8-2014

Comparison of Triangle and Tetrad Discrimination Methodology in Applied, Industrial Manner

Sara Lyn Carlisle
University of Tennessee - Knoxville, scarlisl@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Sara Lyn Carlisle entitled "Comparison of Triangle and Tetrad Discrimination Methodology in Applied, Industrial Manner." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Food Science and Technology.

Marjorie P. Penfield, Major Professor

We have read this thesis and recommend its acceptance:

David A. Golden, Arnold M. Saxton

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Comparison of Triangle and Tetrad Discrimination Methodology in an Applied, Industrial Manner

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Sara Lyn Carlisle
August 2014

Acknowledgements

I would like to express much gratitude to my instructors, family, and friends for their guidance and support throughout my academic career. I especially wish to thank Dr. Marjorie Penfield for introducing me to the field of sensory science and providing me with ample opportunities to learn and grow both as a person and as a sensory scientist over the last three years. Thank you to Dr. Arnold Saxton for serving on my committee and providing much appreciated statistical support during my research. I would also like to thank Dr. David Golden for his confidence in me both as an instructor and as a member of my committee.

I would also like to thank the many Sensory Lab workers for their assistance in panel preparation and serving panelists. Thank you to all of the many panelists who volunteered to take part in this study. Their cooperation and willingness to participate is very much appreciated. Much gratitude is extended to all of the companies who expressed interest in and provided products for this research. Without their support this project would not have been possible.

Special thanks are reserved for my parents, Jane and Terry Carlisle for their unwavering love and confidence in me. I also thank my brothers, Dylan and Dakota Carlisle, for their love and encouragement. Thank you to all my friends and family for their willingness to act as a sounding board and unending support that made this study possible.

ABSTRACT

The triangle method has been widely used in the food industry for many years when conducting sensory discrimination testing. Recently, however, another discrimination testing method, the tetrad, has begun to gain popularity. Based on currently published research, the tetrad method possesses statistical advantages over the triangle and would require fewer panelists, reduce testing time, and use less sample material. More testing is needed to confirm these advantages in an applied, industrial approach on a wider range of products. Over thirty triangles and thirty tetrads with untrained panelists were completed in order to compare the two methods. Products tested ranged from canned vegetables and fresh fruits to deli meats and baked goods. Panels conducted thus far have provided contradictory results. Inconsistencies were found within and across product categories. Significant differences were seen with the triangle method but not in the tetrad in a few cases. In one specific instance, the same products were tested alone and then again with a carrier. Panelists were able to perceive the difference between the products with both methods when the product was served alone but were unable to do so when a carrier was present with the tetrad. Effect size and test power for each test were also calculated and produced similar results. In eight of the experiments completed, the reduction in effect size for the tetrad offset the statistical power advantage, making the triangle method more beneficial for these products. Significant differences ($p < 0.05$) were found between methods when the degree of difference was measured

between samples for each test with a larger difference found using the triangle in a few cases. Participating panelists were also asked to compare the two methods in terms of difficulty on a structured scale and in an open-ended fashion. Overall, panelists perceived the two methods as very similar in terms of method difficulty with very little mean separation between experiments. Panelists noted that the product being tested affected their impression of the tests in multiple experiments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Discrimination testing is a type of sensory testing used to determine if a difference exists between products and is used in an array of situations. When an ingredient in a product needs to be replaced, new equipment has been installed, or deviations from usual protocol during production have occurred, discrimination testing can be used to determine if the final product has been noticeably affected. The type of discrimination test to be used can depend on a number of factors like the complexity of the product, test sensitivity, and panelists to be used. The triangle and tetrad are common discrimination testing methods used in industry. Recently, the tetrad method has been receiving praise as a more sensitive testing method than the more traditional triangle that could save companies money by reducing the number of panelists and amount of samples required (ASTM 2011; Ennis 2013). Even some consulting firms who perform sensory testing have begun to advertise tetrad testing on their websites (Food Safety International Network 2012; Leatherhead Food Research 2014; Sensory Dimensions 2013).

Many concerns surrounding the tetrad methodology have been presented in literature. The addition of a fourth sample could lead to panelist fatigue and a reduction in sensitivity to the stimulus (Ennis 2012). Products with strong seasonings, spice heat, or lingering flavors may overpower panelist memory and have too much carryover between samples to make the tetrad method effective

(Ennis 2012). Unlike the p-value, which can be easily calculated, d′ value tables are not widely available and only provide an approximate value since rounding to the nearest proportion correct is often required. There have also been disagreements among sources that are currently available. Delwiche and O'Mahony (1996) found the specified method, in which a specific attribute, like sweetness or bitterness, is addressed, more statistically advantageous than the unspecified, while Masuoka (1995) found no difference in the two. Very few direct comparisons have been conducted between the triangle and tetrad methods. O'Mahony (2013), using Delwiche and O'Mahony's data (1996), did so with conflicting results. When looking at Yip's (1996) thesis work, he found the tetrad methodology to have a lower d′ than the triangle, indicating that the tetrad was theoretically more powerful than the former. The purpose of this study is to address these concerns in an applied, industrial approach using existing protocol to compare triangle and tetrad test results. These tests were completed in a single session to determine if differences exist. Qualitative data were also gathered to gain insight on panelist perception of the testing methods, which could be helpful to companies trying to decide whether to make the switch from triangle to tetrad or not.

CHAPTER II

Review of Literature

Sensory testing is used to collect information on the properties of wide ranges of products in order to improve the product, maintain a certain level of quality, gauge current market reactions, and aids the research and development process of food companies (Amerine *et al.* 1965). Sensorial characteristics experienced in foods can arise from and be affected by a number of different factors like the genetic makeup of the product, agricultural influences, pre and post-mortem handling conditions, processing methodology, packaging and storage practices, and quality standards in place (Amerine *et al.* 1965). Knowledge of these properties along with sensory testing can help companies to fully understand their products and assess consumer response. Different types of sensory tests are used depending on the information wished to be gained from the experiment. There are three main categories of sensory testing: affective, descriptive, and discrimination. All three serve a different purpose and provide companies with different information and answer different questions. This information can be used to help minimize the risk in making business decisions.

Sensory Test Methods

Affective Testing

Affective testing is also known as consumer acceptance or preference testing. Affective tests may be performed after discrimination testing when a

statistically significant difference has been established in the products. This type of testing can be used for a number of purposes like determining which product consumers will react most favorably toward. This type of testing answers preference questions by measuring panelists' degree of liking for products (Lawless and Heymann 1998).

Descriptive Testing

Descriptive testing can answer a variety of questions and is most often utilized during the initial stages of product development to gauge consumer desires or how similar a new product is to an ideal. This type of testing often produces objective, qualitative descriptions of product attributes and most often involves trained panels. Depending on the questions asked and the way the data are interpreted, some quantitative information can also be learned from this type of testing. Quantitative data can be gained from this type of testing when measuring perceived attribute intensities by panelists (Lawless and Heymann 1998).

Discrimination Testing

Discrimination testing is used to determine whether consumers can perceive products as different. This specific type of testing is often employed when an ingredient substitution is needed or a change in processing has been made. For this type of testing, the null hypothesis is that “the products are not different” when testing for a difference or “the products are the same” when

testing for similarity. After testing, it can be determined whether the null hypothesis should be accepted or rejected using predetermined statistical significance levels (p-value). Discrimination testing works best with two products that differ slightly. If the products vary greatly and a difference is known, discrimination testing may not be the best option.

