It can be seen from Table 3.3 that when C was served first the mean scores for C in (I) were 4.24 and 4.01, which are significantly lower than 4.46 and 4.48 in (II), respectively for OT and OL. This suggests that the poorest quality juice C carried its treatment effect to the second stage by significantly inflating the mean score for C. The hedonic score of the subsequent sample tended to be higher than those of when A or B served first.

Furthermore, the serving-order bias in stage 2 given the juice served in stage 1 was tested. Table 3.4 presents the results for the attribute OL. It was interesting to note that, if in stage 1 juice C was served, then in stage 2, (1) the mean for juice A when served first was 7.92 higher than when served second (7.13) or third (7.37); (2) the mean for juice B when served first was 7.33, higher than when served second 6.42 or third 6.73; (3) the mean for juice C when served first was 4.70, higher than when served second 4.72 or third 3.14.

Table 3.4 Hedonic mean scores for OL in different serving orders for the split-plot experiment

I	II	Position 1	Position 2	Position 3
A	A	7.50$\pm$0.97	7.30$\pm$1.06	6.87$\pm$1.60
	B	6.17$\pm$1.55	5.97$\pm$1.71	6.50$\pm$1.54
	C	4.10$\pm$1.62	3.57$\pm$1.86	3.60$\pm$1.82
B	A	7.43$\pm$1.33	7.60$\pm$1.18	7.60$\pm$1.25
	B	6.33$\pm$1.50	6.35$\pm$1.46	6.77$\pm$1.64
	C	***4.43***$\pm$1.83	3.50$\pm$1.98	3.27$\pm$1.72
C	A	***7.92***$\pm$1.20	7.13$\pm$1.57	7.37$\pm$1.61
	B	***7.33***$\pm$1.13	6.42$\pm$1.73	6.73$\pm$1.67
	C	***4.70***$\pm$1.92	4.72$\pm$1.39	3.14$\pm$1.59

1) Sample calculation template $7.50 = \frac{A_1{:}A_{21} + A_1{:}A_{22}}{60}$ (Table 3.4)

2) Bold and italic numbers in each row indicate that at alpha=.05 the corresponding juices differ significantly in their mean scores among three positions.

However, if juice A was served in stage 1, no serving-order bias was found in stage 2. Similar findings were also observed for OT and OL, which are not shown here for brevity. This conforms the finding of Lawless and Heymann (1999), Stone and Sidel (1993), Kermadec and Pages (2005), and Brockhoff and Sommer (2008). These findings suggest that the poorest quality juice C carried its effect to the second stage by significantly inflating the mean scores for all juice samples served in the second stage. The split-plot experiment reduces the experimental error variance by accounting for the carry-over bias by including the interaction between the whole-plot and split-plot factors in the model.

Power analysis was conducted to determine the sample size per treatment to attain 80% power for the ANOVA F test for all experiments (I, II and III). Table 5 presents the power values given seven different sample sizes ranging from 20 to 70 and three different significance levels of alpha = .05, .10, and .15. It shows a gradual increase in power of analysis obtained from a split-plot experiment. Figure 1 presents the power curve. The results were compared and it can be seen that at alpha = 0.05, sample size of 30 per treatment was adequate to attain power of 80%

by using the SP design; however, at alpha level higher than 0.05 and a larger sample size over 100, all 3 tested designs yielded similar performance.

Table 3.5 Sample size and power for the (I) ad-hoc repeated-measures experiment with 3 treatments, (II) the ad-hoc repeated-measures experiment with 4 treatments, and (III) split-plot experiment at different alpha levels

Sample size	Att*	α = 0.05			α = 0.10			α = 0.15		
		I	II	III	I	II	III	I	II	III
20	OC	0.741	0.795	0.837	0.836	0.876	0.906	0.884	0.915	0.937
30		**0.905**	**0.937**	**0.958**	**0.949**	**0.968**	**0.980**	**0.968**	**0.981**	**0.988**
40		0.969	0.983	>0.99	0.985	>0.99	>0.99	>0.99	>0.99	>0.99
50		>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99
60		>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99
20	OT	0.598	0.609	0.664	0.719	0.728	0.775	0.788	0.795	0.835
30		**0.792**	**0.802**	**0.849**	**0.873**	**0.880**	**0.913**	**0.912**	**0.917**	**0.942**
40		0.901	0.908	0.939	0.946	0.951	0.969	0.966	0.969	0.981
50		0.956	0.960	0.977	0.979	0.981	0.989	0.987	0.989	>0.99
60		0.981	0.983	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99
20	OL	0.653	0.676	0.735	0.766	0.784	0.831	0.827	0.843	0.881
30		**0.840**	**0.858**	**0.901**	**0.906**	**0.919**	**0.946**	**0.937**	**0.946**	**0.966**
40		0.933	0.944	0.967	0.966	0.972	0.984	0.979	0.983	>0.99
50		0.974	0.979	>0.99	0.988	>0.99	>0.99	>0.99	>0.99	>0.99
60		>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99	>0.99

* Att represents an attribute question; OC = Overall Color; OT = Overall Taste and OL = Overall liking hedonic scores.

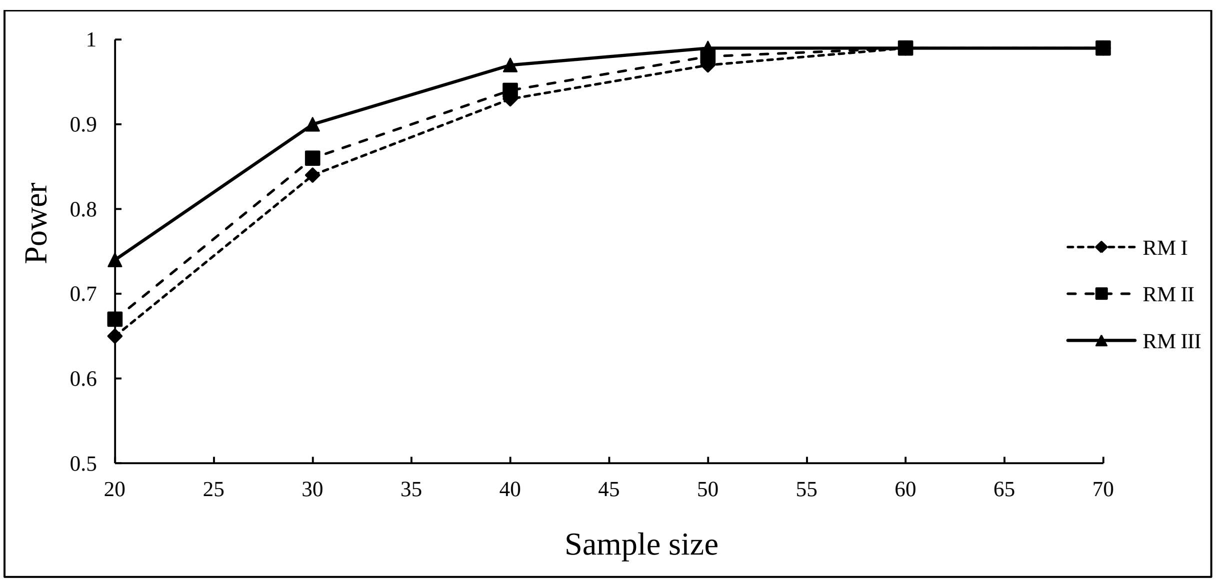

Figure 3.1 Power and sample sizes of the (I) ad-hoc repeated-measures experiment with 3 treatments, (II) the ad-hoc repeated-measures experiment with 4 treatments, and (III) a split-plot experiment

3.6 Conclusion

The results showed that the split-plot experiment could successfully reduce the experimental error variance by removing the carry-over bias from the random errors. By introducing the interaction between the whole-plot and split-plot factors in the model, the poorest quality juice C carried its treatment effect to the second stage by significantly inflating the mean scores for all juice samples served in the second stage. The gradual increase in power of analysis can be obtained from utilizing a split-plot experimental design. To achieve 80% of power with split-plot at alpha = 0.05, a sample size of 30 per treatment was found adequate. However, in reality, one should be aware that the SP design was more effective than the other two designs when testing with a small sample size. When the sample size was greater than 100 consumers, all tested design yielded similar power. All tested design can be performed depending upon the requirement and limitation of each study.

3.7 References

Barabba, V. and Zaltman, G. 1991. ***Hearing the voice of the market: Competitive advantage through creative use of market information.*** Boston: Harvard Business School Press.

Bradley, S.P. and Nolan, R.L. 1998. ***Sense and Response.*** Boston: Harvard Business School Press.

Brockhoff, P.B. and Sommer, N.A. 2008. Accounting for scaling differences in sensory profile data. ***Sensometrics Meeting***: Discovery a new world of data.

Clancy, K.J. and Krieg, P.C. 2000. ***Counterintuitive Marketing: Achieving Great Results Using Common Sense.*** New York: Free Press.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Dine, A.N. and Olabi, A. 2009. Effect of reference foods in repeated acceptability tests: testing familiar and novel foods using 2 acceptability scales. ***Journal of Food Science*** 74 (2): S97-S106.

Ferris, S.J., Kempton, R.A. and Muir, D.D. 2003. Carryover in sensory trials. ***Food Quality and Preference,*** 14: 299-304.

Hottenstein A., Taylor, R. and Carr T. 2008. Preference segments: A deeper understanding of consumer acceptance or a serving-order effect? ***Food Quality and Preference,*** 19: 711-718.

Karalaya, F. and Kobu, B. 1994. New product development process: An investigation of success and failure in high-technology and non-high-technology firms. ***Journal of Business Venturing*** 9(1): 49-66.

Kermadec, F.H. and Pages, J. 2005. Methodology to analyse rank and carry-over effects. ***Food Quality and Preference,*** 16: 600-607.

Kunert, J. and Sailer, O. 2007. Randomization of neighbour balanced generalized Youden designs. ***Journal of Statistical planning and inference***. 137: 2045-2055.

Lawless, H.T., and Heymann, H. 1999. *Sensory Evaluation of Food: Principal and Practices.* Maryland: Aspen publisher. 827p.

Lawless, H.T., Sinopoli, D. and Chapman, K.W. 2010. A comparison of the labeled affective magnitude scale and the 9-point hedonic scale and examination of categorical behavior. ***Journal Sensory Studies*** 25: 54-66.

Lee, Y.S. and Meullenet, J.F. 2010. Comparison of eliminating first order samples for minimizing first serving order bias to data correction. ***Food Science Biotechnol*** 19 (3): 703-709.

Lee, H.J. and Kim, K.O. 2001. Effect of forgetting on various protocols for category and line scales of intensity. ***Journal Sensory Studies*** 16: 327-342.

Macfie, H.J.H., Bratchell, N., Greenhuff, K. and Vallis, L.V. 1989. Designs to balance the effect of order of presentation and first-order carry-over effects in hall test. ***Journal Sensory Studies*** 4: 129-148.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 2006. ***Sensory Evaluation Techniques.*** 4th edition. Florida: CRC Press. 448p.

Morris, C.E. 1993. Why new product fail. ***Food Engineering*** 65 (6): 132-136.

Pine, B.J.II. and Gilmore, J.H. 1999. ***The experience economy.*** Boston: Harvard Business School Press.

Popper, R., Rosenstock, W., Schraidt, M. and Kroll, B.J. 2004. The effect of attribute questions on overall liking ratings. ***Food Quality and Preference,*** 15: 853-858.

SAS 2003. ***SAS/IML software: Usage and reference***, Version 9. SAS Institute Inc., Cary, NC, USA.

Schlich, P. 1993. Use of change-over designs and repeated measurement in sensory and consumer studies. ***Food Quality and Preference,*** 4: 223-235.

Stanton, J.L. 1997. Who's no. 1? Consumer! ***Food Processing*** December: 55-57.

Stone H. and Sidel, J.L. 1993. ***Sensory Evaluation Practises.*** 2nd. Academic press, Redwood City, California. 338 pages.

Vickers, Z.M., Christensen, C.M., Fahrenholtz, S.K. and Gengler, I.M. 1993. Effect of questionnaire design and the number of samples tasted on hedonic rating. ***Journal Sensory Studies,*** 8: 189-200.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. ***Food Quality and Preference***, 16: 691-703.

Wakeling, I.N. and Macfie, H.J.H. 1995. Designing consumer trials for first and higher orders of carry-over effect when only a subset of k samples form P may be tested. ***Food Quality and Preference,*** 6: 299-308.

William, E.J. 1948. Experimental designs balanced for the estimation of residual effects of treatments. ***Australian Journal of Scientific Research,*** A(2): 149-169.

Zaltman, G. 2003. ***How customer think.*** Boston: Harvard Business School Press.

CHAPTER 4 ANALYZING CONTRAST EFFECTS IN CONSUMER ACCEPTANCE TESTS USING LOGISTIC REGRESSION

4.1 Justification

One of the most perceptual biases that consumers have when sequentially tasting and evaluating foods in consumer acceptance taste tests is “contrast effect.” Consumers tend to contrast between samples on several dimensions (rather than on the attribute of interest), and evaluate them against one another. Several publications have confirmed that the contrast effect is presented during the taste testing period; however, currently there have no method to quantify the probability that the contrast effects will happen.

Both regression and logistic regression can be used for prediction of outcomes. The differences between two methods are the types of outcome variables and the method used to analyze a result. Theoretically, the hedonic scale can be classified as either a nominal, an ordinal or interval scale depending on the assumption. However, it has been found that the hedonic data are hard to interpret based on each categorical value corresponding with a word description. Even though many researchers have claimed that the hedonic scale is evenly spaced with an equal distance between points from 1 to 9, it actually has no emotionally equal space due to human psychological perception. Utilizing a baseline logistic regression (nominal response) may be more reasonable in this case so that we can compare the effect of differently rated product classifications on the change in the hedonic score when increasing or decreasing the magnitude of difference. More detail regarding the two different logistic regressions can be found in literature reviews (chapter 2).

4.2 Introduction

In consumer acceptance taste tests, the contrast effect is one of the most perceptual biases that consumers have when sequentially tasting and evaluating foods. Whether the test sample may be perceived as worse or better than it should be can depend on whether it is served before or after another good or bad test sample. For product acceptance, the constrast effect occurs if the food receives a lower/higher hedonic score the preceeding food has good/poor-quality (Meilgaard, *et al.*, 1999). As such, one way to measure the contrast effect is to take the difference in the mean hedonic score between sequentially served foods (Ball, 1997; Boss and Stufken, 2007; Clouser- Roche *et al.*, 2008; Elss *et al.*, 2007; Ferris *et al.*, 2003; Kamenetzy, 1959; King *et al.*, 2003; Walter and Boakes, 2009). The analysis of variance (ANOVA) has been used to test for the contrast effect or if the mean hedonic scores for the foods are the same regardless of which food is previously served (Cordonnier and Delwiche, 2008). Several scales have been used to access degree of liking including the labeled affective magnitude (LAM), labeled magnitude scale (LMS), labeled hedonic scale (LHS) and oral pleasantness and unpleasantness (OPUS) (Cardello *et al.*, 2008; Dine and Olabi, 2009; Green *et al.*, 1993; Guest, *et al.*, 2007; Lim, *et al.*, 2009; Schutz and Cardello, 2001; Warnock *et al.*, 2006); however, the attempting to overcome a 9 point hedonic categorical scale has not been successfully. In this study, we will apply this scale to determine the sensory contrast effects.

For ANOVA F testing, when significance was found, one may wish to know it the contrast effect had occurred, and if so, what would be its possibility. Currently there is no literature dealing with this issue. This paper served this purpose in part. To accomplish this task, we conducted an experiment using three grape juice brands (A, B, C). Based on preliminary ranking test, A, B and C were classified as “good,” “moderate” and “bad” quality products, respectively. Each participant tasted sequentially two juices and evaluated three attributes of each juice: the overall color (OC), overall taste (OT) and overall liking (OL), on a 9-point hedonic scale. For each participant, we took the difference in the hedonic score between the two juices, and transformed these differences according to whether the juice served second has a higher hedonic score than the juice served first, coded as -1 (if lower), 1 (if higher), or 0 (if equal). We used logistic regression to fit these categorical data with the outcome (-1, 0, 1) and the covariate, the first juice. With the fitted model we calculated the odds ratios as well as predicted the probability that each juice would have a higher/lower hedonic score than the first served juice.

4.3 Research specifics

Product: 3 commercial grape juices

Target population: Adults (student, staffs and faculty) age ≥ 18 years

Sampling: a convenience sampling method without specifying genders, races or ages.

Sample size: 540 consumers

Test locations: LSU campus

Risk involved: minimal risks, except for allergy, were involved in the test. Commercially available food products were used for the research. The identity of the individuals was not revealed. It was impossible to connect the results presented to the subjects who participated in the test.

4.4 Material and Method

4.4.1 Material

In this study, three grape juice brands were pre-selected, Welch's 100% (A), Welch's light (B), and Juicy Juice DHA (C), according to 5 critical sensory attributes: color, transparency, grape flavor, sweetness, and sourness. Five hundred and forty participants were recruited at Louisiana State University, including 269 women and 271 men with an age range of 20-60 years old. Clear lidded plastic cups 60 mL (2.0 oz.) (Propak™ Soufflé clear plastic, Comercializado Por Independent Marketing Alliance, Houston, TX) were used. Each cup was half filled with each brand of juice. Samples were presented "blind" in a counter-balanced order. Juices were labeled with a random three-digit number to prevent ordering bias. To reduce the presentation protocol errors (Cordonnier and Delwiche, 2008), both juices were simultaneously presented to each participant. Water and unsalted crackers were served as palate neutralizers during the experiment. Re-tasting was allowed to refresh memory only if necessary (Lee and Kim, 2001).

4.4.2 Statistical experimental design and analysis

In this experiment two out of three grape juices were sequentially served to each participant following one of the 9 random serving orders: AA, AB, AC, BA, BB, BC, CA, CB, and CC, so that each random order was assigned to 60 participants (9*60=540). After tasting, each participant rated 3 attributes of each juice including overall color (OC), overall taste (OT) and overall liking (OL) using a 9-point hedonic categorical scale, where 1 = dislike extremely and 9 = like extremely. This resulted in a repeated measures experiment with 3 treatments (A, B, C) and 540(2) = 1080 observations, so that each treatment has 1080/3=360 replicates.

4.5 Objectives and their corresponding statistical analysis

The objectives were to quantify the probability of the sensory contrast effects as follows 4.5.1. Testing the contrast effect of differently rated product classifications (good (A), moderate (B) and bad (C) quality) on consumer acceptance scores, 4.5.2) Quantifying a contrast effect in consumer testing. Three different categories were used for each pair of sample: negative (hedonic score decreases for the second served sample), positive (hedonic score increases for the second served sample) and unchanged score (hedonic scores were the same; used as baseline). Two fixed factors were defined as first juice serving (FJ) and second juice serving (SJ) with three levels each (classified as good, moderate and poor quality juice) and 4.5.3) Predicting the probabilities of paired products using LRA regarding contrast effects.

4.5.1 Testing the contrast effect

Theoretically, the constrast effect occurs if the juice hedonic score tends to be inflated when a poor-quality juice is served before a good-quality juice (Lawless and Heymann, 1999; Kermadec and Pages, 2005). To verify this theory, the difference in the mean hedonic score between two juices was calculated for each participant, and then ANOVA was conducted on these difference scores. The ANOVA F value was calculated to test if the contrast effect existed. If the ANOVA F test was significant, the Tukey's procedure was subsequently conducted to calculate the confidence interval (CI) estimate of the mean hedonic score for the second juice. These CI estimates were then compared to assess the contrast effect. The ANOVA approach was conducted in the MIXED procedure in the statistics software SAS (SAS, 2003).

4.5.2 Analyzing the contrast effect using logistic regression

For each participant, the difference in the hedonic score was taken between the two juices (SJ - FJ) and transformed to coded data: -1 (if lower), 1 (if higher), or 0 (if same). The outcome variable, denoted by y, was assumed to have a multinomial distribution with 3 response categories: (-1, 0, 1).

Table 4.1 The template for the coded data.

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	.	.	540
FJ	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	.	.	C
SJ	A	A	A	B	B	B	C	C	C	A	A	A	B	B	B	.	.	C
Y	-1	0	1	0	-1	1	0	0	1	0	1	-1	-1	0	0	.	.	0
Probability	P_{11}	P_{21}	P_{31}															

FJ: The first-served juice product; SJ: the second-served juice product; Y: the outcome variable

Table 4.1 presents a template for the coded data. The y = 0 category was chosen as the reference level or baseline, which was the standard against the other two categories. The odds as an expression of the relative probability or the ratio of the probability that y falls into the jth (j=1, 2) category over the probability that y falls into the reference category (j=0).

The logit model is a regression model that predicts the odds of different categories of Y given a set of explanatory variables or covariates x (which may be continuous or discrete). Here, to predict the odds using the covariate x, the first juice (A, B or C), and the reference category (i.g. juice C). This created two dummy variables: x_1 and x_2, each taking 0 or 1, as follows:

First Juice	x_1	x_2
A	1	0
B	0	1
C	0	0

For the ith (i = 1,..,540) participant and the jth (j=0, 1, 2) outcome category, the probability that the ith observation fall into the jth outcome category, denoted by P_{ij} can be written as

$$P_{ij} = P(y_i = j) = \frac{\exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i})}{1 + \sum_{j=1}^{2} \exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i})}$$

where y_i and x_i are the corresponding observed values of y and x for the ith participant.

The model has an equivalent formulation. The logit of P_{ij}, defined as the natural log of the odds P_{ij} / P_{i0}, is modeled as a linear function of the x_{1i} and x_{2i}:

$$\text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{P_{i0}}\right) = \beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} \qquad (4.1)$$

For each first juice A, B, or C, the logit model is as follows (equation 4.2):

$$\begin{array}{lll} \text{C}: x_1 = x_2 = 0 & \log(odds_C) = \beta_{0j} & \\ \text{A}: x_1 = 1, x_2 = 0 & \log(odds_A) = \beta_{0j} + \beta_{1j} & (4.2) \\ \text{B}: x_1 = 0, x_2 = 1 & \log(odds_B) = \beta_{0j} + \beta_{2j} & \end{array}$$

Hence, the coefficient β_{0j} was the natural log of the odds for the reference category, juice C which referred to a bad quality juice or disliked product in this study. To be able to compare all events, two equations were estimated. $\beta_{1j} = \log(odds_A) - \log(odds_C) = \log\left(\frac{odds_A}{odds_C}\right)$ can be interpreted as the log odds ratios of the odds for A over the odds for C. Likewise, β_{2j}, was interpreted as the log odds ratio for B over the odds for C. These unknown β coefficients were typically estimated iteratively by the maximum likelihood through reweighted least squares.

4.5.3 Prediction of the probability for paired products using LRA

To predict the probability that each of the 9 paired juices: AA, AB, AC, BA, BB, BC, CA, CB, and CC fall into a category of the outcome variable, we chose the reference category as CC and created four dummy factors: x_1- x_4, each taking 0 or 1, as follows:

First Juice	x_1	x_2
A	1	0
B	0	1
C	0	0

Second Juice	x_3	x_4
A	1	0
B	0	1
C	0	0

And as in 4.3, the probabilities were modeled as the function of the x_i's.

$$P_{ij} = P(y_i = j) = \frac{\exp(\beta_{0j} + \beta_{1j}x_{1i} + \dots \beta_{4j}x_{4i})}{1 + \sum_{j=1}^{4} \exp(\beta_{0j} + \beta_{1j}x_{1i} + \dots \beta_{4j}x_{4i})} \qquad (4.3)$$

For example, for the pairs BC and AB, the probabilities were calculated as:

$$\text{BC}: \ x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0 \qquad P_{ij} = \frac{\exp(\beta_{0j} + \beta_{2j})}{1 + \sum_{j=1}^{4} \exp(\beta_{0j} + \beta_{2j})}$$

$$\text{AB}: \ x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1 \qquad P_{ij} = \frac{\exp(\beta_{0j} + \beta_{1j} + \beta_{4j})}{1 + \sum_{j=1}^{4} \exp(\beta_{0j} + \beta_{1j} + \beta_{4j})}$$

4.6 Results and Discussion

4.6.1 Testing the contrast effect

To test the existence of sensory contrast effects, the CI estimates of each pair are shown in Table 4.2

Table 4.2 Hedonic scores for the second served juice for a given first served juice for OL, OC and OT

Second Juice (SJ)		First Juice (FJ)		
		A	B	C
A	Overall liking	7.37± 1.09	7.43 ± 1.33	7.37 ± 1.61
B		5.98 ± 1.78^{b}	6.33 ± 1.50^{b}	**7.33 ± 1.13^{a}**
C		4.32 ± 1.72	4.43 ± 1.83	4.70 ± 1.92
A	Overall color	7.35 ± 1.20	7.50 ± 1.23	7.78 ± 0.89
B		6.17 ± 1.54^{b}	6.70 ± 1.09^{b}	**7.45 ± 0.95^{a}**
C		**3.90 ± 1.56^{a}**	3.38 ± 1.56^{b}	**4.50 ± 1.99^{a}**
A	Overall taste	7.30 ± 1.25	7.37 ± 1.56	7.20 ± 1.77
B		6.10 ± 1.71^{b}	6.35 ± 1.66^{b}	**7.23 ± 1.41^{a}**
C		4.48 ± 1.66	4.63 ± 2.00	4.80 ± 2.00

Different letters in each row indicated that at alpha = 0.05, the hedonic scores differ significantly compared across FJ.

The ANOVA F test statistics for testing the contrast effect for the three outcome variables OL, OC, and OT were, respectively, 31.73, 43.97, and 91.98. Because each F value was greater than the critical value F (9-1, 548-8) = F (8,530) = 2.22, we concluded that the contrast effect existed significantly. We then used the Tukey's procedure to calculate the confidence interval (CI) estimate for the mean hedonic score for the second juice. We compared these CI estimates, as shown in Table 4.2, to assess the contrast effect. In each row of Table 4.2, if a mean score was highlighted in bold, it was different from the other two mean values. For example, for OL and the second juice B, the CIs were estimated at 5.98 ± 1.78, 6.33 ± 1.50, and **7.33 ± 1.13** when juices A, B, and C were first served, respectively. Thus indicates that juice B has a significantly lower hedonic mean score when served after juice A or B than when served after juice C. Similarly, for OC and OT the CIs for juice B when served after juice C were, respectively, estimated at 7.45 ± 0.95 and 7.23 ± 1.41 which were those greater than when served after juice A or B. These results indicated the contrast effect was significant, resulting in inflated score particularly for the CB (moderate-poor quality juices). These findings confirmed the finding of Lawless and Heymann, (1999); Kermadec and Pages (2005) and Villanueva, *et al.* (2005). However, the contrast effect was not for the second served observed A or C samples. For testing the second-served C product, a high score fluctuation (high standard deviations) may contribute to an insignificant result. Human psychological biases and/or individual preferences may contribute to cause this phenomenon.

4.6.2 Analyzing the contrast effect using logistic regression

The estimates of the parameters shown in model (4.2) are shown in Table 4.3.

Table 4.3 Parameter estimates, β, for OL, OC and OT for odd estimation

	β_{01}	β_{02}	β_{11}	β_{12}	β_{21}	β_{22}
OL	-0.1483	1.4689	1.3721	-1.766	0.9624	-0.8346
OC	-1.3858	1.3218	2.3552	-2.1102	2.1670	-0.8818
OT	0.4925	1.7130	0.9545	-1.5486	0.6189	-0.6536

A multinomial distribution with 3 categories: (-1, 0, 1) where β_{01} referred to parameter estimates for sample C and the -1 outcome category; β_{02} referred to parameter estimates for sample C and the +1 outcome category; β_{11} referred to parameter estimates for sample A and the -1 outcome category; β_{12} referred to parameter estimates for sample A and the +1 outcome category; β_{21} referred to parameter estimates for sample B and the -1 outcome category; β_{22} referred to parameter estimates for sample B and the +1 outcome category;
* Odds estimation is not shown. To calculate odds, for example, odds A for OL = exp ($\beta_{01}+\beta_{11}$) = exp (-0.1483+1.3721)

The odds and odds ratios can be manually calculated accordingly, as shown in Table 4.3-4.4. For example, for OL and the -1 outcome category, the odds for juice C can be calculated as exp (β_{01}) = Exp (-0.1483) = 0.86, indicating the odds that the juice second-served C has a lower score than C were 0.86. For OC, the odds ratio for A vs. C was exp (β_{11}) = exp (2.3552) = 3.944, indicating the odds when juice A was served were about approximately 4 times as great as the odds when juice C was served. In other words, the chances of second served sample to have score inflation when juice A served first was 4 times higher than that of with juice C. For OL, the odds ratio for B vs. C can be calculated as exp(β_{21}) = exp(0.9624)= 2.618, indicating the likelihood of second served sample to be inflated when juice B served first was roughly 3 times higher than that of when juice C was served. For OC, the odds and odds ratios (Table 4.4) were exp (-1.385) = 0.25 (C), 10.54 (A vs. C) and 8.73 (B vs. C) and for OT, there were exp (0.4925) = 1.64 (C), 2.6 (A vs. C) and 1.86 (B vs. C), respectively. On the other hand, for the +1 outcome category, the odds and odds ratios were calculated as follows: for OL, exp (β_{02}) = exp(1.4689) = 4.34 (C), 0.17 (A vs. C) and 0.43 (B vs. C); for OC, exp (1.3218) = 3.75 (C), 0.12 (A vs. C) and 0.41 (B vs. C); for OT exp(1.7130)= 5.55 (C) , 0.21 (A vs. C) and 0.52 (B vs. C), respectively.

Table 4.4 Odds ratios for attributes OC, OT and OL

First Juice	Outcome category	OL	OC	OT
A vs. C	-1/0	3.94	10.54	2.60
	+1/0	0.13	0.06	0.18
B vs. C	-1/0	0.17	0.12	0.21
	+1/ 0	0.43	0.41	0.52

* Odds estimation can be manually calculated from Table 4.3. To calculate odds ratios, for example, odds A vs. odds C for OL = exp ($\beta_{01}+\beta_{11}$)/exp (β_{01}) = exp (β_{11})

The above results indicate that the odds of -1 (hedonic scores tended to decrease after the first sample served) were greatest if the first juice served was the good quality juice. Likewise the odds of +1 (hedonic scores tended to increase after the first sample served) were greatest if the first juice served was the bad quality, i.e., juice C. Based on this study, it proved that the contrast effects had pronounced and it contributed to score inflation and/or deflation depending on the previous sample served.

4.6.3 Prediction of the probability for paired products using LRA

Table 4.5 presents the parameter estimates for the probability models (4.2) given in order to determine the reliability of result for duplicating sample testing. The method of maximum likelihood was approached to estimate the β parameters. With these estimates, we then can calculate the predicted probabilities that each of the 9 paired juice AA, AB, AC, BA, BB, BC, CA, CB, and CC fall into each outcome category (-1, 0, 1), for each attribute question (OL, OC and OT) as given in Table 4.6.

Table 4.5 Parameter estimates, β, for OL, OC and OT for probability estimation

Parameter	Outcome category	OL	OC	OT
β_{01}	-1	0.20	-1.00	0.79
β_{02}	+1	0.65	0.25	1.05
β_{11}	-1	1.67	3.63	1.12
β_{12}	+1	-2.08	-2.88	-1.69
β_{21}	-1	1.15	3.11	0.71
β_{22}	+1	-1.05	-1.53	-0.74
β_{31}	-1	-1.19	-2.70	-1.09
β_{32}	+1	1.45	2.09	0.92
β_{41}	-1	-0.73	-1.76	-0.24
β_{42}	+1	1.07	1.61	1.10

* Probability estimation = odds / odds +1

For example, for OL, when juice A was served first, the probabilities that 3 juice pairs AA, AB and AC falling into the -1 outcome category were, respectively, 0.49, 0.65 and 0.84. When juice C was served first, the probabilities that 3 juice pairs CA, CB and CC fall into the +1 outcome category were 0.86, 0.78 and 0.46, respectively.

Table 4.6 Probabilities of falling (-1, 0, 1) response categories for each paired samples

Attributes	Paired juice	-1	+1	0
OL	AA	0.49	0.26	0.25
	AB	0.65	0.14	0.21
	AC	0.84	0.03	0.13
	BA	0.23	0.57	0.20
	BB	0.39	0.40	0.21
	BC	0.70	0.12	0.18
	CA	0.04	0.86	0.10
	CB	0.08	0.78	0.14
	CC	0.30	0.46	0.24
OC	AA	0.37	0.23	0.40
	AB	0.64	0.10	0.26
	AC	0.93	0.00	0.07
	BA	0.15	0.59	0.26

Table 4.6 Continued

Attributes	Paired juice	-1	+1	0
OC	BB	0.37	0.37	0.26
	BC	0.87	0.03	0.10
	CA	0.00	0.91	0.09
	CB	0.01	0.86	0.13
	CC	0.14	0.48	0.38
OT	AA	0.49	0.29	0.22
	AB	0.67	0.20	0.13
	AC	0.82	0.06	0.12
	BA	0.25	0.58	0.17
	BB	0.41	0.47	0.12
	BC	0.65	0.20	0.15
	CA	0.08	0.81	0.11
	CB	0.15	0.76	0.09
	CC	0.36	0.47	0.17

A multinomial distribution with 3 categories: (-1, 0, 1) where -1response category means the second-served hedonic scores tended to be decreased; +1 response category means the second-served hedonic scores tended to be increased and 0 response category means the second-served hedonic scores remained the same values.

These results observed that when a poor-quality juice C was served first, juices A and B are 81% and 76% likely to have a higher hedonic score than juice C; when a good-quality juice A was served first, juices B and C are 65% and 84% likely to have a lower hedonic score than juice A. Similar findings were also obtained from OT and OC. Furthermore, the predicted probabilities were larger when the paired juices containing C, i.e., the AC, CA, BC, or CB.

Interestingly, for the probability of getting a same score for the two identical samples testing, there was 20-25% (for overall liking testing of AA; BB and CC) of consumers tended to change their score. This implies that there is much variation in consumer testing. Approximately 75-80% of consumers change their score even for the two identical samples were tested. However, it was found a higher probability of getting same score for the identical samples tested on overall color (26-40%) than overall taste (12-22%). This implies that rating a more complex/multidimensional attribute (i.e., taste) tended to be more affected by the contrast effects than a simpler attribute (i.e., color).

4.7 Conclusion

In this study, we measured the contrast effect by taking a difference in the hedonic mean score between two juices sequentially served to each participant. The ANOVA approach was used to test for the existence of contrast effects. We categorized these difference scores into 3 categories (-1, 0, 1) and used logistic regression to fit these categorical data. With the fitted model, we predicted the probability that the juice would have a higher/lower hedonic score than the juice served before it. Results showed that the odds were largest if the first juice had the best/worst quality, and the predicted probabilities were largest when two juices were strongly contrasted. A careful experimental design and proper product selection must be applied in order to minimize the contrast effect in consumer testing.

4.8 References

Ball, R.D. 1997. Incomplete block designs for the minimization of order and carryover effects in the sensory analysis. ***Food Quality and Preference*** 8: 11-118.

Boss, M. and Stufken, J. 2007. Optimal crossover designs when carryover effects are proportional to direct effects. ***Journal of Statistical Planning and Inference*** 137: 3291-3302.

Cardello, A., Lawless, H.T. and Schutz, H.G. 2008. Effect of extreme anchors and interior label spacing on labeled affective magnitude scales. ***Food Quality and Preference*** 19: 473-480.

Clouser- Roche, A., Johnson, K., Fast, D. and Tang, D. 2008. Beyond pass/fail: A procedure for evaluating the carryover effect in bioanalytical LC/MS/MS method. ***Journal of Pharmaceutical and Biomedical Analysis*** 47:146-155.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Dine, A.N. and Olabi, A. 2009. Effect of reference foods in repeated acceptability tests: testing familiar and novel foods using 2 acceptability scales. ***Journal of Food Science*** 74 (2): S97-S106.

Elss, S., Kleinhenz, S. and Schreier, P. 2007. Odor and taste thresholds of potential carry-over/off-flavor compounds in orange and apple juice. ***LWT*** 40: 1826-1831.

Ferris, S.J., Kempton,R.A. and Muir, D.D. 2003. Carryover in sensory trials. ***Food Quality and Preference*** 14: 299-304.

Green, B.G., Shaffer, G.S. and Gilmore, M.M. 1993. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. ***Chemical Sense.*** 18: 683-702.

Guest, S., Essick, G., Patel, A., Prajapati, R. and MaGlone, F. 2007. Labeled magnitude scales for oral sensations of sweetness, dryness, pleasantness and unpleasantness. ***Food Quality and Preference*** 18: 342-352.

Kamenetzy, J. 1959. Contrast and convergence effect in rating of foods. ***Journal of Applied Psychology*** 43(1)L: 47-52.

Kermadec, F.H. and Pages, J. 2005. Methodology to analyze rank and carry-over effects. ***Food Quality and Preference,*** 16: 600-607.

King, B.M., Arents, P and Duineveld, C.A.A. 2003. A comparison of aspartame and sucrose with respect to carryover effects in yogurt**. *Food Quality and Preference*** 14: 75-81.

Lawless, H.T., and Heymann, H. 1999. ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher. 827p.

Lee, H.J. and Kim, K.O. 2001. Effect of forgetting on various protocols for category and line scales of intensity. ***Journal Sensory Studies*** 16: 327-342.

Lim, J., Wood, A. and Green, B.G. 2009. Derivation and evaluation of a labeled hedonic scale. ***Chemical Senses*** 34: 739-751.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 1999. ***Sensory Evaluation Techniques***. 3rd edition. Florida: CRC Press. 416p.

SAS 2003. ***SAS/IML software: Usage and reference,*** Version 9. SAS Institute Inc., Cary, NC, USA.

Schutz, H.G. and Cardello, A.V. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. ***Journal Sensory Studied.*** 16: 117-159.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. ***Food Quality and Preference***, 16: 691-703.

Walter, F. and Boakes, R.A. 2009. Long-term range effects in hedonic ratings. ***Food Quality and Preference*** 20(6): 440-449.

Warnock, A.R., Shumaker, A.N. and Delwiche, J.F. 2006. Considering of Thurstonian scaling of rating data. ***Food Quality and Preference*** 17: 556-561.