Sensory scientists have many different discrimination methodologies at their disposal including alternative forced choice (AFC), 2-out-of-5, duo-trio, triangle, and tetrad. All of these tests, though executed differently, can be used to determine if panelists perceive a difference in samples. This study will focus on discrimination testing methods, specifically triangle and tetrad tests.

In triangle testing, the subject is simultaneously presented with three samples. Of these three, two of the samples are alike and one is different or “odd”. The subject is asked to identify the odd sample in an unspecified triangle. A specified triangle test would ask the panelists to choose the different sample based on the differing attribute, like sweetness or bitterness. The samples must be presented in six different arrangements (AAB, ABA, BAA, BBA, BAB, and ABB) to prevent psychological errors in judgment. The probability of correctly guessing the odd sample by chance is 1/3.

After testing, a p-value can be calculated using binomial distribution (Lawless and Heymann 1998) to determine if a difference does exist. For most

difference testing, the predetermined statistical significance level is usually set at $p = 0.05$. If the resulting p-value is less than 0.05, the null hypothesis is rejected and it can be assumed that the products are in fact different from one another. When testing for similarity, the p-value is set at $p = 0.10$.

In tetrad testing, subjects are given four samples- two from one group and two from another group with six different arrangements (AABB, ABAB, ABBA, BBAA, BABA, and BAAB). Subjects are then instructed to sort the samples into two groups of two based on similarity using either a specified or unspecified approach. The tetrad method employs the same type of statistical modeling and results as the triangle test but is said to do so with fewer panelists because of increased testing power (Ennis 2012). For difference tests, testing power can be explained as the probability of correctly finding a difference. Power for a test depends on a number of variables: the effect size (δ), the chosen alpha (0.05), and number of panelists used (Lawless and Heymann 1998).

Like the triangle method, the unspecified tetrad test has a guessing probability of 1/3, but the tetrad design offers a few theoretical advantages over the former. These advantages are fueling the push to replace the triangle with the tetrad. Theoretically, switching from the triangle method to tetrad should result in a decrease of 1/3 in effect size and 50% increase in perceptual noise (Ennis 2012) explained using the Thurstonian theory.

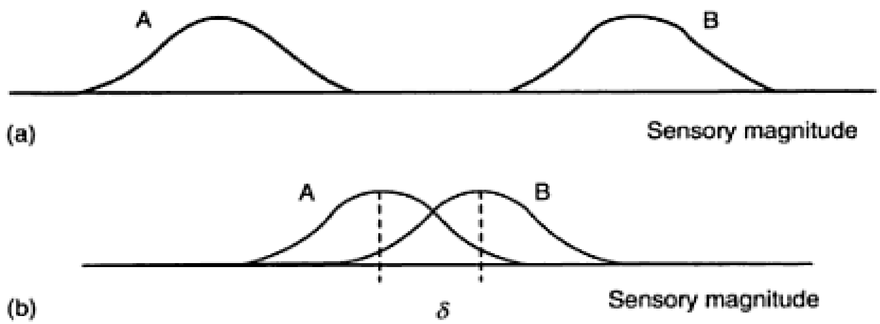

Figure 1 – Representation of Thurstonian discriminal differences (Meilgaard *et al.* 2006).

Effect size can be thought of as a ratio between signal, or the perceived difference intensities, and noise with noise equaling one (Garcia *et al.* 2013). Effect size is an estimate of the amount of perceived difference between samples. Figure 1 depicts this theory by showing the effect size of two products. The difference in sensory magnitude of (a) is larger because the two products are very different, corresponding to a larger δ. The two products shown in (b) are much more similar and therefore have a smaller δ. After testing products, the test statistic d′ is used to estimate the effect size (δ) seen in the experiment (Meilgaard *et al.* 2006). In order for the tetrad to be more powerful than the triangle, the perceptual noise increase between the two tests must be less than or equal to 50% and effect size cannot decrease by more than 1/3 (Ennis 2012). The d′ value for the tetrad should theoretically always decrease because the addition of the fourth sample in the method inherently adds noise to the test.

The tetrad test has been gaining great popularity in recent years with the development of more exact analysis tables. The tetrad's increased test sensitivity could lead to a much lower number of panelists and amount of sample required to test, saving many companies money. Even major international companies like General Mills have decided to convert from the more traditional triangle discrimination method to the tetrad as a way to reduce cost associated with testing (Gelski 2013). Tetradic principles have been the focus of five recent studies using a variety of testing mediums. Garcia *et al.* (2012) utilized a large group of children to compare the tetrad and triangle methods using apple juice. Delwiche and O'Mahony (1996) used chocolate pudding as a medium when comparing discrimination tests. Masuoka *et al.* (1995) compared triad and tetrad discrimination methods using beer bitterness sensitivity. A fourth study by Yip (1996) involved a study of NaCl thresholds in purified water. Ishii *et al.* (2014) compared triangle and tetrad methods using dilutions of orange and apple juices.

All of the experiments utilized Thurstonian discriminal difference modeling. The Garcia *et al.* (2012) and Ishii *et al.* (2014) studies, however, were the only experiments that directly compared the triangle and tetrad test. The latter three researchers used the Thurstonian hypothesis to compare 3-AFC and triangle performances. The d′ values confirmed that the 3-AFC out-performed the triangle when the same judges and mediums were used. Specified and unspecified tetrads were also evaluated. These experiments found that the two versions of

the test were not significantly different despite the differences in chance probabilities, 1/6 for specified and 1/3 for unspecified.

Garcia *et al.* (2012) conducted an experiment with 404 elementary school children using pure and diluted apple juice to compare tetrad and triangle tests. The children performed one tetrad and two triangles per session with example demonstrations before each session. After testing, in addition to p-values, the effect size and its variance for the products in each test type was determined using Thurstonian modeling. The estimates for effect size were expressed as d′ values using tables provided by Ennis (1993) and Ennis *et al.* (1998) while variance values were obtained from Bi *et al.* (1997) and Ennis (2012). After analysis, it was found that the tetrad produced a higher proportion correct than the triangle. The d′ values for the triangle tests (1.41) were higher than the tetrad (1.18). The same trend was noted with the variances of d′ values. This difference, however, was not significant with a p-value of 0.07. The effect size was also reduced by 16% but remained higher than 2/3 when compared to the triangle with a value of 0.837. Overall, the conclusion was made that the tetrad test was more statistically powerful than the triangle.

Delwiche and O’Mahony (1996) compared the triangle, 3-AFC, and both the specified and unspecified tetrad methods using instant pudding and pie filling mixes. In this trial, plain and sweetened “target” chocolate pudding samples were used along with several flavor-added pudding samples as distractors. Thirteen

panelists completed twelve blocks of ten tests with four triads and six tetrads being done in a single block. Subjects were given the option of completing up to two blocks within a single session. In this experimental trial, the 3-AFC performed better than the triangle method ($p= 0.0005$), but the performance of neither was compared to the tetrad. Instead the specified and unspecified tetrads were compared to each other. Delwiche and O’Mahony (1996) found that the unspecified method performed worse than the specified ($p<0.015$).