CHAPTER 5 PERFORMANCE TESTING OF THE 9-POINT HEDONIC CATEGORICAL, LINE AND LABELED AFFECTIVE MAGNITUDE SCALES TO ASSESS FOOD LIKING/DISLIKING

5.1 Justification

From our previous studies (chapter 4), contrast and positional effects were strongly presented in some treatments. For example, when product A or B was first presented, a hedonic mean score of subsequent samples was not significantly affected (i.e., insignificant fluctuation). However, when product C was presented first, a score of the subsequent samples was significantly higher than when without C. Several other factors may affect consumer responses. One of the possible factors is a number of categories on the 9-point hedonic categorical scale. Jones and Thurstone (1955) said a 9-point hedonic categorical scale was developed with a neutral middle point to balance out a category that was evenly spaced, but in fact it was unequal psychologically. The scale can be classified as 4 points bipolar scale: 4 points for negative and 4 points for positive. Is this number reasonable? Does the categorical behavior show an impact on score ratings? Do we have an alternative scale that could minimize the categorical behavior and/or ceiling effects? Is the length of the scale appropriate? The use of hedonic scale is so far remained unclear in many technical aspects. We thus conducted a further experiment to determine the effects of the types of scale, lengths or product impression on hedonic ratings. The aim of this chapter was to determine a performance of 9 point hedonic categorical scale and alternative scales with different lengths on hedonic ratings.

5.2 Introduction

The use of an inappropriate protocol may alter the liking scores and prevent experimenters from getting true responses. Lack of an appropriate scale and/or an attribute question causes panelists to find ways to report their irrelevant perception. The consequences contribute to the enhancement effects called "response restrictions or the dumping effects" (Lawless and Heymann, 1999). Proper scales and experimental and protocols to assess food liking/disliking may help increase the discriminating power of a consumer testing experiment.

The hedonic scale was first developed by David Peryam and his colleagues to measure the food preferences and acceptances of soldiers in 1954. Considering among all scales, the hedonic scale is a unique method in terms of general applicability. Three main advantages of using the hedonic scale are (1) simple and can be applied for a wide range of population, (2) not requiring experienced panelists, (3) getting meaningful result and the data can be analyzed by the various parametric statistics (Cordonnier and Delwiche, 2008; Peryam and Pilgrim, 1957). A few years after the 9-point hedonic scale was created, around 1955, Jones and Thurstone developed a balanced 9-point hedonic scale (with a neutral point) that was believed to have an even space physically. However, the highest and lowest scale points were frequently unused and the frequent use of the midpoint remained unclear. Forty-five years later, Schutz and Cardello introduced one of the most popular scales, a Labeled Affective Magnitude (LAM) scale. The LAM was found to be more sensitive with better discriminative power than a 9-point hedonic scale for a well-liked product (Dine and Olabi, 2009; Greene *et al.*, 2006; Shutz and Cardello, 2001). It permits the use of areas above the dislike/like extremely categories (Lawless *et al.*, 2010a). However, the disadvantage was found on the limited use of the scale by performing the categorical behavior on continuous LAM scale (Cardello *et al.*, 2008; Lawless *et al.*, 2010a; Lawless *et al.*, 2010b).

Another alternative scale besides the 9 point hedonic categorical scale and LAM is the 9-point hedonic line scale. It has been selected as an alternative choice because it provides "a zone of psychological comfort" for participants (Lawless and Heymann, 1999), to lessen a categorical behavior, to perform better than the 9-point hedonic categorical scale, to reduce the contextual effect and to provide more freedom (Giovanni and Pangborn, 1983). It also has lower deviation from normality (McPhearson and Randall, 1985), has higher discriminative power than a scale anchored with the best and worst samples (Villanueva *et al.*, 2005), provides a good correlation between ratings of acceptability and preferences (Resano *et al.*, 2009), is more accurate due to no favorite number of scale categories can be made (Meilgaard *et al.*, 2006) and more sensitive than a 9-point hedonic scale for testing an off flavor (Greene *et al.*, 2006). However, Lawless and Malone (1986) mentioned the disadvantageous in terms of time consuming and being user unfriendly.

Scale development has been proposed for many decades and several scales have been developed intending to improve the reliability of methods of measuring the degree of liking. Such scales including the labeled affective magnitude (LAM), labeled magnitude scale (LMS), labeled hedonic scale (LHS), magnitude estimation (ME), unstructured line, self-adjusting, ranking scale, hybrid hedonic, oral pleasantness and unpleasantness (OPUS), and positional relative rating (PRR) (Cardello *et al.*, 2008; Cordonnier and Delwiche, 2008; Dine and Olabi, 2009; Giovanni and Pangborn, 1983; Green *et al.*, 1993, 1996; Guest, *et al.*, 2007; Lim and Green, 2009; McPhearson and Randall, 1985; Peryam and Pilgrim, 1957; Schutz and Cardello, 2001; Warnock *et al.*, 2006; Villanueva *et al*, 2000, 2005, 2009) has been created . However, efforts to replace a unique 9-point hedonic categorical scale have not been completely successful (Lawless *et al.*, 2010b; Stone and Sidel, 2004).

Currently literature reporting the performance of three scales including 9-point hedonic categorical, 9-point hedonic line and LAM scale associated with their scale lengths are limited. Hein *et al.* (2008) mentioned that even thought the LAM has some potential advantages over the

9-point hedonic categorical scale; it is unclear whether it is because of more phases added at the end or the line length.

5.3 Research Specifics

Product: 3 commercial grape juices

Target population: Adults (student, staffs and faculty) age ≥ 18 years.

Sampling: a convenience sampling method without specifying genders, races or ages.

Sample size: 60 consumers

Test locations: LSU campus

Risk involved: minimal risks, except for allergy, were involved in the test. Commercially available food products were used for the research. The identity of the individuals was not revealed. It was impossible to connect the results presented to the subjects who participated in the test.

5.4 Material and Method

5.4.1 Material

The study was conducted at the Louisiana State University (LSU). The consent form was approved by LSU Institutional Review Board before experimentation began. Sixty panelists (40 females and 20 males, age range 20-40) who were familiar with grape juices were recruited from LSU. Three commercial grape juices were used as the test products: Welch's 100% (A), Welch's light (B), and 50% diluted Welch's light (C), which were pre-screened to cover a range of sensory characteristics: transparency, grape flavor, sweetness, and sourness. Hedonic testing and ranking were preliminarily performed to ensure a proper product selection. They were then categorized as good, moderate, and bad, respectively, in terms of quality. Each participant was randomly assigned to one of the four groups (AB and AC; BA and CA; AC and AB; CA and

BA). Each received a different order of sample presentation. The clear lidded plastic cups 60 mL (2.0 oz.) (Propak™ Soufflé clear plastic, Comercializado Por Independent Marketing Alliance, Houston, TX) were used. Each cup was half filled with each brand of juice and labeled with 3-digits blinding codes. For each session samples were kept in a refrigerator at 4^0C until served and trashed after the end of that day. The attribute questions including overall color (OC), overall taste (OT) and overall liking (OL) were included in the questionnaire. To accomplish this task, the sample presentation orders were in a counterbalanced randomized complete block design (RCBD) to ensure that each of the samples would be evaluated in each position an equal number of times. Once the instructions were given, two sets of samples (a total of four) were served. There was a 5 minutes mandatory break in between each set. Panelists made judgments in partitioned booths. They were given a warm-up session for taste instruction emphasizing that they could mark the score in between the phrases for the LAM scale (Lawless and Malone, 1986) and then asked to evaluate samples from left to right. All products were blind tasted; brand names were revealed only after the completion of tasting. To reduce the presentation protocol errors (Cordonnier and Delwiche, 2008), each participant was exposed to all products at the same time. Water and unsalted crackers were served as palate neutralizers during the experiment. Re-tasting products was allowed to refresh memory only if necessary (Lee and Kim, 2001).

5.4.2 Statistical experimental design and analysis

In this experiment, four juices were served. Each participant was given one of the 4 possible random serving orders: AB and AC, BA and CA, AC and AB or CA and BA. In each session the duplicated sample A was served to determine the consistency of the scale. The total of 6 possible permutations (6 sessions) derived from three different scale types (9-point categorical scale, 9-point line scale and LAM scale) and two scale lengths (100 and 300 mm). To see the effect of scale types and scale lengths, all 6 sessions (1-week interval) were performed and compared within the same group of panelists. The standardized value (Z value) was calculated within each scale. The data could be used to compare across all scale types and lengths. Each panelist was assigned one of the four possible random presentation orders, which was repeated in all 6 sessions; however, with different sets of 3-digit blinding codes. Data from panelists who did not complete all 6 sessions were discarded. An average number of panelists were 60.

After tasting, each participant rated 3 attributes of each product including overall color (OC), overall taste (OT) and overall liking (OL) using (1) a 9-point hedonic categorical scale (CAT) and (2) a 9-point hedonic line scale (LIN) where 1 = dislike extremely and 9 = like extremely (Lawless and Heymann, 1999) and (3) a LAM scale (a horizontal line) where 1 referred to greatest imaginable dislike and 100 = greatest imaginable like. The interior phrases and space were created following the published values of Cardello and Schutz (2004). Each scale had two different lengths: 100mm and 300 mm long, in order to determine the effect of lengths on hedonic responses.

5.5 Objectives and their corresponding statistical analysis

The objectives were to determine the scale performance associated with different scale lengths in terms of (5.5.1) the sensitivity, (5.5.2) the reliability, (5.5.3) the consistency when testing with two identical samples, and (5.5.4) the neutral responses' behavior of each scale. Theoretically, the LAM scale tends to have a higher reliability and sensitivity and can minimize the frequent use of categorical rating. The longer line scale may produce a higher hedonic score (Cardello *et al.*, 2008; Dine and Olabi, 2009; Green *et al.*, 2006; Lawless *et al.*, 2010a, b; Schutz and Cardello, 2001). To verify this theory, four objectives were performed.

5.5.1 The ability to differentiate products and/or the sensitivity

Two different methods to measure the sensitivity of an experiment were described below.

5.5.1.1 Discriminative power: The differences in the mean hedonic scores among products could be analyzed by the ANOVA, F test. To compare across all three scale types with different units (1-9 and/or 1-100), the Z value of each scale was calculated to help estimating the differences in order to test the scale effects. If the ANOVA F test was significant, the Tukey's procedure was carried out to calculate the confidence interval (CI) estimated for the mean score for each scale in each treatment (juice). The interaction between scale types and lengths was also determined. The graphic plot of estimated marginal means of each length was provided.

5.5.1.2 Sensitivity: The differences of two lengths for each treatment among three scales were calculated. ANOVA was conducted to investigate the scale sensitivity. If the ANOVA F test was significant, the Tukey's procedure was conducted to calculate the confidence interval (CI) estimated for the mean hedonic score. These estimated CI were then compared to assess the sensitivity affected by the scale types and lengths on the mean hedonic score for each treatment. The variances for attributes of each factor were estimated by Proc Mixed (SAS, 2003).

5.5.2 The reliability of the scale

The reliability in term of the consistency of responses based on the different scale types and lengths among three attribute questions was approached by using the Cronbach's alpha. The higher the Cronbach's alpha value, the higher the reliability. This implied that consumers tended to rate all attribute questions correspondingly.

5.5.3 The consistency of the scale when testing with two identical samples

The Pearson correlation coefficients were used to measure the reliability of the scale when testing with two identical samples. The higher the correlation coefficient, the higher the consistency, meaning that consumers tended rate the same products similarly with minimum level of the variation affected by scale types and lengths.

5.5.4 The neutral responses' behavior

The percentages of the neutral responses which were defined as "5" on CAT were counted. To minimize an unfair count from LIN and LAM scale, the measurement of neural responses included all hash marks that placed between "4.5-5.5" and "45-55" for LIN and LAM scale, respectively (Schutz and Cardello, 2001). The histogram was also used in conjunction with the neutral response counts to determine if such response was meaningful.

5.6. Results and Discussion

5.6.1 The ability to differentiate products and/or sensitivity

The ANOVA F test calculated from Z values was used to approach the discriminative power among treatment and the ANOVA F test based on hedonic score was used to approach the sensitivity of each scale associated with the length effects.

5.6.1.1 Discriminative power

The ANOVA F test statistics and the associated p value are shown in Table 5.1. This F value (Table 5.1) was used for testing the main effects (scales, lengths and interaction of both) calculated based on standardized Z values. Theoretically, it is suggesting that the higher the F value, the better the discriminative power of the scale as mentioned by Hein *et al.*, (2008). However, the interpretation of F value in this study was also associated with the scale length effects. The higher the F values the more susceptible to the length. Length effects are considered as another type of physic psychological biases that could sway the true responses. Therefore, the lower the F value in this case, the higher the performance of scale. For OC and OT, there was neither significant main effect nor significant interaction effect observed. Considering an overall liking attribute, there was an interaction of scale types and lengths which was likely found in good and moderate quality products. Statistically, when an interaction was significant, the conclusion regarding each main effect was meaningless. The graphic plots between scales versus lengths were needed to clarify how the difference behaves.

Table 5.1 Discriminative power of scales based on the Z value

Factors	P value of each sample			
	$A_{(B)}$	B	$A_{(C)}$	C
Overall liking				
Scale	0.99	0.62	0.99	0.99
Length	0.32	0.95	0.00	0.19
Scale*length	0.14	0.17	0.02	0.67
Overall color				
Scale	0.99	0.99	0.99	1.00
Length	0.70	0.93	0.26	0.13
Scale*length	0.46	0.78	0.70	0.11
Overall taste				
Scale	0.71	0.34	1.00	1.00
Length	0.97	0.83	0.47	0.38
Scale*length	0.42	0.75	0.42	0.93

1) $A_{(B)}$ = a good quality sample that paired with a moderate quality sample; $A_{(C)}$ = a good quality sample that paired with a bad quality sample; B = a moderate quality sample and C = a bad quality sample.
2) The results were tested at alpha = 0.05 ($P<0.05$).

Graphic plots between scale types and lengths are shown below (Figure 5.1-5.4) to determine consumer behavior. For CAT (Figure 5.1), the shorter line tended to have a higher score and it was consistent across both good samples and moderated quality juices (Figure 5.1-5.3). However, with LAM, the longer line was likely to have a higher value than the shorter one but the magnitude of the difference was smaller than that of CAT. This is true for both good juices regardless of the pairing sample (Figure 5.1-5.2). Interestingly, both CAT and LAM yielded similar results when testing with a bad juice i.e., no length effect was observed (Figure 5.4). A further exploration will be discussed. With LIN, there was opposite trends for a good juice that was paired with different samples (Figure 5.1 and 5.2). The conclusion could not be drawn at this point; however, the longer LIN line tended to yield a higher value.

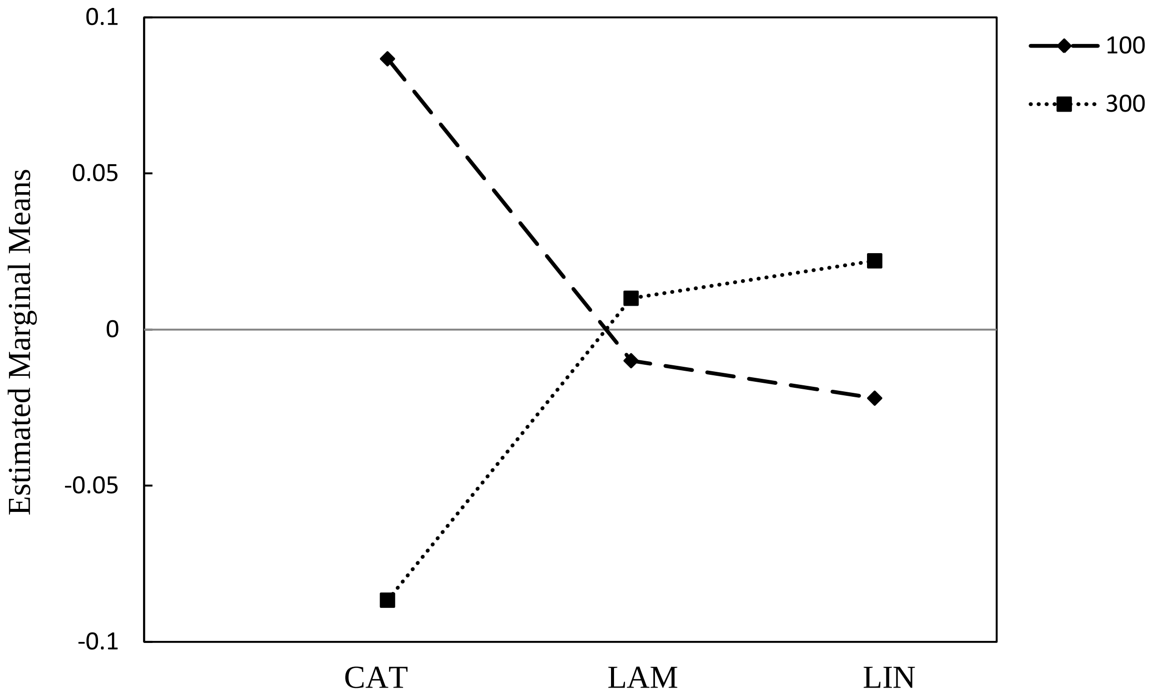

Figure 5.1 A graphic plot of an interaction between scales and lengths for overall liking scores of a good quality juice that was paired with moderate quality juice (A_B) in a counter balance presentation (A_B and $_BA$)

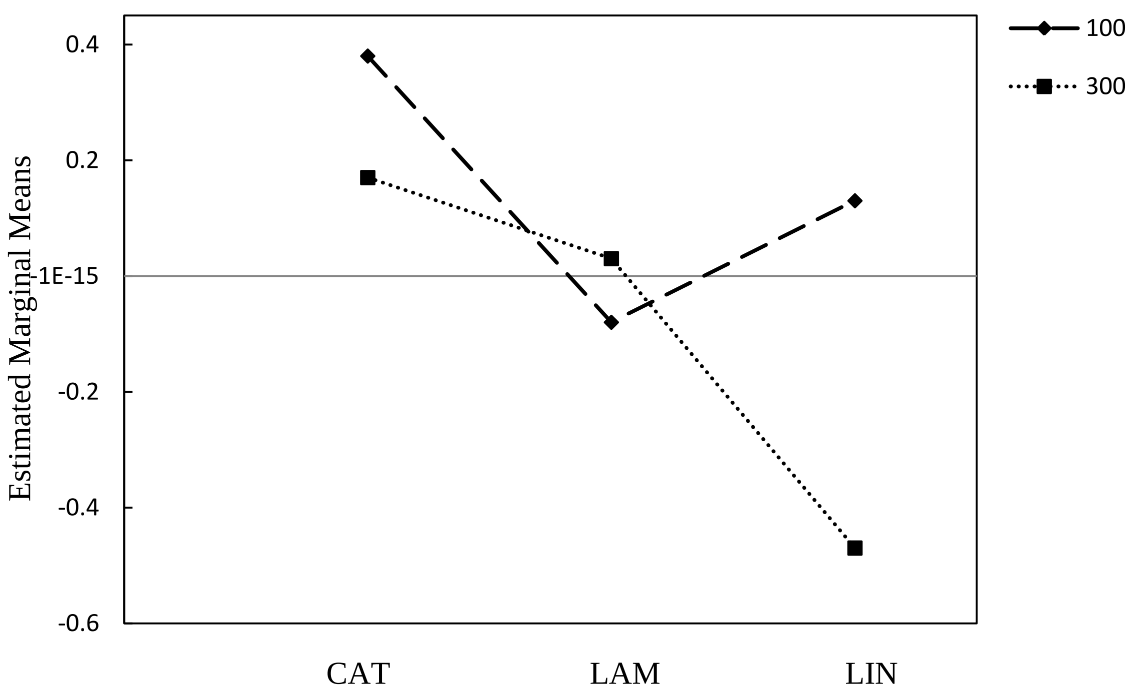

Figure 5.2 A graphic plot of an interaction between scales and lengths for overall liking of a good quality juice that was paired with bad quality juice (A_C) in a counter balance presentation (A_C and $_CA$)