O’Mahony (2013) used the information form this study to compare d′ values for the triangle and unspecified tetrad since d′ tables were not available at the time of the experiment. When doing so, the tetrad provided a smaller d′ value (2.18) than the triangle (2.36). These values, however, were not significant ($p=0.49$) but produced variances in d′ with the same pattern (0.022 for tetrad and 0.044 for triangle). O’Mahony (2013) did find a significant difference ($p=0.007$) in the d′ values for the specified and unspecified tetrad methods (1.64 versus 2.18).

Masuoka *et al.* (1995) conducted a study using beer with various bitterness levels in a two-part experiment to compare triangle and 3-AFC methods as well as specified and unspecified tetrad methods. For both parts of the study, nine judges with 12 replications were used. Specific bitterness levels were determined for each individual panelist prior to both experiments based on their sensitivities to bitterness. Each experimental part was performed over eight sessions with six tests per session, three of which were distractor tests. In this

study, no statistical difference was found when comparing the specified and unspecified tetrad tests.

Like in the Delwiche and O'Mahony (1996) experiment, O'Mahony (2013) took another look at this experiment, comparing d′ values for the triangle and unspecified tetrad tests. Although the values proved to not be significant, it is interesting to note that unlike the Garcia *et al.* (2012) study, the triangle produced a lower d′ (1.26 compared to 1.43). The variances in d′ for the two test types were more in line with previous predictions with 0.02 for the tetrad and 0.07 for the triangle. Because these experiments occurred in two parts with different judges, it is difficult to determine the validity of these values (O'Mahony 2013).

In a thesis study performed by Yip (1996), specified and unspecified tetrad tests along with triangles and 3-AFCs were compared using 26 panelists with NaCl solutions and purified water. This experiment also included distractor tests. Panelists performed 12 of each type of testing and 24 distractor tests in all with each test type being performed in a separate session. Using d′ tables (Ennis *et al.* 1998), Yip (1996) was able to compare the triangle to the unspecified tetrad. This experiment did produce significant differences ($p=0.005$) for the d′ values between the two types (1.66 for the triangle and 1.17 for the tetrad), suggesting that the tetrad was more powerful (O'Mahony 2013).

Ishii *et al.* (2014) tested the efficiency of the triangle versus the tetrad using very small dilutions of orange and apple juices and the effect of resampling on the end results. Each of the 456 panelists completed four tests in a single session with retasting being allowed in two. Panelists performed all tests in a one on one fashion and gave their responses verbally to an interviewer. In all four scenarios, the tetrad received a higher proportion correct than the triangle. Triangle and tetrad test values were not found to be significantly different, but it was concluded that the tetrad was still a more powerful method since the d′ did not decrease by more than 1/3. The tests in which panelists were allowed to retaste produced larger d′ values (0.90, 0.84, 1.35, and 1.14) than single tasting sessions (0, 0.44, 1.02, and 0.91). This verified that retasting allowed for better discrimination by panelists.

There are still some concerns surrounding the methodology, however. Traditionally, tests with more than three samples have been viewed as an option with very limited use. The psychological strain and fatigue associated with an increase in the number of samples presented have steered sensory scientists away from these tests and to simpler tests with fewer samples. Multiple sample tests have been reserved for visual difference testing. Taking away the tasting component reduces concerns like fatigue and memory effects (Amerine *et al.* 1965).The addition of the fourth stimulus could lead to fatigue, a drop in practical sensitivity from adaptation, and a change in cognitive strategies (Lawless and Heymann 1998) used by panelists. Because of these factors a small drop in

effect size is expected when using the tetrad. As long as the effect size drops by no more than 1/3 and the perceptual noise increases by no more than 50%, though, the tetrad should still be statistically more powerful than the triangle or 3-AFC methods (Ennis 2012).

There are also some questions surrounding the instructional approach that should be taken when administering the tetrad. The way in which the instructions are presented to panelists can affect the statistical significance of the test. For example, in the specified tetrad, the panelist would be instructed to find the two sweet samples. In the unspecified tetrad, panelists would be told to find the two most similar samples. The specified tetrad, in which the difference between products was stated, has a chance probability of 1/6 while the unspecified tetrad, where no specific difference is stated, has a chance probability of 1/3 (O'Mahony 2013). In some experimental trials it was found that neither the specified or unspecified instructional method had an advantage over the other (Masuoka *et al.* 1995). More recently, however, it has been found that the unspecified tetrad is more powerful than the specified (Rousseau and Ennis 2013).

Based on recent research (Ennis 2012; Garcia *et al.* 2012; Masuoka *et al.* 1995; O'Mahony 2013), the tetradic method should produce the same results as the triangle with fewer panelists because of increased testing power. If this is true, switching from the triangle to the tetrad could reduce testing time, number of panelists needed, and sample required for testing. Although a lot of experimental

research has been done to confirm the theoretical advantages of the tetrad method, not any practical industry style panels have been run.

In one study, Masuoka *et al.* (1995), specific levels of the differing attribute were chosen for each panelist. This might be useful when comparing an individual panelist's performance but is not very practical in a large-scale difference testing operation. Garcia *et al.* (2012) utilized 404 children as panelists and 456 panelists participated in the Ishii *et al.* (2014) study. The extremely large number of panelists used in both studies is not very feasible in ordinary industrial application. Other than this, most of the studies presented have been conducted with a small number of panelists and replicated to provide a larger base. More trials completed in an industry manner are needed to prove tetrad′s statistical advantage over triangle methodology as it applies to the food industry.

Another question that has yet to be answered is, "For which products is the tetrad methodology applicable?" A more sensitive test, like the tetrad, may not always be the best option for all products (O'Mahony and Rousseau 2002, Ishii *et al.* 2014). For products that are simple in nature, the tetrad has been shown to be more statistically powerful that the more traditional triangle test. Ennis (2012) notes that tests involving products with lingering sensorial characteristics like fragrances and high spice levels, as well as products containing alcohol or tobacco, may not be suitable for tetrad testing. The addition of a fourth stimulus hinders the subjects' ability to evaluate these more complex

products. Many factors like individual panelist decision criteria, dimensionality or nature of the sensory difference, and the amount of re-tasting allowed could affect the outcome of the test (Lawless 2013). In this experiment, a wide array of products was tested in order to gauge which products may be considered too complex for tetrad testing.

Experienced versus inexperienced panelists differences may influence method results. Many companies use already established panels for sensory testing. These panelists are very familiar with current testing methods like the triangle. It is well known that familiarity with method improves the ability to discriminate. If panelists are less comfortable with the methodology, the findings could be negatively affected. Learning experienced panelists' perceptions of the tetrad method versus that of the less experienced panelist when compared to the triangle could be very helpful for industrial companies when trying to decide whether to make the switch from triangle to tetrad or not. Learning panelists' perceptions of the two methodologies in general, could add a lot of insight to the manner in which they approach the test.

CHAPTER III

Materials and Methods

All experiments in this study were conducted in the University of Tennessee at Knoxville Sensory Lab in individualized booths using FIZZ by Biosystemes (2009) computer programming. Samples were presented in balanced orders using randomly chosen 3-digit codes.