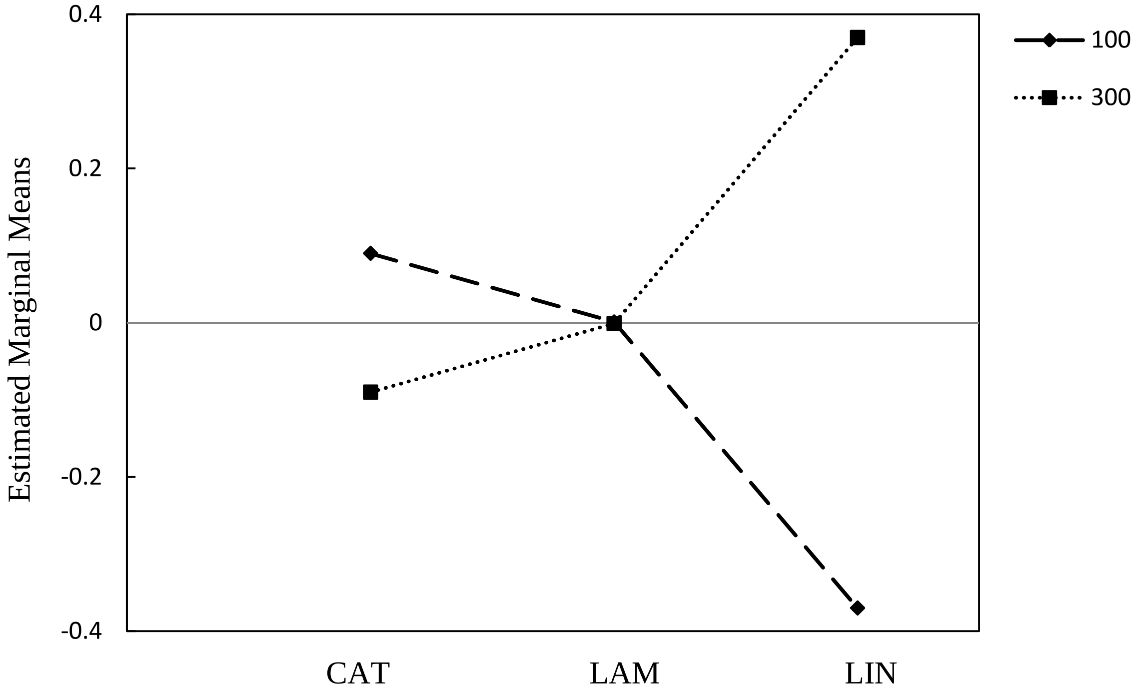

Figure 5.3 A graphic plot of an interaction between scales and lengths for overall liking of a moderate quality juice (B) that was paired with a good quality juice in a counter balance presentation (B_A and $_AB$)

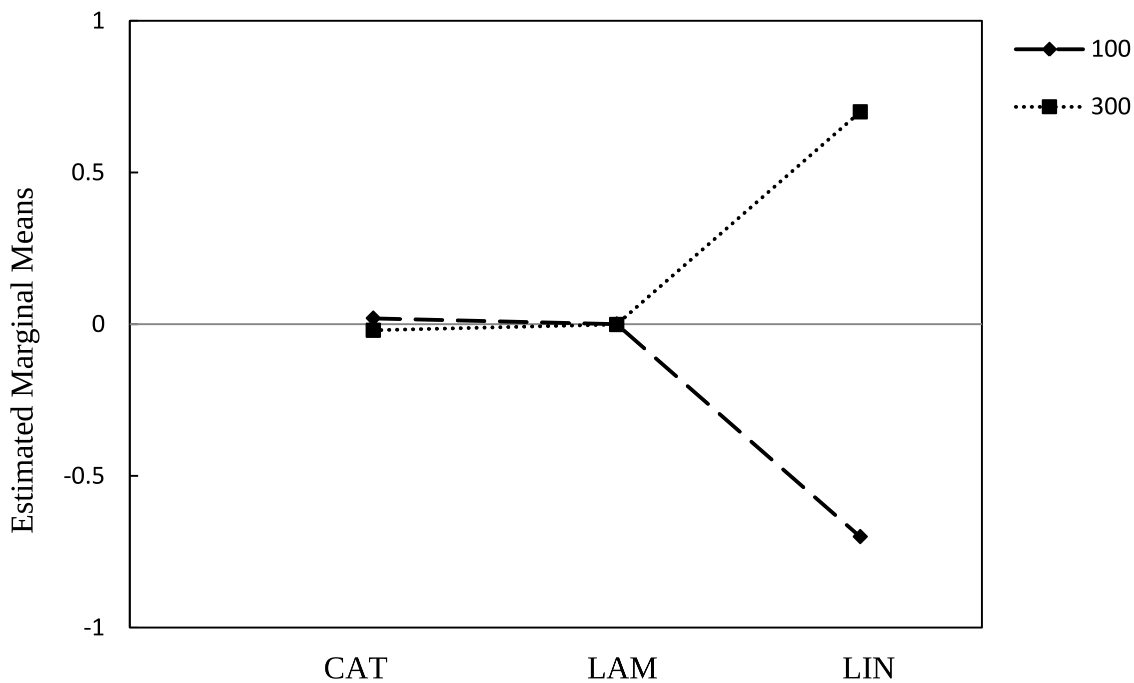

Figure 5.4 A graphic plot of an interaction between scales and lengths for overall liking of a bad quality juice (C) that was paired with a good quality juice in a counter balance presentation (C_A and $_AC$)

When the results were compared across all scale types (Figure 5.1-5.4), LAM seemed to have an advantage over CAT and LIN because it was less sensitive to scale lengths. The results were somehow agreed with Cordonnier and Delwiche (2008) who reported that although the hedonic categorical scale seemed to have better discrimination, the differences on both scales (CAT and LIN) were not obvious. This study found that scale types had no significant effects. The length effects on hedonic variation were slightly observed; the longer scale tended to yield a higher score for LIN but the shorter scale tended to have a higher score for CAT. However, a further exploration will be discussed using the histogram in order to conclude if these differences are meaningful.

5.6.1.2 Sensitivity

This sensitivity is defined as the number of differentiation of a pair of mean (Green *et al.*, 2006; Lawless *et al.*, 2010b). The more the significant number of pairs the higher the discriminative power of the treatment which in turn could help to avoid the type II error in an experiment (Gravetter and Wallnau, 2008; Lawless *et al.*, 2010b).

Table 5.2 Sensitivity of scales

		Overall Liking (OL)			
Scale	Lengths(mm.)	$A_{(B)}$	B	$A_{(C)}$	C
CAT	100	7.74±0.76	6.75±0.90	7.70±0.99	3.65±1.32
	300	7.36±0.97	6.28±1.29	7.12±1.12	3.40±1.53
	P-Value	0.03	0.03	0.04	ns
LIN	100	7.27±1.56	6.23±1.64	7.16±1.58	3.78±1.64
	300	7.35±1.35	7.36±1.35	6.35±1.54	3.40±1.63
	P-Value	0.77	<0.00	0.01	0.02
LAM	100	76.87±15.19	64.46±15.92	76.46±13.78	35.30±19.31
	300	77.15±13.42	63.62±15.78	77.83±15.18	35.05±20.78
	P-Value	0.44	0.38	0.60	ns
		Overall Color (OC)			
Scale	Lengths(mm.)	$A_{(B)}$	B	$A_{(C)}$	C
CAT	100	7.71±0.89	6.03±1.76	7.66±0.99	4.32±1.80
	300	7.51±1.12	5.88±1.64	7.66±0.92	3.61±1.53
	P-Value	0.27	0.82	1.00	0.01

Table 5.2 Continued

Overall Color (OC)					
Scale	Lengths(mm.)	$A_{(B)}$	B	$A_{(C)}$	C
LIN	100	7.63±1.02	5.88±1.47	7.31±1.43	3.87±1.53
	300	7.58±1.01	6.01±1.45	7.49±1.12	4.07±1.80
	P-Value	0.79	0.62	0.44	0.50
LAM	100	77.97±16.57	61.20±16.16	79.12±16.32	41.69±19.18
	300	79.70±10.26	61.65±16.68	81.96±8.64	38.22±15.47
	P-Value	0.49	0.90	0.24	0.33
Overall Taste (OT)					
Scale	Lengths(mm.)	$A_{(B)}$	B	$A_{(C)}$	C
CAT	100	7.46±1.10	6.44±1.36	7.46±1.25	3.54±1.57
	300	7.32±1.27	6.15±1.54	7.12±1.25	3.46±1.70
	P-Value	0.53	0.36	0.14	0.39
LIN	100	7.27±1.41	5.93±1.61	7.00±1.53	3.67±1.52
	300	7.21±1.48	6.34±1.54	7.09±1.29	3.44±1.56
	P-Value	0.81	0.16	0.74	0.42
LAM	100	75.85±16.27	62.80±16.70	75.78±14.41	35.92±20.13
	300	76.33±14.66	63.70±15.30	75.47±13.65	34.38±20.32
	P-Value	0.40	0.90	0.91	0.68

1) P value < 0.05 in the column indicated that at alpha=.05 their corresponding juices differ significantly in their mean scores.
2) "ns" indicated that at alpha=.05, the corresponding juices were not significantly different in their mean scores.
3) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).
4) $A_{(B)}$ = a good quality sample that paired with a moderate quality sample; $A_{(C)}$ = a good quality sample that paired with a bad quality sample; B = a moderate quality sample and C = a bad quality sample.

Mentioned earlier, the sensitivity of the scale length to hedonic score differences was not considered as a good sensitivity. The lower the F value in this case, the higher the performance of scales. Results (Table 5.2) showed that regardless of sample quality, CAT was most susceptible (poor performance) to length effects follow by LIN and LAM. According to the value in the table, overall liking score of LIN showed a significant difference for good quality that paired with bad quality but found no significance in good quality that paired with moderate quality. The result was inconclusive whether the variation came from scale lengths or product impression. The further histogram chart will be used to clarify this phenomenon. A total of 12 pairs (4 products * 3 attribute questions) were evaluated for significant difference. There were four significant pairs of difference from CAT, whereas three and zero pairs of differences for LIN and LAM scale, respectively. Again, CAT was most susceptible to the length effects yielding 4 pairs out of 12 pairs accounting for 33.33% comparing to that of for LIN (25%) and LAM (0%). Thus far regarding the discriminative power and the sensitivity of the scale, LAM seemed to have an advantage over CAT and LIN scale. Also LAM and CAT had showed less variation from normality comparing to LIN (Figure 5.5). The graph showed the distribution of overall liking scores from three scale types and four juices (treatments) with 100 mm length. Both 100 and 300 mm scale lengths (not shown) yielded similar distribution. Based on the results, by comparing across scale types, the distribution of hedonic scores from both CAT and LAM were in a good bell shape with lower standard deviation compared to that from LIN. This was true for good and moderate quality juices. The implication from a previous section discussing about the sensitivity to the scale length was meaningful as it was also evidenced from this histogram. Thus, one should be aware of length effects when testing good and moderate samples quality using CAT as well as when using LIN to test a good quality juice that was paired with a bad quality. The score tended to be higher for the shorter scale. The paradox of LIN results (Figure 5.1 and 5.2) may be due to the length effects in conjunction with scale complexity (Lawless and Malone, 1986). The instruction of how to response on the scale should be more clearly given especially for an online survey where the width of the computer screen (i.e., different scale lengths) can influence the results. The scale length should be consistent across an entire experiment in order to minimize length effects that could sway an accuracy of an experiment.

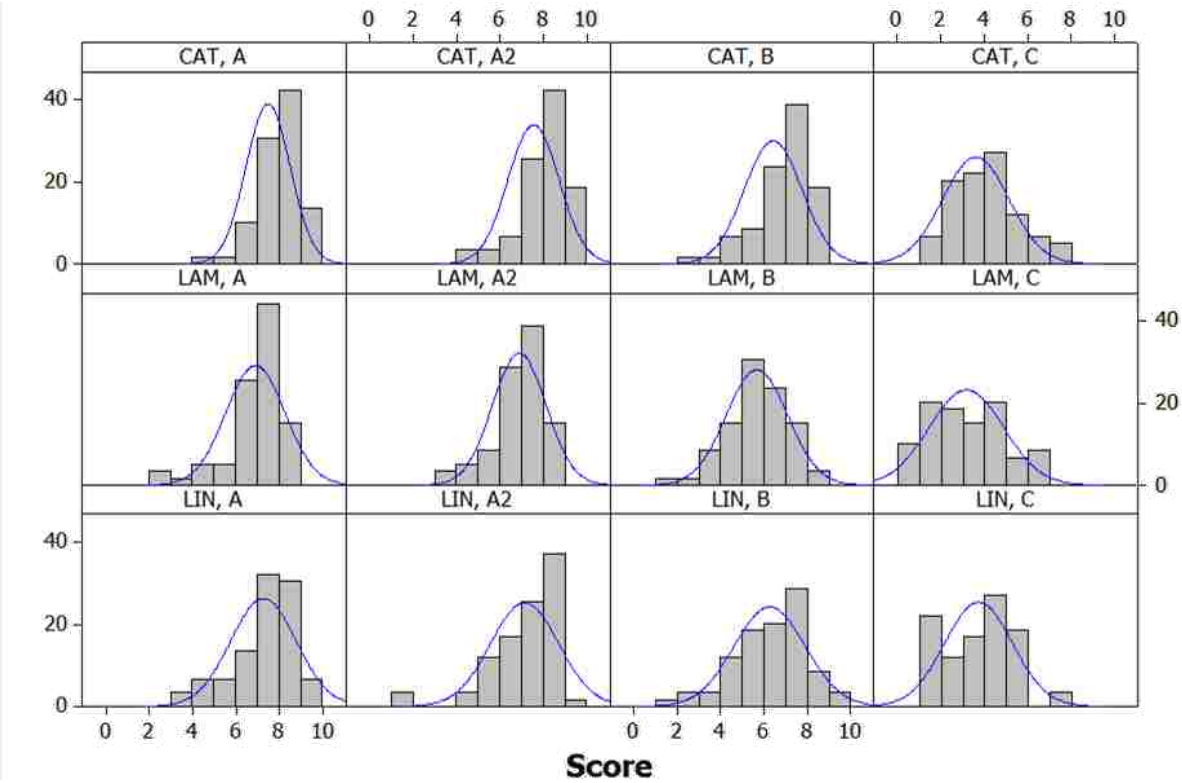

Figure 5.5 Histogram of Overall liking of grape juices for a good quality juice paired with a moderate quality juice (A), good quality juice paired with a bad quality (A2), moderate quality juice (B) and a bad quality juice(C) for each scale (CAT, LAM and LIN) with 100 mm. length.

1) $A_{(B)}$ = a good quality sample that paired with a moderate quality sample; $A_{(C)}$ = a good quality sample that paired with a bad quality sample; B = a moderate quality sample and C = a bad quality sample.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

However, when testing a bad juice, all scales yielded similar distribution. The standard deviation was higher making the bell shape flatter which may result in an insignificant difference. Lawless *et al.* (2010b) observed the higher variation in LAM scale as well but the cause still remained unclear. The result was scattered and it was inconclusive to draw a conclusion.

5.6.2 The reliability of scales

Taking into account of the scale reliability, this experiment tested the reliability in two difference ways which were (1) the reliability in terms of the responses' consistency based on the scale types and lengths among three attribute questions (OL, OC and OT) approached by the Cronbach's alpha value (Table 5.3) and (2) the reliability in terms of the consistency of responses from the two identical samples approached by the Pearson's correlation coefficient (Table 5.4).

Table 5.3 Reliability test for three scale types and two scale lengths

Lengths (mm.)	Cronbach's alpha				
	Scale	$A_{(B)}$	B	$A_{(C)}$	C
100	CAT	0.879	0.827	0.845	0.888
	LIN	0.777	0.833	0.855	0.909
	LAM	0.859	0.843	0.816	0.896
300	CAT	0.853	0.892	0.849	0.914
	LIN	0.785	0.856	0.850	0.891
	LAM	0.880	0.898	0.781	0.919

1) $A_{(B)}$ = a good quality sample that paired with a moderate quality sample; $A_{(C)}$ = a good quality sample that paired with a bad quality sample; B = a moderate quality sample and C = a bad quality sample.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

Based on the results (Table 5.3), the Cronbach's alpha values of all three scale types with two lengths were similar. Consumers tended to rate OL, OC and OT question consistently regardless of sample quality. There was less or no conflict of interest within same products. The variation from scale types did not impact hedonic ratings. Regarding overall liking question, values were ranging from 0.777-0.919 which implied a consistent scoring on all scales. There was no specific pattern could be observed between two lengths. However, there was somewhat higher (0.891-0.919) Cronbach's alpha values for a bad quality sample. The higher correlation among the two attribute questions and the overall liking score for a disliked product implied that consumers tended to rate a disliked product more consistently than the well-liked and/or moderate-liked sample.

5.6.3 The consistency of the scale when testing with two identical samples

The correlation coefficient between the two identical samples presented within the same test was used to determine the consistency of ratings (Table 5.4). This reliability was defined as the consistency of the responses based on scale types and lengths from two identical tested samples (Cardelllo and Maller (1982) and Lawless *et al.* (2010b)). The higher the Pearson's correlation coefficient of identical samples testing, the higher the reliability of the scale (Lawless *et al.*, 2010b).

The Pearson coefficient was carried out to determine which of these scales and lengths were strongly correlated. The length effects under this experiment were observed with an unclear explanation. The coefficient values were ranging from 0.32-0.75 which was relatively low. Wannita *et al.* (2011) reported that there is much variation within each consumer. Even for the same sample served, approximately 75% of consumers changed their responses.

Table 5.4 Pearson's correlation coefficients between the two identical samples

Scale	Lengths (mm.)	OL	OC	OT
CAT	100	0.504*	0.569*	0.745*
	300	0.424*	0.655*	0.580*
LIN	100	0.485*	0.398*	0.580*
	300	0.377*	0.742*	0.323**
LAM	100	0.451*	0.747*	0.390*
	300	0.558*	0.694*	0.527*

* showed the significant difference at alpha level = 0.01
** showed the significant difference at alpha level = 0.05
CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

By comparing across all scales, the data showed trivial differences. For LIN scale, it yielded the lowest correlation coefficient when disregarding the length effects (0.43-0.57). The LAM scale seemed to have a slightly higher coefficient (0.46-0.72) than CAT and LIN scale whereas the CAT scale held a moderate value between 0.46-0.66. Our results fall in the same range observed by Lawless *et al.* (2010b) for the category scale (+0.62); however, it was dissimilar for the LAM scale (+0.34). Nevertheless, Lawless *et al.* (2010a) discovered differently; the correlation coefficients for CAT and LAM were +0.52 and +0.52, respectively, and the results were agreed with Schutz and Cardello (2001). This study also agreed that the

LAM scale was slightly better than a categorical scale in terms of reliability of testing two identical products.

5.6.4 The neutral responses' behavior

The percentages of the neutral responses are shown (Table 5.5). The sample quality showed an immense influence on neutral response counts. Theoretically, the better the scale performance in achieving hedonic responses, the lower the percentage of neutral tendency.