Panelists

Panelists participating in this study were recruited using the University of Tennessee at Knoxville Sensory Lab email database, which includes roughly 400 university staff, faculty, and students. All panelists received an email announcing the test type, number of panelists required, product to be tested, and a list of potential allergens prior to the testing date. To participate, panelists must be 18 years or older and willing to taste the product. Prior to each test, panelists were asked to sign a consent and confidentiality form and were then given another brief description of the products and testing method. This study was certified for exemption by IRB review for research involving human subjects.

After completing each test, panelists were asked to record their gender and age within a range. Most participants were University of Tennessee staff, faculty, and graduates students. For the most part, the age ranges presented reflect this with about 63% of the panelists being 18-34 years of age. The

exceptions include the smaller apple juice and applesauce experiments as strictly naïve panelists were recruited, accounting for the pronounced skew in the distributions. Panelists who were considered naïve had no experience participating in either triangle or tetrad testing prior to participating in the study. The ratio of male to female participants fluctuated slightly over the course of the study, but on average, close to 30% of the participants was male while the remaining 70% was female. These distributions can be seen in Appendix A.

Products

In order to simulate the practical application of the tetrad method in industry, this research was conducted in the same manner companies would approach difference testing. Because of this, testing was done in a single test with a variety of products to better encompass the many facets of the food industry. Products tested included canned vegetables, carbonated beverages, fruit juices, dairy products, food colorings, fruit and vegetable sauces, fresh fruits, cereals and crackers, and sliced lunchmeats. Table 1 contains descriptions of control and test samples for each experiment. Product names used in later results tables can also be found in the table following the control description. To maintain the proprietary nature of the data for companies that provided products, specific brands names are not mentioned. Materials used in the experiments that were not provided by companies were purchased from local supermarkets.

Table 1. PRODUCT DESCRIPTIONS FOR TETRAD AND TRIANGLE TESTS

Product	Control Description	Test Description
Black beans	Commercially processed black beans in brine	Commercially processed black beans with added seasoning
Kidney beans	Commercially processed dark red kidney beans in brine	Commercially processed dark red kidney beans in brine with change made during processing step
Chili Beans	Commercially processed pinto beans in a hot chili sauce with garlic, onion, and other spices.	Commercially processed pinto beans in a mild chili sauce with garlic, onion, and other spices.
Pinto beans	Commercially processed pinto beans in brine	Commercially processed seasoned and regular pinto beans mix: 3 cans seasoned to 1 can regular
Baked beans	Commercially processed navy beans in a thick brown sugar sauce with bacon (Baked beans BB)[a]	Commercially processed navy beans in a thick brown sugar sauce with bacon and reduced pork flavoring
	Commercially processed navy beans in a smoky sauce with brown sugar (BB Smoky 1)	Commercially processed navy beans in a smoky sauce with brown sugar and level 1 reduced pork flavor
	Commercially processed navy beans in a smoky sauce with brown sugar (BB Smoky 2)	Commercially processed navy beans in a smoky sauce with brown sugar and level 2 reduced pork flavor
	Commercially processed navy beans in a tomato and brown sugar sauce with bacon and spices (BB Veg)	Commercially processed navy beans in a vegetarian tomato and brown sugar sauce with spices
	Brand A commercially processed navy beans in a brown sugar and molasses sauce with bacon and spices (BB Molasses)	Brand B commercially processed navy beans in a brown sugar and molasses sauce with bacon and spices
	Commercially processed navy beans in a brown sugar sauce with bacon and spices (BB + liquid smoke)	Commercially processed navy beans in a brown sugar sauce with bacon and spices and 0.6 g of liquid smoke added per 28 oz can
	Commercially processed navy beans in a brown sugar sauce with bacon and spices (BB + brown sugar)	Commercially processed navy beans in a brown sugar sauce with bacon and spices and 36 g of dark brown sugar added per 28 oz can
	Commercially processed navy beans in a brown sugar sauce with bacon and spices (BB + BBQ sauce)	Commercially processed navy beans in a brown sugar sauce with bacon and spices and 30 g of BBQ sauce added per 28 oz can

[a] Terms in parentheses following control description correspond to product names in results tables.

Table 1 (continued).

Product	Control Description	Test Description
Lemon-Lime soda	Regular lemon-lime flavored soda in 12 oz aluminum can	Zero Calorie lemon-lime flavored soda in 12 oz aluminum can
Apple Juice	Store brand 100% apple juice in plastic gallon container	Store brand 100% apple juice in plastic gallon container diluted 25% by volume with spring water
Milk	Reduced fat (2%) light-oxidized milk in plastic gallon jug	Reduced fat (2%) non-oxidized milk in plastic gallon jug
Milk with coloring	Reduced fat (2%) milk with 10 mL annatto cheese coloring per gallon milk (Milk with color 1)	Reduced fat (2%) milk with 8 mL annatto cheese coloring per gallon milk
	Reduced fat (2%) milk with 10 mL annatto cheese coloring per gallon milk (Milk with color 2)	Reduced fat (2%) milk with 7.5 mL annatto cheese coloring per gallon milk
Applesauce	Store brand regular applesauce (Applesauce 1)	Store brand regular and no sugar added applesauce mix: 70% regular to 30% no sugar
	Store brand regular applesauce (Applesauce 2)	Store brand regular and no sugar added applesauce mix: 80% regular to 20% no sugar
Tomato Sauce	Traditional tomato sauce in 26 oz glass jar	Traditional tomato sauce in 42 oz plastic jar
	Traditional tomato sauce in 26 oz glass jar	Traditional tomato sauce in 42 oz plastic jar, cubes of white bread used as carrier
Cantaloupe	Fresh cantaloupe of variety A cubed to uniform size	Fresh cantaloupe of variety B cubed to uniform size
Cheese crackers	Cheddar flavored baked snack crackers	Reduced fat cheddar flavored baked snack crackers
Wheat crackers	Reduced fat whole grain snack crackers	Reduced salt whole grain snack crackers
Oat Cereal	Name brand toasted oats cereal	Store brand toasted oats cereal
Lunch meat	Thinly sliced oven roasted turkey lunch meat from Plant A individually folded into cup (Lunch meat 1)	Thinly sliced oven roasted turkey lunch meat from Plant B individually folded into cup
	Thinly sliced oven roasted turkey lunch meat from Plant A individually laid flat on plate (Lunch meat 2)	Thinly sliced oven roasted turkey lunch meat from Plant B individually laid flat on plate

Specific protocols followed when performing experiments for each product category are presented in Table 2. Tests where samples could be prepared in advanced occurred on the same day in a balanced fashion, half of the panelists received the tetrad first and half received the triangle first. If serving was time dependent, such as carbonated beverages and milk, or heating was required immediately before serving, testing occurred over two days with the triangle method occurring on the first day. All experiments, with the exception of the BB Molasses and Oat cereal, were conducted with white fluorescent lighting in each booth. The BB Molasses and Oat cereal experiments utilized red lighting in each booth to minimize obvious visual differences.