Table 5.5 Neutral tendency percentage

Overall Liking (OL)					
Lengths(mm.)	Scale	$A_{(B)}$	B	$A_{(C)}$	C
100	CAT	1.69	8.47	3.39	11.86
	LIN	6.78	8.47	11.86	23.73
	LAM	1.69	15.25	5.08	18.64
300	CAT	1.69	11.86	6.78	15.25
	LIN	8.47	20.34	11.86	23.73
	LAM	10.17	16.95	1.69	16.95
Overall Color (OC)					
Lengths(mm.)	Scale	$A_{(B)}$	B	$A_{(C)}$	C
100	CAT	3.39	5.08	1.69	5.08
	LIN	6.78	22.03	8.47	27.12
	LAM	1.69	11.86	3.39	25.42
300	CAT	1.69	10.17	1.69	13.56
	LIN	5.08	25.42	3.39	18.64
	LAM	1.69	18.64	0.00	22.03
Overall Taste (OT)					
Lengths(mm.)	Scale	$A_{(B)}$	B	$A_{(C)}$	C
100	CAT	0.00	8.47	0.00	13.56
	LIN	8.47	20.34	13.56	20.34
	LAM	1.69	13.56	6.78	27.12
300	CAT	1.69	11.86	3.39	5.08
	LIN	6.78	18.64	16.95	16.95
	LAM	6.78	16.95	10.17	22.03

1) $A_{(B)}$ = a good quality sample that paired with a moderate quality sample; $A_{(C)}$ = a good quality sample that paired with a bad quality sample; B = a moderate quality sample and C = a bad quality sample.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

Consumers seemed to have a similar frame of acceptance over grape juice quality. The good quality sample was rated higher with a lower percentage of neutral responses. However, the moderate and bad quality product had a higher count of neutral response. Several reasons could be used to explain this behavior. Some consumers may choose the mid point as a safe space to express a perception (dumping effect) (Lawless and Heymann, 1999). Consumers weren't sure about how to think or response to the product. The selected attribute question may not be a good indicator for this product category or the complexity of the scale could drive consumers' confusion. Regarding Table 5.5, the results based on attribute questions were inconclusive. The neutral counts tended to be higher for the overall taste question than overall color but there was a contradiction for a number of cases. The scale lengths showed a minor influence while the scale types showed more effects. There was no specific pattern could be observed based on the length effects; however, the longer lengths tended to have a higher count of the mid scale responses.

When rating a bad quality sample, it tended to have a higher neutral response count all scale types. Interestingly, the hedonic CAT scale yielded the lowest amount of neutral responses (1.69-10.17) following by LIN scale (7.63-14.41) and LAM scale (4.66-16.10) when the data were average among the product, ignoring the length effects for each attribute. One of the possible reasons is that CAT is a categorical scale whereas LIN and LAM is a continuous line scale. The complexity of scale could lead to a higher count of neutral responses. Based on this result, CAT seemed to be a good choice to assess a degree of liking/disliking with a lower rate of neutral response.

Considering a discrimination power, sensitivity, reliability and neutral tendency, CAT and LAM seemed to be more superior to LIN for assessing a degree of food liking/disliking. However, CAT was more susceptible to length effects while LAM was not. In addition to the advantages of using LAM over the 9- point hedonic scale, it could help minimize a ceiling effect when using CAT by giving more flexibility of permitted space above the "like extremely" category and below the “dislike extremely” category. This experiment observed the frequency of usage these areas. It was found that to evaluate the well-liked sample, panelist tended to rate more often in the area above like extremely. Consumers (28.82%) rated more frequently above the like extremely category on a longer line scale than on a shorter scale (17.80%) for well-liked sample. In contrast, the moderately-liked sample had a lower count approximately (3.39-5.08%) and was not affected by of scale length. For testing the disliked product, the usage of area below “dislike extremely” category was observed roughly at 15.25% and was not affected by scale lengths. This result corresponded to those reported by Cardello *et al.* (2008), Lawless *et al.* (2010a) and Schutz and Cardello (2001); the values were 19-30%, 10-20% and 17% for the usage of area above the like extremely category. Therefore, the LAM scale may be beneficial from these permitted areas when testing well-liked or disliked products.

5.7 Conclusion

LAM had a high discriminative power, sensitivity and reliability with a higher Cronbach’s alpha value and a higher Pearson’s correlation coefficient. The benefit from LAM is observed when testing with well-liked and disliked products. The longer scale length was more beneficial as consumers rated more often in the area above/below like/dislike extremely category. The hedonic CAT scale was sensitive to length effects; however, it yielded high reliability, and high Pearson’s correlation with the lowest neutral tendency rate. The hedonic LIN scale had a moderated discriminative power, sensitivity, reliability, Pearson correlation’s but yielded the highest rate of neutral response counts. Both lengths yielded similar reliability Cronbach’s alpha values which implied no superiority over each other. Based on this study, it is suggested that a regular length (100 mm.) be used as a standard for hedonic testing and LAM as an alternative choice to a 9-point categorical scale.

5.8 References

Cardello, A., Lawless, H.T. and Schutz, H.G. 2008. Effect of extreme anchors and interior Label spacing on labeled affective magnitude scales. ***Food Quality and Preference*** 19: 473-480.

Cardello, A.V. and Maller, O. 1982. Relationships between food preferences and food acceptance ratings. ***Journal of Food Science*** 47: 1553-1557.

Cardello, A.V. and Schutz, H.G. 2004. Research Note Numerical scale-point locations for constructing the LAM (Labeled affective magnitude) scale. ***Journal Sensory Studies.*** 19: 341-346.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Dine, A.N. and Olabi, A. 2009. Effect of reference foods in repeated acceptability tests: Testing familiar and novel foods using 2 acceptability scales. ***Journal of Food Science.*** 74: S97-S105.

Giovanni, M.E. and Pangborn, R.M. 1983. Measurement of taste intensity and degree of liking of beverages by graphic scaling and magnitude estimation. ***Journal of Food Science.*** 48: 1175-1182.

Gravetter, F.J. and Wallnau, L.B. 2008. ***Statistics for the behavioral sciences***. 8th. Wadworth, Cengage Learning, Belmont, CA.

Green, B.G., Dalton, P., Cowart, B., Shaffer, G.S., Rankin, K. and Higgings, J. 1996. Evaluating the "labeled magnitude scale" for measuring sensations of taste and smell. ***Chemical Sense.*** 21: 323-334.

Green, B.G., Shaffer, G.S. and Gilmore, M.M. 1993. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. ***Chemical Sense.*** 18: 683-702.

Greene, J.L., Bratka, K.J., Drake, M.A. and Sanders, T.H. 2006. Effective of category and line scales to characterize consumer perception of fruit fermented flavors in peanuts. ***Journal Sensory Studies*** 21: 146-154.

Guest, S., Essick, G., Patel, A., Prajapati, R. and MaGlone, F. 2007. Labeled magnitude scales for oral sensations of sweetness, dryness, pleasantness and unpleasantness. ***Food Quality and Preference*** 18: 342-352.

Hein, K.A., Jaeger, S.R., Carr, B.T. and Delahunty, C.M. 2008. Comparison of five common acceptance and preference methods. ***Food Quality and Preference*** 19: 651-661.

Jones, L.V. and Thurstone, L.L. 1955. The psychophysics of semantics: An experimental investigation. ***Journal Applied Psychophysic*** 39: 31-36.

Lawless, H.T. and Malone, G.J. 1986. A comparison of scaling methods: Sensitivity, replicates and relative measurement. ***Journal Sensory Studies*** 1: 155-174.

Lawless, H.T., and Heymann, H. 1999. ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher. 827p.

Lawless, H.T., Popper, R. and Kroll, B.J. 2010a. Comparison of the labeled affective magnitude (LAM) scale, an 11-point category scale and the traditional nine-point hedonic scale. ***Food Quality and Preference*** 21: 4-12.

Lawless, H.T., Sinopoli, D. and Chapman, K.W. 2010b. A comparison of the labeled affective magnitude scale and the 9-point hedonic scale and examination of categorical behavior. ***Journal Sensory Studies*** 25: 54-66.

Lee, H.J. and Kim, K.O. 2001. Effect of forgetting on various protocols for category and line scales of intensity. ***Journal Sensory Studies*** 16: 327-342.

Lim, J., Wood, A. and Green, B.G. 2009. Derivation and evaluation of a labeled hedonic scale. ***Chemical Senses*** 34: 739-751.

McPhearson, R.S. and Randall, D.E. 1985. Line length measurement as a tool for food preference research. ***Ecology of Food and Nutrition*** 17(2): 149-156.

Meilgaard, M.C., Civille, G.V. and Carr, B.T. 2006. ***Sensory Evaluation Techniques***. 4th edition. Florida: CRC Press. 448p.

Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preference. ***Food Technology*** *(September)*, 9-14

Resano, H., Sanjuan, A.I. and Alibisu, L.M. 2009. Consumer acceptability and actual choice. An exploratory research on cured ham in Spain. ***Food Quality and Preference*** 20: 391-398.

SAS 2003. ***SAS/IML software: Usage and reference***, Version 9. SAS Institute Inc., Cary, NC, USA.

Schutz, H.G. & Cardello, A.V. 2001. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. ***Journal Sensory Studied.*** 16: 117-159.

Stone, H. and Sidel, J. 2004. ***Sensory Evaluation Practices***. 3rd edition. California: Academic Press.

Villanueva, N.D.M., Peternate, G.J. & Da Silva, M.A.A.P. 2000. Performance of three affective methods and diagnosis of the ANOVA model. ***Food Quality and Preference*** 11: 363-370.

Villanueva, N.D.M., Petenate, A.J., and Da Silva, M.A.A.P. 2005. Performance of the hybrid hedonic scale as compared to the traditional hedonic, self-adjusting and ranking scales. ***Food Quality and Preference***, 16: 691-703.

Villanueva, N.D.M. and Da Silva, M.A.A.P. 2009. Performance of the nine-point hedonic, hybrid and self-adjusting scales in the generation of internal preference maps. ***Food Quality and Preference***, 20: 1-12.

Jirangrat, W., Wang, J, and Prinyawiwatkul, W. 2011. Quantifying contrast effect in consumer acceptance tests using the logistic regression analysis. IFT Oral competition: Sensory and Consumer Science division. ***IFT Annual Meeting*** in New Orleans, LA from June 11 – June 15, 2011.

Warnock, A.R., Shumaker, A.N. and Delwiche, J.F. 2006. Considering of Thurstonian scaling of rating data. ***Food Quality and Preference*** 17: 556-561.

CHAPTER 6 VARIATIONS IN HEDONIC RATINGS CHARACTERIZED BY SCALE POLARITY/TYPES AND ATTRIBUTE QUESTIONS

6.1 Justification

According to chapters 3-5, several selected factors have been proven to cause variations in hedonic ratings. The confounding effects caused by the presentation order, contrast effects and different scale types that were observed previously could prevent experimenters from getting reliable sensory outcomes. However, the author doubts that would these effects be pronounced for all attribute questions? Which type of hedonic scales is impacted the most? Is it true for both unipolar and bipolar hedonic scales? If this can be answered properly, it may help experimenters to interpret the results better. This, in turn, can help sensory scientists to plan an experiment more appropriately. In this chapter, this experiment was done to test if these factors (attribute questions and scale polarity) are a major concern when using CAT, LIN and LAM scales.

Sensory attribute questions should be properly selected when testing a product. If an experimenter chooses an irrelevant attribute to be included in a questionnaire, the result will not be meaningful. No direction for product improvement will be gained based on an irrelevant attribute. Selecting proper attributes for a product testing will help avoid a dumping effect. There are two different types of attribute questions, i.e., positive and negative attribute questions. In this study, overall liking is a positive attribute. In contrast, bitterness perception of products is a negative attribute (Bartoshuk, 1979).

The question is why we are interested in bitterness? Can consumer perform similarly on both negative and positive attribute questions? Will the score be the same for both scale polarity? At this point, we have no published data for these questions. There are limited numbers of publication associated with negative attribute on hedonic scale performance; however, Lawless and Heymann (1999) said "The negative side of hedonic scaling is not as fully differentiated as the positive side" resulting in losing the discriminative power. More research is needed in this area to verify the sensory assumptions.

6.2 Introduction

The hedonic scale was first developed by David Peryam and his colleagues to measure the food preferences of soldiers (Peryam, 1954). Considering all scales and procedures, the hedonic scales occupies a unique methodology in terms of general applicability (Cordonnier and Delwiche, 2008; Peryam and Pilgrim, 1957). However, the usage of the traditional 9-point hedonic scale has been questioned regarding its accuracy and validity when testing negative sensory attributes. Lawless and Heymann (1999) mentioned that consumers are likely to rate opinion about what they like more than dislike so that "The negative side of hedonic scaling is not as fully differentiated as the positive side."

Several scales have been applied for product testing associated with positive attributes such as overall liking of the product. Some attributes have been asked in conjunction with overall liking in order to establish a relationship among them that can be used for product improvement. The more we get to know the product, the better chance to deliver acceptable products. Scientists have been struggling to find an appropriate scale and protocol to be used. Several problems such as the end-use avoidance (a tendency to omit or avoid using the extreme or end-categories in order to spare an extreme response for further samples) have been an issue leading to a new scale development. Because of the categorical behavior nature of consumers when performing in consumer testing, several scales have been developed to serve this purpose. For example, LAM has been developed and it was concluded to be more sensitive with better discrimination power than CAT (Dine and Olabi, 2009; Greene *et al.*, 2006). It may help reduce categorical behavior responses comparing to CAT.

Several trials have been conducted to test an application of the traditional 9 point hedonic scale for assessing food liking/disliking. Unfortunately, no study has been reported to evaluate a scale performance as affected by attribute question (negative versus positive attributes) and scale polarity (unipolar versus bipolar scales). In this study, the three widely used scales (labeled affective magnitude scale (LAM), traditional 9-point categorical hedonic (CAT), and 9-point continuous hedonic line scale (LIN)0 were used to determine the effects of attribute question and scale polarity.

6.3 Research specifics

Product: chicken soup containing Potassium Chloride (KCL Mortan® Salt Substitute for salt-free diet, Chicago, IILINOIS) and Sodium Chloride (NaCL Mortan® Salt Substitute, Chicago, IILINOIS) and commercial grape juices.

Target population: Adults (student, staffs and faculty) age $\geq$ 18 years.

Sampling: a convenience sampling method without specifying genders, races or ages.

Sample size: 216 consumers

Test locations: LSU campus

Risk involved: minimal risks, except for allergy, are involved in the test. Commercially available food products will be used for the research. The identity of the individuals was not revealed. It will be impossible to connect the results presented to the subjects who participated in the investigation.

6.4 Material and Method

6.4.1 Material

The study was conducted at the Louisiana State University (LSU). All procedures and the consent forms were approved by LSU Institutional Review Board before the experiment was started. Two sub-experiments were conducted separately; one for testing the positive attributes and the other for testing the negative attributes. To accomplish this task, the sample presentation orders followed a counterbalanced randomized complete block design (RCBD) to ensure that each of the samples would be evaluated in each position an equal number of times.

For testing positive attributes, sixty panelists who were familiar with grape juices were recruited from LSU (students, staffs and faculty). Each participant was randomly assigned to one of the four groups: AB and AC, BA and CA, AC and AB or CA and BA. Each received a different order of sample presentation. Three commercial grape juices were evaluated: Welch's 100% (A), Welch's light (B), and 50% diluted Welch's light (C), which were pre-screened to cover a range of sensory characteristics: transparency, grape flavor, sweetness, and sourness. There were categorized as good, moderate and bad quality juices, respectively. Clear lidded plastic cups 60 mL (2.0 oz.) (Propak™ Soufflé clear plastic, Comercializado Por Independent Marketing Alliance, Houston, TX) were used. Each cup was half filled with each brand of juice and labeled with 3-digits blinding codes. During the test, samples were kept in a refrigerator at 4^{0}C until served and trashed after the end of that day. The attribute questions including overall color (OC), overall taste (OT) and overall liking (OL) were included in the questionnaire.

For testing negative attributes, two hundred and sixteen panelists who are familiar with chicken broths were recruited from LSU (students, staffs and faculty). Each participant was randomly assigned to one of the twelve groups derived from different orders of sample presentation, scale types and scale polarity shown in Appendix D). Two chicken soups were evaluated: one with a high level of salt substitution (Potassium Chloride: KCL Mortan® Salt Substitute for salt-free diet, Chicago, IL) at 2% by weight classified as strong bitterness broths (S) and the other with a mixture of a regular table salt (Sodium Chloride: NaCl Mortan® Salt

Substitute, Chicago, IL) and a low level of salt substitution (Potassium Chloride: KCL Mortan® Salt Substitute for salt-free diet, Chicago, IL) at the ratio of NaCl: KCl (2:1) at 1.3 % by weight classified as mild bitterness broths (M). The formulation, process, codes and questionnaire are shown in the appendix C and D. Clear lidded plastic cups 60 mL (2.0 oz.) (Propak™ Soufflé clear plastic, Comercializado Por Independent Marketing Alliance, Houston, TX) were used. Each cup was half filled with a formulation of chicken soup and labeled with 3-digits blinding codes. During the test, samples were kept warm at 50^{0}C until served and trashed after three hours. The overall bitterness perception was included in the questionnaire.

Once the instructions had been given, two sets of samples were served. There were 5 minutes mandatory break in between the set of samples. Panelists made judgments in partitioned booths. They were informed that they could mark the score in between the phrases for LAM scale (Lawless and Malone, 1986) and then asked to evaluate samples from left to right. All products were blind tasted; brand names were revealed only after the tasting had been completed. To reduce the presentation protocol errors (Cordonnier and Delwiche, 2008), each participant was served with both products at the same time. Water and unsalted crackers were served as neutralizers during the experiment. Re-tasting products was allowed to refresh memory if necessary (Lee and Kim, 2001).

6.4.2 Statistical experimental design and analysis

For positive attribute testing: each panelist was served with one of the four possible random serving orders: AB and AC, BA and CA, AC and AB or CA and BA. Panelists were asked to participate in all three sessions to complete evaluation of three different scale types (CAT, LIN and LAM). Each panelist received the same order of juice presentation in all 3 sessions, however, with different sets of 3-digit blinding codes. Data from panelists who did not complete all 3 sessions were discarded, giving a final sample size, N = 59 panelists participated in this study.

For negative attribute testing: one of the two possible random serving orders: MS and SM was served to each panelist. Two sets of questionnaire: one for bi-polar scales and the other for uni-polar scales were included during the test. Twelve sets of questionnaire were derived from three scale types (CAT, LIN, LAM; CAT LAM, LIN; LIN, CAT, LAM; LIN, LAM, CAT; LAM, CAT, LIN and LAM, LIN, CAT) and two permutation of serving order (SM or MS). Each received the same presentation order in both sets of questionnaire (unipolar and bipolar scales), however, with different sets of 3-digit blinding codes. Data from panelists who did not complete all two sessions were discarded.

After tasting grape juices, each participant rated 3 attributes of each product, including overall color (OC), overall taste (OT) and overall liking (OL). After tasting chicken broths, each participant rated the bitterness perception of broths. The rating scores were on (1) a 9-point hedonic categorical scale (CAT) and (2) a 9-point hedonic line scale (LIN) where 1 = dislike extremely and 9 = like extremely (Lawless and Heymann, 1999) and (3) a labeled affective magnitude scale (LAM) (a horizontal line) where 1 = greatest imaginable dislike and 100 refers greatest imaginable like. The interior phases and space were created following the published values of Cardello and Schutz (2004).

6.5 Objectives and their corresponding statistical analysis

The aims of this study were to investigate an impact of negative (N) versus positive (P) product attributes on three different scale types [categorical (CAT), line (LIN) and labeled affective magnitude (LAM) scale] and two different scale polarity [uni-polar (U) and bi-polar (B)] on degree of liking and/or disliking. Two sub-objectives were:

6.5.1 An effect of attribute questions

The objective was to determine effects of negative question (N) versus positive question ratings (P) obtained from three above mentioned scales on the sensitivity as related to confounding effects (contrast and panelist effects (CP)).