Test Instructions

Panelists were asked to taste samples from left to right in both the tetrad and triangle tests. For the triangle tests, panelists were asked to "Indicate which sample is the odd (different) sample by checking the box next to the appropriate code number." For tetrad tests, instructions given were as follows: "Sort the samples into two groups of two. Check the sample codes from ONE of your groups." Re-tasting was allowed in both tests. After completing each test, panelists were asked to rate the degree of difference they perceived on a 5-point interval scale ranging from "very slight" to "extremely large" difference. Panelists were also given the opportunity to mark "no difference". "No difference" choices were not considered in mean score calculations.

Table 2. PROTOCOL FOR TETRAD AND TRIANGLE TESTS BY PRODUCT CATEGORY

Product category	Sample preparation	Test order	Sample size	Container	Serving temperature
Canned beans	Cans opened and, if present, bacon removed before mixing thoroughly	Separate panels with triangle on day 1	47.3 mL	6-oz. white Corelle® rice bowls	Heated in 1100 watt microwave at 100% power (Triangle: 30 sec, Tetrad: 40 sec) immediately prior to serving
Carbonated beverages	Samples poured directly from cans and served immediately	Separate panels with triangle on day 1	45 Ml	3-oz white plastic Great Value™ cups	20-22ºC
Fruit juices	Prepared day before and stored in refrigerator overnight; samples stirred morning of prior to serving	Balanced design[a]	45 mL	3-oz white plastic Great Value™ cups	20-22ºC
Dairy products	Care taken to ensure minimal exposure to light prior to serving	Separate panels with triangle on day 1	30 mL	5-oz opaque plastic Great Value™ cups	2-4ºC
Visual Milk Test	Samples prepared and mixed thoroughly morning of test	Balanced design	20 mL	Standard shot glass	20-22ºC
Fruit and vegetable sauces	Mixed thoroughly	Balanced design	18.5 mL	2-oz opaque plastic Solo® cups	20-22ºC
Fresh fruits	Cubed day before and stored in lidded serving container in refrigerator overnight	Balanced design	2 cubes	2-oz opaque plastic Solo® cups with lids	2-4ºC
Cereals and crackers	Poured from box into large bowl where broken and/or burnt pieces removed	Balanced design	3-4 crackers; 2.2 g cereal	2-oz opaque plastic Solo® cups	20-22ºC
Lunch meats	Care taken to ensure minimal exposure to light prior to serving	Balanced design	2 slices	4-oz opaque plastic Solo® cups; 6-inch Styrofoam® plates	2-4ºC

[a] Balanced design: half received triangle test first, half received tetrad first.

Following the completion of both tests for experiments with a balanced design or the tetrad for single day tests, panelists were asked to compare the two methods in terms of difficulty level using a 5-point interval scale anchored by “much easier” and “much more difficult”. If panelists had not completed a triangle test before, they were asked to indicate so. After rating the difficulty level, panelists were able to explain their answer using their own words.

Data Analysis

Significance levels (p-values) were calculated between samples using FIZZ software (Biosystemes 2009) for triangle tests. As explained by Ennis (2012), guessing probability for both methods is 1/3. Therefore, the same principle for calculating p-values for triangle tests can be used for the unspecified tetrad tests used in this study. Since this option was not available in FIZZ, a discrimination test analysis tool provided by Carr Consulting (1998) was used to determine p-values for the tetrad in Microsoft Office Excel®. FIZZ was also used to collect degree of difference, difficulty level, age, and gender distributions. Data was exported from FIZZ to an Excel® file for each test for further analysis.

Standards for estimating discriminal differences have been published by ASTM (2009) with regard to the triangle method, but have yet to be published for tetrad (ASTM 2011). Thurstonian theory was used to estimate the effect size (d′) as well as the variance of d′ using tables provided in Ennis *et al.* (2011). Test power was calculated using an Excel® program provided by Teixeira *et al.* (2009)

and was based off a formula with $\alpha = 0.05$ and $\beta = 0.20$. Means and significance levels for degree of difference between samples were calculated in Microsoft Office Excel® using Student's t-test. SAS version 9.3 (SAS Institute 2011) was used to determine if significant differences existed between products for method difficulty levels using PROC GLIMMIX with the PDIFF option. Additional formulas and SAS code are provided in Appendix B.

CHAPTER IV

Results and Discussion

The analysis portion of this study consisted of comparing 31 tetrad tests and 31 triangle tests in a stepwise fashion. The significance level (p-value) for each discrimination test was first calculated to determine if a difference existed between samples. To determine the extent of that difference, effect sizes (d′) were then calculated using Thurstonian theory as described previously. To further compare the methods, power, degree of difference perceived, and ease of method were also determined. Panelist comments were also collected to qualitatively compare the methods.

Significance Levels (p-value)

The p-value is the likelihood of producing results as extreme, or more extreme, as the results observed in the test given the null hypothesis is true. For the purposes of this study, the null hypothesis is that the control and test samples for each product do not differ. The p-value is based on the proportion of correct responses with a significance level set at 0.05. Therefore, if a test's resulting p-value is less than 0.05, the null hypothesis can be rejected, and it is assumed that the two samples do differ. This calculation is widely used in industry to determine if a difference exists between existing and new or altered products. The p-values for the triangle and tetrad tests for this study can be found in Table 3.

Table 3. PROBABILITY OF DIFFERENCE RESULTS FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS

	Tetrad tests			Triangle tests		
Product	*N*	*Pc*[a]	*P-value*	*N*	*Pc*	*P-value*
Black beans	84	0.57	<0.001	78	0.55	<0.001
Kidney beans	60	0.33	0.548	54	0.30	0.762
Chili beans	54	0.57	<0.001	54	0.57	<0.001
Pinto beans	54	0.56	0.001	54	0.57	<0.001
Baked beans (BB)	60	0.42	0.110	60	0.45	0.040
BB Smoky 1	54	0.33	0.551	54	0.41	0.156
BB Smoky 2	54	0.48	0.017	54	0.52	0.004
BB Vegetarian	54	0.48	0.017	54	0.59	<0.001
BB Molasses	54	0.81	<0.001	54	0.76	<0.001
BB + liquid smoke	54	0.67	<0.001	54	0.56	0.001
BB + brown sugar	54	0.91	<0.001	54	0.69	<0.001
BB + BBQ sauce	54	0.83	<0.001	54	0.89	<0.001
Lemon-Lime soda	72	0.72	<0.001	72	0.61	<0.001
Apple juice	150	0.83	<0.001	150	0.68	<0.001
Apple juice[b]	29	0.86	<0.001	29	0.69	<0.001
Apple juice combined	179	0.83	<0.001	179	0.68	<0.001
Milk	72	0.50	0.001	72	0.65	<0.001
Milk with color 1	90	0.86	<0.001	90	0.59	<0.001
Milk with color 2	90	0.87	<0.001	90	0.77	<0.001
Applesauce 1	78	0.38	0.199	78	0.46	0.013
Applesauce 2	78	0.36	0.355	78	0.42	0.061
Applesauce 2[b]	31	0.68	<0.001	31	0.42	0.203
Applesauce 2 combined	109	0.45	0.008	109	0.42	0.033
Tomato sauce	72	0.53	0.001	72	0.46	0.018
Tomato sauce w/ carrier	72	0.40	0.131	72	0.51	0.001
Cantaloupe	54	0.54	0.002	54	0.50	0.008
Cheese crackers	78	0.44	0.038	78	0.42	0.061
Wheat crackers	78	0.83	<0.001	78	0.79	<0.001
Oat cereal	78	0.55	<0.001	78	0.63	<0.001
Lunch meat 1	78	0.54	<0.001	78	0.53	<0.001
Lunch meat 2	54	0.83	<0.001	54	0.52	0.004

[a]Pc: proportion correct = (N correct/ N total).