6.5.2 An effect of scale polarity

The objective was to compared effects of bi-polar scale (B) versus uni-polar scale (U) (only a negative side and negative attribute ratings) on the sensitivity as related to confounding effects [contrast and panelist effects (CP)].

For both objectives, ANOVA was carried out (SAS, 2003). If the ANOVA F test was significant, the Tukey's procedure was further conducted to calculate the confidence interval (CI) estimated for the mean hedonic score for each treatment. These estimated CI values were then compared to assess the sensitivity as affected by the scale types and polarity. The variances for attributes of each factor were estimated using a Proc Mixed procedure.

6.6 Results and Discussion

6.6.1 An effect of attribute questions

Table 6.1 presents the Coefficient of determination from each scale. The Coefficient of determination (R^2) can be calculated from the ANOVA F test statistics table (not shown). First, sum up all mean square (MS) values and use it as a denominator [Denominator = ∑MS (Contrast+Treatment+Panel+Residual)]. The R^2 of each factor was then calculated from a ratio of each mean square factor and denominator e.g., [R^2 of Treatment = MS (Treatment) *100/ Denominator] (below example). Theoretically, the higher the coefficient of determination (R^2), the higher the discriminative power (Hein *et al.*, 2008), implying the higher explained variance accounted for by such factors.

Table 6.1 Coefficient of determination (R^2) calculated for each factor

Attributes	Scale	R^2			
		Treatment	Contrast	Panel	Residual
Positive (A_B) [Bi-polar]	CAT	98.75	0.43	0.11	0.71
	LIN	77.71	12.43	3.78	6.09
	LAM	2.56	19.16	72.57	5.70
Positive (A_C) [Bi-polar]	CAT	99.14	0.37	0.10	0.40
	LIN	98.75	0.43	0.11	0.71
	LAM	98.01	1.34	0.10	0.55
Negative [Bi-polar]	CAT	99.61*	0.03	0.09	0.30
	LIN	99.51	0.08	0.01	0.39
	LAM	99.36	0.00	0.23	0.40
Negative [Uni-polar]	CAT	98.86	0.31	0.38	0.45
	LIN	97.29	0.55	0.50	1.66
	LAM	88.60	0.73	10.09	0.58

1) $A_{(B)}$ = a well-liked sample that paired with a moderately-liked sample; $A_{(C)}$ = a well-liked sample that paired with a disliked sample.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

*Example: To calculate a coefficient of determination from ANOVA table shown below:

Source	DF	SS	MS	F	Pr>F
Treatment	1	793.500	793.500	335.36	<0.0001
Contrast	1	0.019	0.019	0.01	0.9296
Panel	1	0.695	0.695	0.29	0.5883
Residual	212	501.620	2.366		

Denominator = ∑MS (Contrast+Treatment+Panel+Residual)
= (0.019+793.5+0.695+2.366)

R^2 of Treatment = MS (Treatment)/ Denominator*100
= 793.5/796.58]*100
= 99.61

According to Table 6.1, the treatment effect showed the highest R^2 value which implied that the variation of hedonic ratings came mostly from the product impression (good, moderate and bad quality) rather than the biases (contrast, panelist and unexplained variance factors). For a positive attribute (a bi-polar scale; overall liking) of grape juices, CAT yielded the highest R^2 value which implied the highest ability to differentiate among products. This finding agreed with Hein *et al.* (2008) and Lawless and Malone (1986) who observed that the 9-point hedonic categorical scale exhibited the highest sensitivity. Comparing across all three scales, CAT ranked first followed by LIN and LAM for both A_B and A_C combinations. However, it can be seen that

LAM showed a very low sensitivity in this experiment for a small magnitude of difference (similar products; A_B).

For a negative attribute (a bi-polar scale; bitterness), although the negative side of the scale is not as fully differentiated as the positive side (Lawless and Heymann, 1999), this result showed higher R^2 values for all bipolar scales: (positive vs. negative: CAT (98.75 vs. 99.61), LIN (77.71 vs. 99.51) and LAM (2.56 vs. 99.36), respectively. The individual preference is varied for a positive attribute (or liked products) whereas when it comes to a negative attribute (or disliked products), consumers have similar frame of preferences. This finding confirmed the notion that individuals differed in their perception to suprathreshold of bitterness (Horne *et al.*, 2002) and hedonic responses could not be used to differentiate products at a high level of bitterness (Drewnowski *et al.*, 1997).

Even though this experiment was conducted by a balance randomized design, the extraneous error theoretically was canceled out. From Table 6.1, LAM was affected by the confounding effects (Contrast and Panelist: CP) the most; however, these effects were less pronounced when scale was applied with a negative attribute. The R^2 values of CP effects and unexplained variance ranged from 0.09-0.23 and 0.3-0.4, respectively. For positive attribute testing, CAT yielded the highest sensitivity in differentiating products due to the lowest confounding effects and unexplained variance (residual) observed, followed by LIN and LAM. The R^2 value of confounding effects was 0.54, 16.21 and 91.73, respectively, for small product differences (A_B) and 0.47, 0.54 and 1.44, respectively for large differences (A_C). CAT or LIN can be applied for testing positive attributes with subtle difference product. However, when testing negative attributes, all three scales yielded similar performance.

To compare across scale types and to observe a contrast effect, Table 6.2 was calculated based on the Z values.

Table 6.2 Analysis of variance for testing effects of scale types and contrast effect

Factors	P value of positive attribute (testing overall liking)			
	$A_{(B)}$	B	$A_{(C)}$	C
Scale	0.99	0.99	0.99	0.99
Contrast	0.01	0.61	0.26	0.14
Scale*Contrast	0.77	0.37	0.81	0.56
Factors	**P value of negative attribute (testing bitterness perception)**			
	Mild		Strong	
Scale	1.00		1.00	
Contrast	0.80		0.74	
Scale*Contrast	0.61		0.73	

1) $A_{(B)}$ = a well-liked sample that paired with a moderately-liked sample; $A_{(C)}$ = a well-liked sample that paired with a disliked sample; B = a moderately-liked sample and C = a disliked sample.
2) The results were tested at alpha = 0.05 (P<0.05).
3) For scale effects, the results were tested among CAT, LIN and LAM scale based on the Z values. For length effects, the results were tested between 100 and 300 mm based on the Z values.

Based on the ANOVA result, it was found that an interaction between scale types and contrast effects (Scale*Contrast) was not observed. For main effects, scale types showed no impact on the hedonic ratings tested on either positive or negative attributes (P>0.05). However, the contrast effects significantly affected hedonic ratings of well-liked products (good/moderate) and tended to affect (though not significant) hedonic scores of disliked products (bad). When the magnitude of differences was small (good vs. moderate), consumers may get confused. However, it did not impact moderately liked products (moderate) for positive attribute testing.

To conclude, CAT, LIN or LAM can be applied in hedonic procedures for testing negative attribute questions such as bitterness. To test a positive attribute, researchers should all be cautious with contrast effects when testing products having small product difference.

To determine how contrast effects affected scale types and attribute question, Table 6.3 was calculated. For positive attribute, three levels of product impression: well-liked (Good: A), moderately liked (moderate: B) and disliked (bad: C) samples were tested. In this case, the significant difference (low in P value) means that products differed in hedonic scores due to contrast effects.

Table 6.3 Analysis of variance for three scales and two attribute questions

Positive attribute				
Sample	Contrast	CAT	LIN	LAM
	AB	7.72+0.80	7.68+1.13	78.78+9.80
Good/Moderate	BA	7.30+1.18	6.68+1.82	75.03+19.02
	P value	0.11	**0.04**	0.32
	AC	7.58+1.24	7.38+1.09	77.63+12.51
Good/Bad	CA	7.53+1.14	6.94+1.94	75.33+15.03
	P value	0.86	0.28	0.62
	BA	6.31+1.17	6.29+1.64	65..25+16.15
Moderate	AB	6.53+1.48	6.16+1.65	61.73+15.78
	P value	0.66	0.77	0.51
	CA	3.79+1.63	3.70+1.43	31.66+19.15
Bad	AC	3.37+1.43	3.71+1.73	39.67+19.07
	P value	**0.01**	0.31	0.14
Negative attribute				
Sample	Contrast	CAT	LIN	LAM
	MS	6.48+1.92	6.39+1.67	62.92+19.0
Mild	SM	6.29+1.38	6.65+1.50	63.56+19.97
	P value	0.56	0.95	0.64
	SM	2.67+1.39	2.76+1.71	23.85+15.75
Strong	MS	2.44+1.38	2.78+1.91	24.49+18.23
	P value	0.39	0.95	0.85

1) P value < 0.05 in the column indicated that at alpha=.05 their corresponding juices/broths differ significantly in their mean scores within each sample category.
2) $A_{(B)}$ = a well-liked sample that paired with a moderately-liked sample; $A_{(C)}$ = a well-liked sample that paired with a disliked sample; B = a moderately-liked sample and C = a disliked sample.
3) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

It was found that testing a disliked sample with CAT was more affected by contrast effects (significance; P<0.05), followed by testing well-liked (paired with moderately liked; P=0.11) and moderately liked sample (P=0.66); however, the opposite of order was observed when testing with LIN. The well-liked sample (paired with moderately liked) using LIN was more affected by contrast effects (significance; P<0.05), followed by disliked and moderately liked sample. The conclusion of contrast effects on CAT and LIN remains unclear as to which scale or which product is affected the most. Nevertheless, it was likely to happen in a well-liked product for both scales (0.11 vs. 0.04 for CAT vs. LIN). For testing positive attribute with LAM, it seemed to give more consistent results across products but the results had a high standard deviation that may cause a lower F value and, thus, insignificant difference. For moderately-liked product testing, there were no significant effects from all three scales. For negative attribute testing, all three scales showed similar performance in testing samples with a minimum level of CP effects.

To evaluate the CP effects more clearly, Table 6.4 was created to determine the size/magnitude of CP effects without treatment effects being included in the model.

Table 6.4 Shared explained variance for each attribute question

Scale	*Positive attribute*			*Negative attribute*		
	Sample	R^2		Sample	R^2	
		Contrast	Panelist		Contrast	Panelist
CAT	$A_{(B)}$	**4.57**	3.14			
	$A_{(C)}$	0.07	2.51			
	B	**0.69**	0.17	Mild	7.03	**72.34**
	C	**1.93**	0.16	Strong	6.23	**85.52**
LIN	$A_{(B)}$	**76.29**	4.45			
	$A_{(C)}$	**35.71**	35.71			
	B	0.46	**68.87**	Mild	14.18	**60.85**
	C	0.00	**23.01**	Strong	0.003	**2.45**
LAM	$A_{(B)}$	18.56	**63.71**			
	$A_{(C)}$	22.90	**23.5**			
	B	18.45	**73.96**	Mild	2.72	**4.01**
	C	51.52	24.28	Strong	2.02	**44.75**

1)$A_{(B)}$ = a well-liked sample that paired with a moderately-liked sample; $A_{(C)}$ = a well-liked sample that paired with a disliked sample; B = a moderately-liked sample and C = a disliked sample.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

Based on the Z score testing, the results can be compared across product types and scales. The higher coefficient R^2 values represented the importance of such factor on hedonic variation. In this case the higher R^2 value implied that scales/products were more susceptible to contrast/panelist effects. The stronger the contrast/panelist effects, the lower the discriminative power. For a positive attribute, a high R^2 value for well-liked sample was associated with contrast effects for a few cases. Considering among scale types, CAT was prone to contrast effects rather than panelist effects. The ease of scale may contribute to lower panelist effects. The majority of explained variance came from treatment effects rather than biases for CAT which was desirable in this case. The contrast effects observed in LIN (Table 6.4) may result in significant differences for well-liked products shown in Table 6.3. However, moderately liked and disliked products were not significant (Table 6.3) due to susceptibility of panelist effects but not due to contrast effects. The paradox of results from LIN remained unclear. A further experiment can be performed to test a complexity of scale by increasing a sample size and period of warm up session. This experimental design (RCBD) was designed to exclude the panelist effect from the model by treating it as a random factor so that we couldn't observe the effect through the analysis. Likewise for LAM, the score fluctuation came from a panelist effect rather than a contrast effect, except for disliked sample. The contrast effects seemed to impact more on bad quality samples which supported the finding reported in Chapter 4.

For a negative attribute, the result showed that panelist effects were much greater than the contrast effect for CAT (72-85%), LIN (2-60%) and LAM (4-44%); the score fluctuation was mainly caused by panelists rather than contrast effects. No matter which position the sample was served, the score remained constantly low; this was observed on all scale types and on both high and low bitterness concentrations. In conclusion, for positive attribute testing, CAT and LIN scales were affected mostly by contrast effects particularly for well-liked products that were paired with similarly liked products whereas LAM was prone to panelist effects, except for disliked products. However, for testing a negative attribute question, all scales had minimal contrast effects between both samples: a strong bitterness sample (S) and a mild bitterness sample (M) but were affected by panelist effects.

6.6.2 An effect of scale polarity

The coefficients of determination (R^2) values (Table 6.5) were similar for both bipolar and unipolar scales. Comparing the sensitivity among three scales, both bi-polar and uni-polar scales showed a similar trend, CAT> LIN > LAM. The R^2 of treatment effects from bipolar scales (99.36-99.61) was slightly higher, but insignificantly compared with that of unipolar scales (88.60-98.86).

Table 6.5 Coefficient of determination (R^2) for different scale polarities

Attributes	Scale	R^2			
		Treatment	Contrast	Panel	Residual
Negative [Bi-polar]	CAT	99.61	0.03	0.09	0.30
	LIN	99.51	0.08	0.01	0.39
	LAM	99.36	0.00	0.23	0.40
Negative [Uni-polar]	CAT	98.86	0.31	0.38	0.45
	LIN	97.29	0.55	0.50	1.66
	LAM	88.60	0.73	10.09	0.58

CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

Theoretically the higher the R^2 value of a factor, the higher the power of an experiment that can be explained by such factor. Based on this study, polarity of a scale showed a slight effect on treatment differentiation. The unipolar scale had relatively high CP effects especially with LAM (unipolar vs. bipolar: 0.69 vs. 0.11; 1.05 vs. 0.09; 10.82 vs. 0.23 for CAT, LIN and LAM, respectively). LAM was affected by CP the most on both scale polarities. The reason remained unclear at this point. The residual of an experiment for a unipolar scale was also higher for all three scale types (0.45 vs. 0.30; 1.66 vs. 0.39; 0.58 vs. 0.4 for CAT, LIN and LAM, respectively) compared with a bipolar scale. Considering the given space in a unipolar scale, it may allow consumers more room to response. This may contribute to an increased discriminative power as also observed for a longer line scale in chapter 5. The result was corresponding with a previous chapter demonstrating that the longer line (more space) had higher discriminative power; however, it induced a higher standard deviation. To increase a power of an experiment, increasing a sample size may be done. Based on this study, a unipolar scale may not be appropriate because it is sensitive to length effects (proved in chapter 5).

To investigate the contrast effects on each scale, the hedonic mean scores for a negative attribute were analyzed using the statistic F value, the associated P value and the covariance values as shown in Table 6.6. The contrast effect was observed between two bitterness concentrations of chicken broth samples where S represented a strong bitterness sample and M represented a mild bitterness sample.

Table 6.6 Analysis of variance for three scales and two scale polarities

Unipolar scale				
Sample	Contrast	CAT	LIN	LAM
Mild	MS	3.87±1.05	4.06±1.11	40.57±10.32
	SM	3.94±0.96	4.52±3.91	41.59±10.39
	F value	0.98	0.79	0.29
	P value	0.3233	0.3757	0.5937
Strong	SM	1.98±0.94	2.06±1.15	22.53±15.02
	MS	1.83±1.04	1.94±1.20	19.95±12.55
	F value	1.11	0	0.99
	P value	0.2942	0.9945	0.3215
Bipolar scale				
Sample	Contrast	CAT	LIN	LAM
Mild	MS	6.29±1.38	6.39±1.67	62.92±19.0
	SM	6.48±1.92	6.63±1.62	63.56±19.97
	F value	0.34	0.57	0.22
	P value	0.5619	0.4531	0.6434
Strong	SM	2.67±1.39	2.76±1.71	23.85±15.75
	MS	2.44±1.38	2.78±1.91	24.49±18.23
	F value	0.76	0	0.04
	P value	0.3864	0.9523	0.846

1) P value < 0.05 in the column for each sample type (Mild or Strong) indicated that at alpha=.05 their corresponding chicken broths differ significantly in their mean scores within each sample category.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

To determine a positional effect and contrast effects, two hedonic mean scores between two positions were compared (SM and MS). It is known that the higher the F value, the higher the discriminative power for product discrimination. However, the F values in this case were susceptible to contrast effects. The lower the F value, the higher chance of getting the same hedonic scores no matter which position the sample was served; this implied the reliability of getting true responses with minimal level of contrast and/or positional effects involved. Ideally we expected to have an insignificance difference of hedonic scores when testing the same products even with a different presentation order. The phenomena of contrast effects will be observed when presenting a bad sample prior to a good one, the score of good sample will be inflated and vice versa. The result from this study was in agreement with this rule. According to Table 6.6, the hedonic scores of mild samples presented after a strong bitterness sample was slightly higher but not significant (6.29 vs. 6.48, 6.39 vs. 6.63 and 62.92 vs. 63.56 for CAT, LIN and LAM, respectively) for bipolar scales. The hedonic score of strong bitterness sample presented after a mild sample was lower but not significant. This is true for both bi- and unipolar scales. Based on this study, we concluded that positional and contrast effects did not impact hedonic scales (CAT, LIN and LAM) for testing a negative attribute questions tested on both scale polarities. However, contrast and positional effects were likely to have an impact on CAT due to the low P value (higher chance of getting significant difference) observed in this study.

This experiment revealed the effects of scale types and polarities on hedonic ratings. Table 6.7 showed the ANOVA results of hedonic ratings from the same samples using three scales and two scale polarities. The F value in this case was susceptible to different scale polarities. The assumption of this experiment was that the hedonic scores of both polarities were similar or were not significantly different. The lower the F value, the higher the chance of getting the same hedonic scores no matter which scale was used. The P values were calculated based on hedonic means, and the histogram of each polarity was presented.

Table 6.7 Analysis of variance testing scale polarity for negative attributes (Bitterness)

Sample	Scale	CAT	LIN	LAM
Mild	Bi	6.39±1.67	6.49±1.67	63.24±19.4
	Uni	3.91±1.00	4.28±2.86	41.08±10.32
	F value	176.14	27.02	110.78
	P value	<0.0001	<0.0001	<0.0001
Strong	Bi	2.56±1.38	2.77±1.80	24.17±16.92
	Uni	1.91±0.99	1.99±1.17	21.24±13.83
	F value	16.15	14.19	1.99
	P value	<0.0001	<0.0001	0.1595

1) P value < 0.05 in the column for each sample indicated that at alpha=.05 their corresponding chicken broths differ significantly in their mean scores.
2) CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

Polarity effects were obvious for a mild sample showing significantly different results between uni- and bi-polar scales (3.91 vs. 6.39; 4.28 vs.6.49 and 41.05 vs. 63.24) for CAT, LIN and LAM, respectively. Noticeably, the hedonic scores for a mild sample were placed an opposite in category comparing the bi- and uni- polar scales, which certainly result in misunderstanding and misinterpretation of the results. The question is whether the hedonic differences caused by the nature of products or by scale polarity. The histogram chart may help to clarify this assumption which will be explained later. However, when testing with a strong bitterness sample, CAT and LIN showed statistical significance but not LAM. The scores for each polarity were closer than those of mild samples. The F value from CAT was highest followed by LAM and LIN for a mild sample but the order changed with a strong sample: CAT, LIN and LAM. The conflict remained unclear at this point. In conclusion, all three scales yielded similar results (uni- or bi- polar scale) when testing with a mild sample; however, LAM seemed to have a consistency pattern of testing a negative attribute with a strong bitterness sample.

The histogram (Figure 6.1 and 6.2) can be used to determine if the above finding was meaningful. To verify the result, the data distribution was performed on each scale polarity: a unipolar scale (Figure 6.1) and a bipolar scale (Figure 6.2) as shown below. Interestingly, this study found that testing a negative attribute was prone to have a skewed distribution rather than a normal distribution, likely due to panelist biases. The score for the mild bitterness sample was skewed to the left whereas the strong bitterness sample was skewed to the right. Figure 6.1 show that all scales yielded similar distribution for both concentrations (mild and strong bitter) except for LAM (unipolar, strong bitterness sample). Considering within the same unipolar scale, LAM seemed to have a clear and consistent pattern. Figure 6.2 shows that both CAT and LAM yielded similar distribution for testing a mild sample whereas LIN yielded a scattering distribution. However, for testing a strong bitterness sample, all scales yielded similar distribution. This finding confirmed that individuals differed in bitterness perception (Horne *et al.*, 2002). At a high level of bitterness, hedonic responses could not be differentiated (Drewnowski *et al.*, 1997). The conclusion regarding the reliability of testing a mild bitterness sample with CAT and LAM can be concluded. CAT was slightly sensitive to a scale polarity leading to have a misinterpretation of results. LAM seemed to have a clear and consistent pattern over the other two scales but it came with a high standard deviation. To increase a sample size may help to increase a power of an experiment. The result from LIN was inconclusive regarding a high standard deviation and an abnormal distribution. For testing a strong bitterness sample, all scales provided a similar distribution and the distribution is concentrated on the left. Again, LAM seemed to have a consistent pattern of testing a strong bitterness sample.

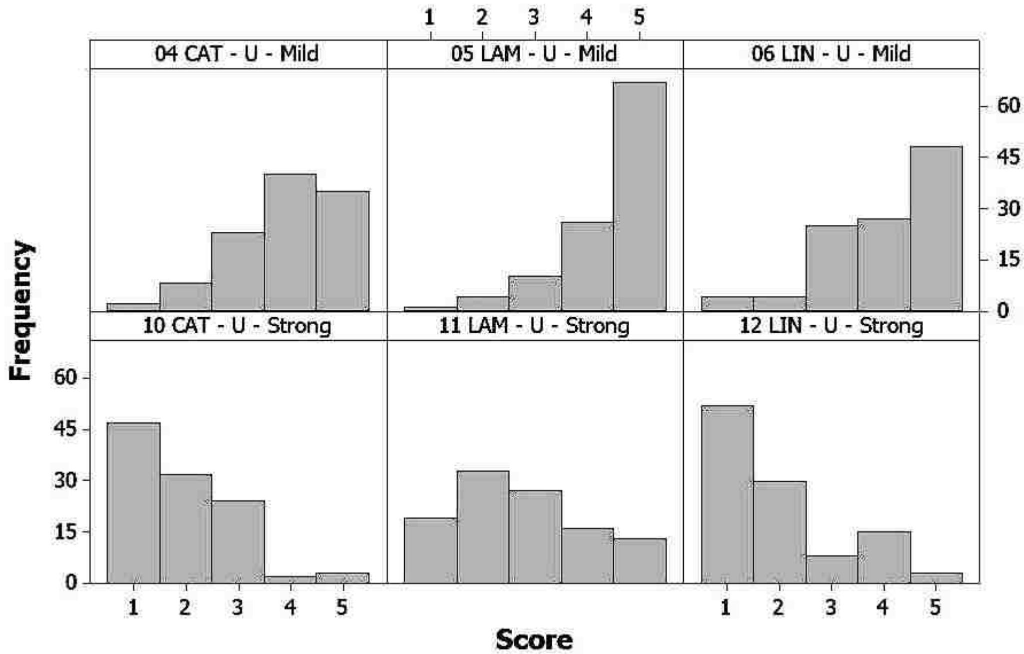

Figure 6.1 Histogram of overall bitterness perception of chicken broths for mild and strong bitterness sample for each scale (CAT, LAM and LIN) from a unipolar scale.

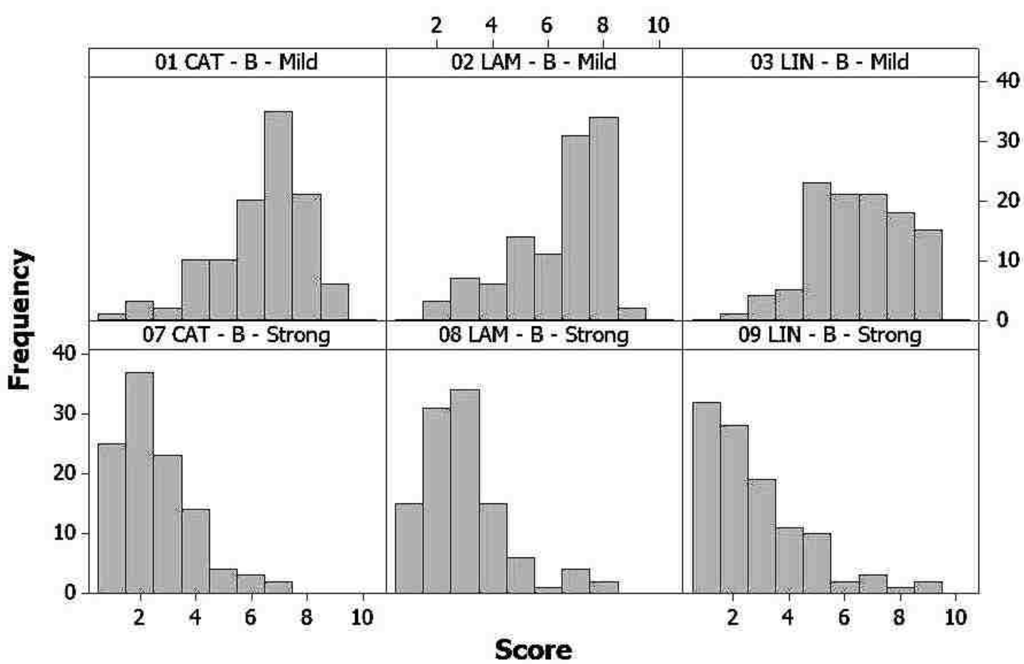

Figure 6.2 Histogram of overall bitterness perception of chicken broths for mild and strong bitterness sample for each scale (CAT, LAM and LIN) from a bipolar scale.

In order to extract the CP effects, the ANOVA approach (Table 6.8) was conducted in the MIXED procedure and the coefficient of determination (R^2) was calculated.

Table 6.8 Shared explained variance for each scale polarity across three scale types

Scale	*Unipolar*			*Bipolar*		
	Sample	R^2		Sample	R^2	
		Contrast	Panelist		Contrast	Panelist
CAT	Mild	10.20	20.41	Mild	7.03	**72.34**
LIN		**36.95**	10.39		14.18	**60.85**
LAM		2.28	**89.78**		2.72	**4.01**
CAT	Strong	27.70	25.82	Strong	6.23	**85.52**
LIN		**20.21**	9.55		0.003	**2.45**
LAM		10.65	**78.61**		2.02	**44.75**

CAT= Categorical hedonic scale (1-9); LIN= Line hedonic scale (1-9); LAM= Labeled Affective Magnitude scale (0-100).

In this study, the counterbalanced experiment was practiced to minimize the presentation biases; however, to investigate the CP effects, the model without the treatment effects was carried out. It was found that roughly 2-90% of CP and unexplained variance affected the score variation for a mild sample, 10-79% for a strong sample when tested using a unipolar scale whereas 3-72% for a mild sample and 0-85% for a strong sample when tested using a bipolar scale. The fraction of contrast effects in unipolar scales was bigger than in bipolar scales (Table 6.1) (1.14 vs. 0.4; 2.39 vs. 0.49; 11.4 vs. 0.64 for CAT, LIN and LAM, respectively). Based on Table 6.7, the CP effects were strongly pronounced for a unipolar scale for LAM (mild, unipolar vs. bipolar: 92.06 vs. 6.73; Strong: 89.26 vs. 46.77) and it was mainly due to panelist effects. The paradox of results was again found with LIN making this result inconclusive. However, when testing a negative attribute with CAT, one should be aware of polarity effects. It was found that using a CAT bipolar scale was more susceptible with CP effects (mild, unipolar vs. bipolar: 30.61 vs. 79.37; Strong: 53.52 vs. 91.75). Although the effects were observed with a unipolar scale, no specific pattern can be observed.

In conclusion testing a product using CAT and LIN was influenced more by contrast effects and panelist effects and scale polarities, whereas LAM was prone to panelist effects on both scale polarities. Comparing between scale polarities, CAT was affected by the CP effects and it was strongly evidenced with a bipolar scale for both strong and mild bitterness products. LIN seemed to have a better score pattern (Table 6.6). When testing products using CAT and LIN a researcher needs to be aware of CP effects while LAM is mainly more affected by the panelist effects.

6.7 Conclusion

Traditional 9-point hedonic scale has been used to assess the degree of liking/disliking since 1940 but it has shown negative and inaccuracy results for negative sensory attributes. This experiment concluded that CAT or LIN yielded similar performance to assess the degree of liking/disliking for both positive and negative attributes; however, LAM can be used for testing a negative attribute. The low sensitivity in testing positive attributes with LAM was mostly due to CP effects; however, more evidences or further experiments are needed to confirm this finding. Testing a well-liked product, a researcher should be aware of contrast effects for positive attributes but for a negative attribute less or no contrast effects but panelist effects played a more important role. This experiment also revealed the effect of scale polarity on hedonic ratings. The bipolar scale exhibited a slightly better performance. The unipolar scale induced a higher score fluctuation due to a more flexible space on the scale. Polarity effects were obvious for a mild sample resulting in misinterpretation. LAM yielded the consistency pattern of testing a negative attribute with both uni- and bi- polar scales whereas CAT was sensitive especially when testing with a mild bitterness sample. However, when using LAM, experimenters should be aware of severe panelist effects. Increasing a sample size may help increase a power of an experiment. It was recommended that when using CAT and LIN, a researcher should be aware of CP effects.

6.8 References

Bartoshuk, L.M. 1979. Bitter taste of saccharin related to the genetic ability to taste the bitter substance 6-n-propylthiouracil. ***Science*** 205: 934-935.

Cardello, A.V. and Schutz, H.G. 2004. Research Note Numerical scale-point locations for constructing the LAM (Labeled affective magnitude) scale. ***Journal Sensory Studies.*** 19: 341-346.

Cordonnier, S.M. and Delwiche, J.F. 2008. An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. ***Journal Sensory Studies*** 23: 284-292.

Dine, A.N. and Olabi, A. 2009. Effect of reference foods in repeated acceptability tests: Testing familiar and novel foods using 2 acceptability scales. ***Journal of Food Science.*** 74: S97-S105.

Drewnowski, A., Henderson, S.A. and Shore, A.B. 1997. Genetic sensitivity to 6-n-proprylthiouracil (PROP) and hedonic responses to bitter and sweet tastes. ***Chemistry Sense*** 22: 27-37

Greene, J.L., Bratka, K.J., Drake, M.A. and Sanders, T.H. 2006. Effective of category and line scales to characterize consumer perception of fruit fermented flavors in peanuts. ***Journal Sensory Studies*** 21: 146-154.

Hein, K.A., Jaeger, S.R., Carr, B.T. and Delahunty, C.M. 2008. Comparison of five common acceptance and preference methods. ***Food Quality and Preference*** 19: 651-661.

Horne, J., Lawless, H.T., Speirs, W. and Sposato, D. 2002. Bitter taste of saccharin and acesulfame-K. ***Chemistry Sense*** 27: 31-38.

Lawless, H.T. and Malone, G.J. 1986. A comparison of scaling methods: Sensitivity, replicates and relative measurement. ***Journal Sensory Studies*** 1: 155-174.

Lawless, H.T., and Heymann, H. 1999. ***Sensory Evaluation of Food: Principal and Practices***. Maryland: Aspen publisher. 827p.

Lee, H.J. and Kim, K.O. 2001. Effect of forgetting on various protocols for category and line scales of intensity. ***Journal Sensory Studies*** 16: 327-342.

Peryam, D.R. 1954. Field testing of armed force rations in Peryam, D.R., Pilgrim, F.L. and Peterson, M.S. ***Food Acceptance rating methodology: A symposium*****.** Advisory board on Quartermaster research and development 75-78. National Research Council, Washington D.C.

Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preference. ***Food Technology*** *(September)*, 9-14

SAS 2003. ***SAS/IML software: Usage and reference,*** Version 9. SAS Institute Inc., Cary, NC, USA.

CHAPTER 7 SUMMARY AND CONCLUSION

Psychological biases induced by sample presentation order including halo effects, stimulus error, logical error, error of central tendency, contrast and convergence effects (Dine and Olabi, 2009; Lawless and Heymann, 1999; Lee and Meullenet, 2010; Meilgaard *et al.* 2006 and Stone and Sidel, 1993) can provoke a negative impact and misinterpretation of the sensory results, thus lowering an experimental power. Several proper practices such as experimental design, methods for selecting samples and data analysis methodology have been proposed to help minimize such biases.

Randomized complete block design (RCBD) has been extensively used in consumer testing; however, it is ineffective in preventing sample presentation biases. This dissertation demonstrated a more efficient experimental design (i.e., Split Plot Repeated Randomized Complete Block Design: SPRCBD) to help minimize positional and FSO (First Serving Order) biases in consumer tests. Results suggested that positional bias was more pronounced for the poor-quality sample. Comparing RCBD and RCBD w/o FSO, there were significant differences between overall liking scores (OL) of the same sample served at the same position. Hence, omitting the first sample score from data analysis was not recommended. The mean-squared error (MSE) of SPRCBD was lower than RCBD and RCBD w/o FSO (2.28, 2.72, and 2.60, respectively), indicating a more powerful design to explain variations in mean differences. SPRCBD extracted more explained variances, resulting in a decreased Type-II error in the model.

Contrast effects are one of the psychological biases mainly observed in consumer acceptance tests. Contrast effects (inflated differences) are caused by the order of sample presentation. It has been proven to pronounce in sensory testing; however, so far there is no published work reporting the methodology to quantify the magnitude of sensory contrast effects. This dissertation demonstrated the use of logistic regression analysis (LRA) to quantify such effects. LRA has been successfully applied and it was found that when the poor-quality sample (C) was presented first, the liking scores of SS (Second Served) was higher significantly ($P<0.05$). This was true for color, taste and overall liking attributes. To quantify the magnitude of contrast effects, the odds ratio was estimated. The estimation procedure and interpretation can be found in chapter 4. Surprisingly, there was a huge variation within consumers. Roughly 20-25% of consumers altered their ratings after the first sample was served even for the identical samples. Rating multidimensional attributes (i.e., taste) tended to be more affected by the contrast effects than did simpler attribute (i.e., color) rating. Factors related to the hedonic test protocol including scale types, scale lengths, product overall impression, scale polarities and attribute questions may influence hedonic scores and could prevent experimenters from getting true responses.

Several scales have been developed as an alternative choice for assessing the degree of food liking/disliking. This study is the very first consumer study to determine a performance of hedonic scales (CAT, LIN and LAM associated with length effects) i.e., sensitivity, reliability, correlation and the neutral response's behavior of each scale. It was found that LAM had a high discriminative power, sensitivity and reliability with a high Cronbach's alpha value and a high Pearson's correlation coefficient when testing two identical samples. The benefit from using LAM is that frequent ratings in the permitted area above like extremely and below dislike extremely categories for well-liked and disliked products, respectively, were observed. Among three scales, the hedonic CAT scale was sensitive to length effects; however, it yielded a similarly high reliability, and a high Pearson's correlation with the lowest neutral tendency rate.

The hedonic LIN scale had a low to moderate discriminative power, sensitivity, reliability, Pearson's correlation but yielded the highest rate of neutral responses. The scale length effects were observed; however, there was no specific pattern for discriminative power and sensitivity. Both scale lengths (100 vs. 300 mm.) yielded similar reliability Cronbach's alpha values (0.777-0.919) which implied no superiority over each other. The percentages of neutral responses were slightly higher for the longer scale. Based on this study, it was suggested that one should use a regular length (100 mm.) as a standard for hedonic testing, and LAM as an alternative choice for assessing a degree of food liking/disliking when the length effects may be a critical issue. Even though a traditional 9-point hedonic scale has been used to assess the degree of liking/disliking for testing a positive attribute for several decades, it has shown negative and inaccurate results for negative sensory attribute testing.

This study revealed some advantages and disadvantages of hedonic scales induced by scale polarity/types and attribute questions. It was found that consumers better differentiated negative-attribute ratings yielding a higher R^2 of treatment effects. The positive attribute question was susceptible to confounding effects. With bipolar scales, CAT or LIN yielded similar performance in assessing the degree of liking/disliking for both positive and negative attributes; however, LAM could be used for negative attribute testing. The low sensitivity in testing positive attributes with LAM was mostly due to CP (contrast and panelists) effects; however, more evidences or further experiments are needed to confirm this finding. When testing a well-liked product a researcher should be aware of contrast effects for positive attributes and panelist effects for a negative attribute. Polarity effects were obvious for a low level of bitterness, showing significant differences between uni- vs. bipolar scales [3.91 vs. 6.39, 4.28 vs. 6.49, and 41.05 vs. 63.24, respectively, for CAT, LIN and LAM]; all ratings from bipolar scales were not on the negative-side leading to a wrong category rating. For the strong bitterness sample, unipolar and bipolar ratings were on the negative side with LAM having more consistent pattern. This study revealed some drawbacks of hedonic scales induced by scale polarity/types and attributes. CAT and LIN were more affected by CP effects while LAM was more affected by panelist effects.

Overall, this study demonstrated only a few factors that affected the hedonic ratings. There are many more factors that could sway sensory responses and prevent experimenters from getting a true outcome. Further studies on fundamental physic-psychological biases are recommended.

APPENDIX A RESEARCH CONSENT FORM

APPROVED BY
LSU AG CENTER
IRB AS HE11-29
ON 12-14-2011

Research Consent Form

I, ________________________, agree to participate in the research entitled "Consumer Acceptance of New Food Products" which is being conducted by Dr. Witoon Prinyawiwatkul, Professor of the Department of Food Science at Louisiana State University, Agricultural Center, phone number (225) 578-5188.

I understand that participation is entirely voluntary and whether or not I participate will not affect how I am treated on my job. I can withdraw my consent at any time without penalty or loss of benefits to which I am otherwise entitled and have the results of the participation returned to me, removed from the experimental records, or destroyed. Up to 150 consumers will participate in this research. For this particular research, about 15-20 minutes participation will be required for each consumer.

The following points have been explained to me:

1. In any case, it is my responsibility to report prior participation to the investigator any food allergies I may have.

2. The reason for the research is to gather information on sensory acceptability of new food products. The benefit that I may expect from it is a satisfaction that I have contributed to quality improvement of these products.

3. The procedures are as follows: 3-5 coded samples will be placed in front of me, and I will evaluate them by normal standard methods and indicate my evaluation on score sheets. All procedures are standard methods as published by the American Society for Testing and Materials and the Sensory Evaluation Division of the Institute of Food Technologists.