[b] Tests using naïve panelists only.

In most cases, the conclusions drawn from p-values for triangle and tetrad for the products agree. There are a few cases, however, where a difference was found in one test but not the other. A difference was found using the triangle but not the tetrad in Baked beans (BB), Applesauce 1, and Tomato sauce with carrier. Conversely, a difference was found with the tetrad but not the triangle method in the smaller Applesauce 2 experiment as well as the Cheese crackers experiment.

The samples used in the Baked beans (BB) experiment varied in pork flavor level. The complex nature of this product involved many different flavors and seasonings that may have proved too overwhelming for the four-sample test ($p = 0.110$) when compared to the triangle ($p = 0.040$). For the Applesauce 1 experiment, the samples differed in sweetness levels with the control being 100% regular applesauce while 30% of the test contained no sugar added applesauce. This product's flavor profile was much simpler than the Baked beans (BB). Again, a significant difference was found using the triangle ($p = 0.013$) and not the tetrad ($p = 0.199$).

The difference in the Tomato sauce with carrier results is especially of interest as both the tetrad ($p = 0.001$) and triangle ($p = 0.018$) results indicated a difference in the product in the previous Tomato sauce experiment with the same samples. The addition of white bread as a carrier added a level of complexity to the tests that proved to be a disadvantage in the tetrad ($p = 0.131$) but not in the

triangle ($p = 0.001$). In this case, the tetrad was more affected by sample presentation than the triangle.

The smaller Applesauce 2 experiment in which a difference was found in the tetrad ($p < 0.001$) but not the triangle ($p = 0.203$) only involved 31 panelists. This number is much lower than what would normally be used for difference testing. It is possible that because of reduced test power, the triangle required more than 31 panelists to find a difference. This would be consistent with findings by Ennis (2012). Another possibility for this outcome is the panelists participating in the experiment. This smaller group was comprised of solely naïve panelists with little to no experience with either method. When the same product was tested using a larger more experienced panelist base, different results were found. No significant differences were found with either test in the larger Applesauce 2 experiment, although the triangle p-value ($p = 0.061$) would be considered trending toward a difference while the tetrad test ($p = 0.355$) rendered no difference. The panelists' familiarity with the testing methodology could have affected the results.

Regular and reduced fat samples were used in the Cheese crackers experiment. The samples were significantly different when the tetrad method was employed ($p = 0.038$) but only trending towards a difference with the triangle method ($p = 0.061$). Unlike the overall flavor of the Baked beans (BB) product, this product's flavor was very simplistic. Since taster fatigue is not as much of a

concern with basic products, this could prove to be a situation in which the tetrad out-performs the triangle method. When Wheat crackers, a very similar product, were used, significant differences ($p < 0.001$) were found with both tests.

The tetrad method was expected to out-perform the triangle in both Milk with coloring experiments, but that was not seen in either experiment. Since visual differences are not as taxing on panelist memory load, it has been hypothesized that the tetrad method would be advantageous in this type of scenario based on information from Amerine *et al.* (1965). In both the triangle and tetrad, however, a significant difference ($p < 0.001$) was found in both experiments. Another visual variation between experiments was done with the lunchmeat experiments in which the same products were presented folded into a cup (Lunch meat 1) and laid flat on a plate (Lunch meat 2). No method disagreement was seen in this case either as a significant difference ($p < 0.05$) was found for the two methods in both experiments.

Effect Size (d′)

Where the p-value indicates if a difference exists, the estimated effect size (d′) indicates how different the samples were perceived using Thurstonian discriminal modeling. The effect size is a signal to noise ratio where the signal is the actual difference between samples and the noise is other distracting factors. This value can be estimated when testing using the d′ statistic (Meilgaard *et al.* 2006). Small d′ values correspond to small perceptual differences or large noise

in the samples. Large d′ values correspond to samples with a large signal, or perceptual difference. The addition of the fourth sample adds complexity to the tetrad method so the d′ values are inherently expected to be lower than the d′ measured with triangle method. The tetrad method is considered to be more powerful than the triangle method as long as the noise does not increase by more than 50% or the d′ does not decrease by more than 1/3 (Ennis 2012). The values for d′ are based on the proportion correct and can be found using tables in Ennis et al. (2011). Tables in the same book can also be used to find the variance of d′. Results relating to d′ can be seen in Table 4.

The d′ values found in this study were not very consistent with theories expressed in the literature (Garcia 2012; O'Mahony 2013). There were a number of cases where the d′ value for the tetrad was higher than the d′ of the triangle. Of the 31 experiments conducted, eight of those resulted in a noise increase of more than 50% and a d′ decrease of more than 1/3. These eight experiments involved a large variety of products including Pinto beans, BB Vegetarian, BB + BBQ sauce, Milk, Applesauce 1, Applesauce 2, Tomato sauce with carrier, and Oat cereal. Some of these products, like the canned beans and Tomato sauce with carrier, were more complex. While others, like the milk, apple sauces, and cereal, were much simpler. Regardless, the tetrad proved to be less powerful in these eight cases.

Table 4. EFFECT SIZE (d′) RESULTS FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS

	Tetrad tests				Triangle tests					
Product	*N*	*Pc*[a]	*d′*	*Var d′*	*N*	*Pc*	*d′*	*Var d′*	⅔ Δd′	Noise increase (%)
Black beans	84	0.57	1.83	0.033	78	0.55	1.73	0.081	1.153	-5.5
Kidney beans	60	0.33	0.00	0.098	54	0.30	<0	***	***	***
Chili beans	54	0.57	1.28	0.051	54	0.57	1.84	0.115	1.227	43.8
Pinto beans	54	0.56	1.21	0.052	54	0.57	1.84	0.115	1.227	52.1
Baked beans (BB)	60	0.42	0.99	0.062	60	0.45	1.19	0.137	0.793	20.2
BB Smoky 1	54	0.33	0.00	0.109	54	0.41	0.93	0.203	0.620	***
BB Smoky 2	54	0.48	1.37	0.057	54	0.52	1.56	0.123	1.040	13.9
BB Vegetarian	54	0.48	0.96	0.062	54	0.59	1.94	0.115	1.293	102.1
BB Molasses	54	0.81	2.16	0.058	54	0.76	2.86	0.136	1.907	32.4
BB + liquid smoke	54	0.67	1.59	0.052	54	0.56	1.75	0.117	1.167	10.1
BB + brown sugar	54	0.91	2.69	0.085	54	0.69	2.42	0.120	1.613	-10.0
BB + BBQ sauce	54	0.83	2.25	0.062	54	0.89	3.90	0.220	2.600	73.3
Lemon-Lime soda	72	0.72	2.63	0.038	72	0.61	2.03	0.086	1.353	-22.8
Apple juice	150	0.83	3.33	0.023	150	0.68	2.39	0.043	1.593	-28.2
Apple juice[b]	29	0.86	2.41	0.127	29	0.69	2.44	0.224	1.627	1.2
Apple juice combined	179	0.83	2.24	0.013	179	0.68	2.40	0.036	1.600	7.1
Milk	72	0.50	1.47	0.041	72	0.65	2.25	0.087	1.500	53.1
Milk with color 1	90	0.86	2.36	0.040	90	0.59	1.92	0.069	1.280	-18.6
Milk with color 2	90	0.87	2.42	0.042	90	0.77	2.90	0.083	1.933	19.8
Applesauce 1	78	0.38	0.59	0.079	78	0.46	1.26	0.100	0.840	113.6
Applesauce 2	78	0.36	0.53	0.164	78	0.42	1.03	0.124	0.687	94.3
Applesauce 2[b]	31	0.68	1.62	0.085	31	0.42	1.01	0.319	0.673	-37.7
Applesauce 2 combined	109	0.45	0.83	0.031	109	0.42	1.03	0.088	0.687	24.1
Tomato sauce	72	0.53	1.61	0.039	72	0.46	1.24	0.109	0.827	-23.0
Tomato sauce w/ carrier	72	0.40	0.90	0.055	72	0.51	1.54	0.093	1.027	71.1
Cantaloupe	54	0.54	1.15	0.054	54	0.50	1.47	0.060	0.980	27.8
Cheese crackers	78	0.44	0.78	0.054	78	0.42	1.03	0.124	0.687	32.1
Wheat crackers	78	0.83	2.25	0.043	78	0.79	3.09	0.103	2.060	37.3
Oat cereal	78	0.55	1.20	0.035	78	0.63	2.12	0.080	1.413	76.7
Lunch meat 1	78	0.54	1.16	0.037	78	0.53	1.60	0.084	1.067	37.9
Lunch meat 2	54	0.83	2.25	0.062	54	0.52	1.53	0.124	1.020	-32.0

[a]Pc: Proportion correct = (N correct/ N total).
[b] Tests using naïve panelists only.

Test Power

The Z value, significance of d′, and power for each test was calculated using Excel® and based off findings by Teixeira *et al.* (2009). The specific formulas used can be found in the appendix. A formula similar to the one used to calculate Z value was used by ASTM (2009) when finding T values to compare Thurstonian discriminal differences. The findings are included in Table 5. The power for the Kidney beans triangle test could not be determined because table values were not available for the variance of d′ so further calculations were not possible.

Significant differences were found between tests when resulting Z values were larger than 1.96 or the d′ p-value was less than 0.05. The 1.96 value was chosen based on a 95% confidence interval as was done in the ASTM Thurstonian discriminal distances standard (ASTM 2009). This occurred in six experiments: BB Vegetarian, BB + BBQ sauce, Apple juice, Milk, Wheat crackers, and Oat cereal. The effect size was significantly larger ($p < 0.05$) with the triangle method in the BB Vegetarian, BB + BBQ sauce, Milk, Wheat crackers, and Oat cereal. These experiments confirm literature findings (Ennis 2012; Garcia et al. 2012) that predicted a drop in effect size when the fourth sample was introduced. The Apple juice experiment, however, contradicts these predictions as the d′ found with the tetrad test (3.33) was significantly higher ($p < 0.05$) than the d′ found with the triangle test (2.39).

Table 5. TEST POWER VALUES FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS

	Tetrad tests		Triangle tests		d′ difference		Tetrad tests	Triangle tests
Product	*d′*	*Var d′*	*d′*	*Var d′*	Z value[a]	p-value	*Test power*	*Test power*
Black beans	1.83	0.033	1.73	0.081	0.296	0.767	**1.000**	0.990
Kidney beans	0.00	0.098	<0	---[b]	---	---	0.050	---
Chili beans	1.28	0.051	1.84	0.115	1.374	0.169	0.980	**1.000**
Pinto beans	1.21	0.052	1.84	0.115	1.540	0.124	0.963	**1.000**
Baked beans (BB)	0.99	0.062	1.19	0.137	0.448	0.654	**0.800**	0.624
BB Smoky 1	0.00	0.109	0.93	0.203	1.665	0.096	0.050	**0.308**
BB Smoky 2	1.37	0.057	1.56	0.123	0.448	0.654	**0.982**	0.883
BB Vegetarian	0.96	0.062	1.94	0.115	2.333	0.020	0.779	**1.000**
BB Molasses	2.16	0.058	2.86	0.136	1.588	0.112	1.000	1.000
BB + liquid smoke	1.59	0.052	1.75	0.117	0.389	0.697	0.998	**1.000**
BB + brown sugar	2.69	0.085	2.42	0.120	0.596	0.551	1.000	1.000
BB + BBQ sauce	2.25	0.062	3.90	0.220	3.109	0.002	1.000	1.000
Lemon-Lime soda	2.63	0.038	2.03	0.086	1.706	0.088	**1.000**	0.998
Apple juice	3.33	0.023	2.39	0.043	3.657	<0.001	1.000	1.000
Apple juice[c]	2.41	0.127	2.44	0.224	0.051	0.960	0.998	**1.000**
Apple juice combined	2.24	0.013	2.40	0.036	0.726	0.468	1.000	1.000
Milk	1.47	0.041	2.25	0.087	2.175	0.030	0.999	**1.000**
Milk with color 1	2.36	0.040	1.92	0.069	1.335	0.182	1.000	1.000
Milk with color 2	2.42	0.042	2.90	0.083	1.361	0.173	1.000	1.000
Applesauce 1	0.59	0.079	1.26	0.100	1.586	0.113	0.317	**0.806**
Applesauce 2	0.53	0.164	1.03	0.124	0.932	0.351	0.152	**0.544**
Applesauce 2[c]	1.62	0.085	1.01	0.319	0.960	0.337	0.975	**1.000**
Applesauce 2 combined	0.83	0.031	1.03	0.088	0.579	0.562	0.918	**1.000**
Tomato sauce	1.61	0.039	1.24	0.109	0.959	0.338	**1.000**	0.755
Tomato sauce w/ carrier	0.90	0.055	1.54	0.093	1.666	0.096	0.775	**0.947**
Cantaloupe	1.15	0.054	1.47	0.060	0.951	0.342	0.939	**1.000**
Cheese crackers	0.78	0.054	1.03	0.124	0.993	0.321	0.662	**1.000**
Wheat crackers	2.25	0.043	3.09	0.103	2.203	0.028	1.000	1.000
Oat cereal	1.20	0.035	2.12	0.080	2.714	0.007	0.995	**1.000**
Lunch meat 1	1.16	0.037	1.60	0.084	1.266	0.205	0.989	**1.000**
Lunch meat 2	2.25	0.062	1.53	0.124	1.671	0.095	1.000	1.000

[a] Z value = $|d'_1 - d'_2|/\mathrm{SQRT}(\mathrm{Var}\ d'_1 + \mathrm{Var}\ d'_2)$.

[b] Table values not available, further calculations not possible.