4. Participation entails minimal risk: The only risk which can be envisioned is that of an allergic reaction common food ingredients [red beans, bell pepper, onion, garlic, celery, thyme, cayenne pepper, bay leaf, pork products, rice and rice products, milk and dairy products, wheat flour, tapioca flour, eggs, table sugar, vanilla, sweet potato, salt (sodium chloride) and salt substitute (potassium chloride), and plain unsalted crackers]. However, because it is known to me beforehand that the food to be tested contains common food ingredients, the situation can normally be avoided.

5. The results of this study will not be released in any individual identifiable form without my prior consent unless required by law.

6. The investigator will answer any further questions about the research, either now or during the course of the project.

The study has been discussed with me, and all of my questions have been answered. I understand that additional questions regarding the study should be directed to the investigator listed above. In addition, I understand the research at Louisiana State University, Agricultural Center, which involves human participation, is carried out under the oversight of the Institutional Review Board. Questions or problems regarding these activities should be addressed to Dr. Michael Keenan, Chair of LSU AgCenter IRB, (225) 578-1708. I agree with the terms above and acknowledge.

I have been given a copy of the consent form.

Signature of Investigator　　　　　　Signature of Participant

Witness: ________________　　　　　　Date: ________________

LSU AgCenter Institutional Review Board (IRB)
Dr. Michael J. Keenan, Chair
School of Human Ecology
209 Knapp Hall
225-578-1708
mkeenan@agctr.lsu.edu

Application for Exemption from Institutional Oversight

All research projects using living humans as subjects, or samples or data obtained from humans must be approved or exempted in advance by the LSU AgCenter IRB. This form helps the principal investigator determine if a project may be exempted, and is used to request an exemption.

- Applicant, please fill out the application in its entirety and include the completed application as well as parts A-E, listed below, when submitting to the LSU AgCenter IRB. Once the application is completed, please submit the original and one copy to the chair, Dr. Michael J. Keenan, in 209 Knapp Hall.

- A Complete Application Includes All of the Following:
  - **(A)** The original and a copy of this completed form and a copy of parts B through E.
  - **(B)** A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1 & 2)
  - **(C)** Copies of all instruments and all recruitment material to be used.
    - If this proposal is part of a grant proposal, include a copy of the proposal.
  - **(D)** The consent form you will use in the study (see part 3 for more information)
  - **(E)** Beginning January 1, 2009: Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing and handling data, unless already on file with the LSU AgCenter IRB. Training link: (http://grants.nih.gov/grants/policy/hs/training.htm)

1) Principal Investigator: Witoon Prinyawiwatkul **Rank:** Professor **Student?** Y/(N)
Dept: Food Science **Ph:** 8-5188 **E-mail:** wprinya@lsu.edu
2) Co-Investigator(s): please include department, rank, phone and e-mail for each

- If student as principal or co-investigator(s), please identify and name supervising professor in this space

None

3) Project Title: Consumer Acceptance of New Food Products
4) Grant Proposal?(yes or (no))____ **If Yes, Proposal Number and funding Agency**____
Also, if Yes, either: **this application completely matches the scope of work in the grant Y/N**__
OR
more IRB applications will be filed later Y/N__
5) Subject pool (e.g. Nutrition Students) LSU Faculty, staff, students, and off-campus consumers

- Circle any **"vulnerable populations" to be used:** (children<18, the mentally impaired, pregnant women, the aged, other). Projects with incarcerated persons cannot be exempted.

6) PI signature ________ ****Date** 12/9/2011 (no per signatures)
****I certify that my responses are accurate and complete.** If the project scope or design is later changed I will resubmit for review. I will obtain written approval from the Authorized Representative of all non-LSU AgCenter institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at the LSU AgCenter for three years after completion of the study. If I leave the LSU AgCenter before that time the consent forms should be preserved in the Departmental Office.

Committee Action: Exempted ✓ Not Exempted____ **IRB#** HE11-29
Reviewer Michael Keenan Signature Michael Keenan Date 12-14-2011

APPENDIX B GRAPE JUICE PRELIMINARY

B1 Screening Test

Three commercial grape juice brands were randomly selected from local markets in Baton Rouge, Louisiana area including Wal-Mart, Winn Dixie, Albertson and Target supermarket. Grape juices were pre-screened based on a sensory attribute including color, transparency, grape flavor, sweetness, and sourness. The pre-observed brands were:

1. Market Pantry (A)
2. Juicy Juice DHA (B)
3. Kool aid (C)
4. Welch's 100% (D)
5. Welch's light (F)
6. Apple ave (E)
7. Great value (Wal-Mart brand) (G)

To minimize the possible biases as well as to ensure a proper product selection, the sensory testing was performed with 7 expert panelists who regularly consume grape juice and have bought these products for the past 3 months. Grape juices were categorized into 3 categories: bad, moderate, and good quality. To determine an effect of product impression affected on hedonic scores, all samples have been tested and were classified as abovementioned. The result was shown in Table B11.

Table B11 Grape juice classifications

	Good	Moderate	Bad
Brand	D, A	F	B,G

For prescreening, Brand C couldn't be categorized as in the same grape juice category as it contains a strong artificial flavor not. Brand E has a strong apple flavor. Considering between the price and availability, three brands, which were D (good), F (moderate), and B (bad) were selected. To ensure the proper product selection, we conducted a small sensory test using a ranking method with three selected samples. The consumers including student, staffs and faculty (n=30) were randomly recruited by convenience sampling method within the department of food science, Louisiana State University.

The result was agreed with the prescreening test. Sample Welch's 100% were ranked at the first place (most like) followed by sample Welch's light and sample Juicy DHA, respectively. The result for each rank sum was shown in Table B12 and B13.

Table B12 Rank sum of grape juices

Rank sum	Treatment		
	Juicy DHA	Welch's 100%	Welch's light
1	0	20	7
2	8	16	42
3	78	6	6
Total	86	42	55

Table B13 Difference of Rank sum

	Treatment 1	Treatment 2	Treatment 3
Treatment 1	-	13	44
Treatment 2		-	31
Treatment 3			-

The estimation was approached by binomial model using the table from Christensen *et al.* (2006). The maximum distance between samples or the critical values of differences between rank sums at 90% significant confident interval was 16. The results showed that the maximum distance between samples was 44. This concluded that the overall difference was observed. Then we proceeded to determine which tested pair was significantly difference by using the table from Christensen *et al.* (2006). Multiple comparisons had been performed. The critical values of differences between rank sums at 90% significant confident interval were 13. It was concluded that all possible pairs were significantly different from one another. The hedonic score of treatment 1 was significantly different from treatment 2 and 3, and treatment 2 was also significant from treatment 3. Different letter showed significantly differences as $T1_c$ $T2_b$ $T3_a$. These three samples were clearly defined the magnitude of product impression in this study. The rank order was corresponded with a prescreening result. To complete an entire experiment, the approximated usage and price of each sample showed in Table B14.

Table B14 Usage and price.

	Juicy DHA (C)	Welch's 100% (A)	Welch's light (B)
Price	2.62/ 1L	2.00/ 1.89L	3.59/ 1.89L
App. Usage	20 mL per cup		
Amount (550 panel x 2sets) plus 180 (each first serve)	1280 cups =25600 mL		
Needed (Bottles)	26	14	14
Prices ($)	68	28	51

B2 Sample presentation

Layout of sample presentation for each experiment (4 and 3 treatment per consumer) showed in Table B21 and B22.

Table B21 Set of four samples

First position (X2)	Subjects	Treatment (X1)		
184	30	701	384	629
184	30	701	629	384
184	30	384	701	629
184	30	384	629	701
184	30	629	701	384
184	30	629	384	701
551	30	701	384	629
551	30	701	629	384
551	30	384	701	629
551	30	384	629	701
551	30	629	701	384
551	30	629	384	701
610	30	701	384	629
610	30	701	629	384
610	30	384	701	629

Table B21 Continued

First position (X2)	Subjects	Treatment (X1)		
610	30	384	629	701
610	30	629	701	384
610	30	629	384	701

Note 184 and 701 is Welch's 100% (A)
551 and 384 is Welch's light (B)
610 and 629 is Juicy DHA (C)

Table B22 Set of three samples

Subjects	Treatment (X1)		
90	518	249	783
90	518	783	249
90	249	518	783
90	249	783	518
90	783	518	249
90	783	249	518
540			

Note 518 is Welch's 100% (A) = 540
249 is Welch's light (B) = 540
783 is Juicy DHA = 540

APPENDIX C CHICKEN BROTHS PRELIMINARY

Table C11 Chicken broth formulation

Ingredient	Weight	Process
Water	5 Liters	Boil on Hi for 20 mins
Chicken	5 pound	@room temp.
Celery	380 g.	Rosted@400^{0}C 20 mins
Onion	260 g.	
Garlic	6 g.	
Tomato	100 g.	
Pepper corn	0.7 g.	
		Simmer for 10 hrs.

Note: N = 216 consumers (108 for Bipolar and 108 for unipolar scale)

Sample: 098, 147, 278, 198, 247, 378
Mild bitter sample with a mixture of NaCl and KCL (2:1) at 1.3 % by weight

Sample: 511, 705, 665, 611, 805, 765
Strong bitter sample with a KCL 2% by weight

Chicken broth screening questionnaire:

Please rank each sample from 1-3 (1= like the most, 2= like moderately and 3= like the least).

Sample	783	518	249
Chicken flavor			
Saltiness			
Overall liking			

How would you rate the **OVERALL LIKING** of the product that you like the most?

Dislike Extremely	Dislike Very much	Dislike Moderately	Dislike Slightly	Neither Like nor Dislike	Like Slightly	Like Moderately	Like Very much	Like Extremely
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

1) How would you rate the **Chicken Flavor** of these products?

Sample	Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
714									
098									
511									

2) How would you rate the **Saltiness** of these products?

Sample	Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
714									
098									
511									

3) Have you perceived the **Bitterness** of these products? **And** how would you rated the **Bitterness** of these products?

Sample	Dislike Extremely	Dislike Very Much	Dislike Moderately	Dislike Slightly	Neither like Nor Dislike	Like Slightly	Like Moderately	Like Very Much	Like Extremely
714									
098									
511									

If **Yes**, please answer the following question (How would you rated the Bitterness)
If **No**, please skip the following question

APPENDIX D SETS OF QUESTIONAIRES

<u>Set 1</u> **Gender** Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	Dislike extremely 1	Dislike very much 2	Dislike moderately 3	Dislike slightly 4	Neither like nor dislike 5	Like slightly 6	Like moderately 7	Like very much 8	Like extremely 9
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

2. Please mark (✍) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	Dislike extremely 1	Neither like nor dislike 5	Like extremely 9
147	|—————	——————	—————|
705	|—————	——————	—————|

3. Please mark (✍) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	Greatest imaginable Dislike 0	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike 50	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like 100
278	|——	|——	|——	|——	|——	|——	|——	|——	|——	|——	|
665	|——	|——	|——	|——	|——	|——	|——	|——	|——	|——	|

Gender Female ( )   Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

Sample	Dislike extremely 1	2	Dislike moderately 3	4	Neither like nor dislike 5
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	Dislike extremely 1	Dislike moderately 3	Neither like nor dislike 5
247			
805			

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	Greatest imaginable Dislike 0	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike 50
378						
765						

<u>Set 2</u> **Gender** Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
Sample	1	2	3	4	5	6	7	8	9
511	( )	( )	( )	( )	( )	( )	( )	( )	( )
098	( )	( )	( )	( )	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	5	9
705			
147			

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
Sample	0					50					100
665											
278											

Gender Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
611	( )	( )	( )	( )	( )
198	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
805			
247			

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
765						
378						

Set 3 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely 1	Dislike very much 2	Dislike moderately 3	Dislike slightly 4	Neither like nor dislike 5	Like slightly 6	Like moderately 7	Like very much 8	Like extremely 9
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike 0	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike 50	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like 100
278											
665											

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely 1	Neither like nor dislike 5	Like extremely 9
147			
705			

Gender Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

Sample	1 Dislike extremely	2	3 Dislike moderately	4	5 Neither like nor dislike
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	0 Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	50 Neither like nor dislike
378	|————	|————	|————	|————	|————	|
765	|————	|————	|————	|————	|————	|

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	1 Dislike extremely	3 Dislike moderately	5 Neither like nor dislike
247	|————	————	|
805	|————	————	|

Set 4 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
Sample	1	2	3	4	5	6	7	8	9
511	( )	( )	( )	( )	( )	( )	( )	( )	( )
098	( )	( )	( )	( )	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Greatest imaginable Dislike | Dislike extremely | Dislike very much | Dislike moderately | Dislike slightly | Neither like nor dislike | Like slightly | Like moderately | Like very much | Like extremely | Greatest imaginable Like

Sample 0 50 100

665 |—|—|—|—|—|—|—|—|—|—|

278 |—|—|—|—|—|—|—|—|—|—|

3. Please mark (✐) along the line the point that best reflects yo **Bitterness perception** of the products.

Dislike extremely | Neither like nor dislike | Like extremely

Sample 1 5 9

705 |————————————|

147 |————————————|

Gender Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

Sample	1 Dislike extremely	2	3 Dislike moderately	4	5 Neither like nor dislike
611	( )	( )	( )	( )	( )
198	( )	( )	( )	( )	( )

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	0 Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	50 Neither like nor dislike
765						
378						

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	1 Dislike extremely	3 Dislike moderately	5 Neither like nor dislike
805			
247			

<u>Set 5</u> **Gender** Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Neither like nor dislike	Like extremely
Sample	1	5	9
147			
705			

2. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
Sample	1	2	3	4	5	6	7	8	9
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
Sample	0					50					100
278											
665											

Gender Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
247			
805			

2. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
378						
765						

<u>Set 6</u> **Gender** Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Neither like nor dislike	Like extremely
Sample	1	5	9
705	├	─	┤
147	├	─	┤

2. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
Sample	1	2	3	4	5	6	7	8	9
511	( )	( )	( )	( )	( )	( )	( )	( )	( )
098	( )	( )	( )	( )	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
Sample	0					50					100
665											
278											

Gender Female ( )   Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
805	|————	————	————|
247	|————	————	————|

2. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
611	( )	( )	( )	( )	( )
198	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
765	|————	+————	+————	+————	+————	————|
378	|————	+————	+————	+————	+————	————|

<u>Set 7</u> **Gender** Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	1 Dislike extremely	5 Neither like nor dislike	9 Like extremely
147			
705			

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	0 Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	50 Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	100 Greatest imaginable Like
278											
665											

3. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	1 Dislike extremely	2 Dislike very much	3 Dislike moderately	4 Dislike slightly	5 Neither like nor dislike	6 Like slightly	7 Like moderately	8 Like very much	9 Like extremely
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

Gender Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	1 Dislike extremely	3 Dislike moderately	5 Neither like nor dislike
247			
805			

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

Sample	0 Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	50 Neither like nor dislike
378						
765						

3. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

Sample	1 Dislike extremely	2	3 Dislike moderately	4	5 Neither like nor dislike
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

Set 8 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely	Neither like nor dislike	Like extremely
	1	5	9
147			
705			

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
	0					50					100
278											
665											

3. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
	1	2	3	4	5	6	7	8	9
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

Gender Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
247	├	──	┤
805	├	──	┤

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
378	├	┼	┼	┼	┼	┤
765	├	┼	┼	┼	┼	┤

3. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

Set 9 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	0 Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	50 Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	100 Greatest imaginable Like
278											
665											

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	1 Dislike extremely	5 Neither like nor dislike	9 Like extremely
147			
705			

3. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

Sample	1 Dislike extremely	2 Dislike very much	3 Dislike moderately	4 Dislike slightly	5 Neither like nor dislike	6 Like slightly	7 Like moderately	8 Like very much	9 Like extremely
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

Gender Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
	0					50
378						
765						

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely	Dislike moderately	Neither like nor dislike
	1	3	5
247			
805			

3. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

Sample	Dislike extremely		Dislike moderately		Neither like nor dislike
	1	2	3	4	5
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

Set 10 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
	0					50					100
665											
278											

2. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely	Neither like nor dislike	Like extremely
	1	5	9
705			
147			

3. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
	1	2	3	4	5	6	7	8	9
511	( )	( )	( )	( )	( )	( )	( )	( )	( )
098	( )	( )	( )	( )	( )	( )	( )	( )	( )

Gender Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
765						
378						

2. Please mark (✐) along the line the point that best reflects your **<u>Bitterness perception</u>** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
805			
247			

3. Please check the parenthesis (√) that best reflects your **<u>Bitterness perception</u>** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
611	( )	( )	( )	( )	( )
198	( )	( )	( )	( )	( )

Set 11 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like
Sample	0					50					100
278											
665											

2. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely
Sample	1	2	3	4	5	6	7	8	9
098	( )	( )	( )	( )	( )	( )	( )	( )	( )
511	( )	( )	( )	( )	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Neither like nor dislike	Like extremely
Sample	1	5	9
147			
705			

Gender Female ( )   Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Greatest imaginable Dislike	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike
Sample	0					50
378						
765						

2. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

	Dislike extremely		Dislike moderately		Neither like nor dislike
Sample	1	2	3	4	5
198	( )	( )	( )	( )	( )
611	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

	Dislike extremely	Dislike moderately	Neither like nor dislike
Sample	1	3	5
247			
805			

Set 12 **Gender** Female ( ) Male ( )

Instruction

Please evaluate **Bitterness Perception** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike 0	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike 50	Like slightly	Like moderately	Like very much	Like extremely	Greatest imaginable Like 100
665											
278											

2. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely 1	Dislike very much 2	Dislike moderately 3	Dislike slightly 4	Neither like nor dislike 5	Like slightly 6	Like moderately 7	Like very much 8	Like extremely 9
511	( )	( )	( )	( )	( )	( )	( )	( )	( )
098	( )	( )	( )	( )	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely 1	Neither like nor dislike 5	Like extremely 9
705			
147			

Gender Female ( ) Male ( )

Instruction

Please evaluate **<u>Bitterness Perception</u>** of the products. **Between the samples, you are required to drink water and eat unsalted cracker to clean your palate!!!**

1. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Greatest imaginable Dislike 0	Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike 50
765						
378						

2. Please check the parenthesis (√) that best reflects your **Bitterness perception** of the products

Sample	Dislike extremely 1	2	Dislike moderately 3	4	Neither like nor dislike 5
611	( )	( )	( )	( )	( )
198	( )	( )	( )	( )	( )

3. Please mark (✐) along the line the point that best reflects your **Bitterness perception** of the products.

Sample	Dislike extremely 1	Dislike moderately 3	Neither like nor dislike 5
805			
247			

VITA

Wannita Jirangrat is originally from Bangkok, Thailand and. She was born on June 7th. She completed her bachelor's and master's degree from Kasetsart University. After receiving her master's degree, she worked as a researcher for 2 years and enjoyed teaching product development. She found herself interested in sensory and has decided to continue her PhD. Her credentials are in the area of sensory science, product development, and statistical analysis as well as being well versed in relevant computer software. On the third year of her doctoral program, she joined a 6 months sensory internship program at Kellogg Company where she has extensive experiences in descriptive analysis, discriminative testing (trained and assisted in execution as well as data interpretation), quantitative and qualitative consumer testing (questionnaire creation, data analysis and interpretation) and shelf life evaluation of beverages, crackers, bars and cereals. She is now a PhD candidate from the department of Food Science at Louisiana State University Agricultural and Mechanical College who will be awarded a doctoral degree in May 2013.