[c] Tests using naïve panelists only

Higher power values shown in bold

Test power was expected to increase for the tetrad based on literature findings (Ennis 2012; Ennis and Jesionka 2011), but that was not seen in all experiments. A majority of the experiments produced very high power values (> 0.90) in both tests. In experiments where one or more tests produced power values < 0.90, the tetrad resulted in higher power in three cases (Baked beans (BB), BB Smoky 2, and Tomato sauce), while the triangle generated higher power values in six experiments (BB Smoky 1, BB Vegetarian, Applesauce 1, Applesauce 2, Tomato sauce w/ carrier, and Cheese crackers). Test power could not be calculated for the triangle method with Kidney beans as table values for variances in d′ were not available. Product dependencies were seen, as power advantage was not consistent within product categories, especially with canned beans and vegetable sauces. The noise increase theory was further confirmed, as the eight tests mentioned in the previous section with high perceptual noise increases (Pinto beans, BB Vegetarian, BB + BBQ sauce, Milk, Applesauce 1, Applesauce 2, Tomato sauce w/ carrier, and Oat cereal), for the most part, produced lower test power (Ennis 2012). The only exception to this was BB + BBQ sauce which had an equivalent power value to the triangle method.

Degree of Difference

Similarly to perceptual difference in effect sizes, panelists were asked to rate the degree of difference they perceived between control and test samples for each test method. The responses of panelists who were able to correctly determine the difference between the two samples as well as mean scores for

each test are shown in Table 6. A Student's t-test was performed for each comparison experiment to determine if the perceived degree of difference means for the testing methods significantly differed for the products.

The mean scores for four of the comparison experiments significantly differed ($p<0.05$) and two more were trending toward a difference. In four of the six experiments that either significantly differed or trended toward a difference, the degree of difference mean was higher for the triangle method. Panelists were able to tell a significantly larger difference between samples with the Black beans ($p < 0.001$), BB Molasses ($p = 0.051$), BB + brown sugar ($p < 0.001$), and BB + BBQ sauce ($p < 0.001$) experiments using the triangle method. The opposite was true with the BB Vegetarian ($p = 0.037$) and small Apple juice ($p = 0.053$) experiments where the tetrad method produced higher degree of difference means.

Overall, the degree of difference perceived by panelists was product dependent. Products with stronger flavors and increased carryover, like the seasoned black beans, baked beans with brown sugar, and beans with added BBQ sauce, all faired better with the triangle. The products with more diluted flavors, like the vegetarian baked beans and juice, faired better with the tetrad. These results confirm findings in the literature (Ennis 2012).

Table 6. FREQUENCIES AND MEANS OF DEGREE OF DIFFERENCE SCORES BETWEEN CONTROL AND TEST SAMPLES.

	Tetrad tests								Triangle tests								
Product	*N correct*	*None*	*Very slight*	*Slight*	*Moderate*	*Large*	*Extremely large*	*Mean Score*[a]	*N correct*	*None*	*Very slight*	*Slight*	*Moderate*	*Large*	*Extremely large*	*Mean Score*	p-value
Black beans	48	0	8	19	14	7	0	1.9	43	1	2	12	14	10	4	3.0	<0.001
Kidney beans	20	0	6	10	3	1	0	2.0	16	1	8	6	0	1	0	1.5	0.131
Chili beans	31	0	2	2	15	10	2	3.3	31	0	4	6	11	8	2	2.9	0.223
Pinto beans	30	0	9	8	11	2	0	2.2	31	0	9	6	12	2	2	2.4	0.428
Baked beans (BB)	25	0	6	11	7	1	0	2.1	27	0	7	11	7	2	0	2.1	0.908
BB Smoky 1	18	0	7	6	3	2	0	2.0	22	2	8	6	5	1	0	1.8	0.498
BB Smoky 2	26	0	6	12	7	0	1	3.2	28	1	6	15	4	2	0	3.0	0.539
BB Vegetarian	26	1	13	6	4	2	0	2.5	32	3	10	9	8	2	0	1.9	0.037
BB Molasses	44	1	10	5	16	12	0	2.1	41	0	8	11	14	7	1	2.6	0.051
BB + liquid smoke	36	4	9	16	6	1	0	2.2	30	3	5	12	9	1	0	2.0	0.349
BB + brown sugar	49	0	6	12	23	7	1	1.9	37	1	4	8	16	8	0	2.7	<0.001
BB + BBQ sauce	45	0	8	10	19	8	0	2.0	48	1	3	13	17	11	3	2.9	<0.001
Lemon-Lime soda	52	1	5	15	21	6	4	2.8	44	0	6	13	14	7	4	2.7	0.857
Apple juice	124	0	12	41	53	17	1	2.6	102	0	6	46	38	11	1	2.6	0.529
Apple juice[b]	25	0	0	8	16	1	0	2.7	20	0	3	9	7	1	0	2.3	0.053
Apple juice combined	149	0	20	51	72	25	1	2.6	122	1	9	59	55	22	4	2.5	0.199
Milk	36	0	14	11	7	4	0	2.0	47	1	18	14	7	5	2	2.1	0.884
Milk with color 1	78	5	44	23	5	1	0	1.4	69	1	38	25	5	0	0	1.5	0.416
Milk with color 2	77	6	54	15	2	0	0	1.2	53	3	29	20	1	0	0	1.4	0.085
Applesauce 1	30	3	9	16	2	0	0	1.6	36	3	12	16	4	1	0	1.7	0.628
Applesauce 2	28	1	9	13	4	1	0	1.8	33	3	11	15	4	0	0	1.6	0.326
Applesauce 2[b]	21	1	9	6	5	0	0	1.7	13	1	5	6	1	0	0	1.5	0.552
Applesauce 2 combined	64	6	68	26	9	4	0	1.8	63	4	14	22	3	0	0	1.6	0.276

[a]Mean based on scale of Very slight= 1, Slight= 2, Moderate= 3, Large= 4, Extremely large= 5.
[b] Tests using naïve panelists only.

Table 6 (continued).

Product	Tetrad tests								Triangle tests								
	N correct	*None*	*Very slight*	*Slight*	*Moderate*	*Large*	*Extremely large*	*Mean Score*	*N correct*	*None*	*Very slight*	*Slight*	*Moderate*	*Large*	*Extremely large*	*Mean Score*	p-value
Tomato sauce	38	0	7	12	17	1	1	2.4	33	0	9	13	10	1	0	2.1	0.150
Tomato sauce w/ carrier	29	1	10	15	2	1	0	1.7	37	1	16	14	5	1	0	1.7	0.916
Cantaloupe	29	1	9	10	5	4	0	2.1	27	0	6	9	8	3	1	2.4	0.253
Cheese crackers	34	2	11	14	7	0	0	1.8	33	1	11	12	7	2	0	1.9	0.734
Wheat crackers	65	0	13	15	24	13	0	2.6	62	0	11	21	21	9	0	2.4	0.505
Oat cereal	43	3	19	12	4	5	0	1.7	49	1	17	16	10	4	1	2.0	0.120
Lunch meat 1	42	0	13	18	10	1	0	2.0	41	0	6	19	14	1	1	2.3	0.065
Lunch meat 2	45	0	16	19	5	5	0	3.0	28	1	5	8	11	3	0	2.4	0.122

Ease of Method

After completing the final test in each experiment, panelists compared the difficulty of completing the tetrad test to the triangle method using a fixed interval scale anchored by “Much easier” and “Much more difficult”. The least squares means from each experiment and mean separation were calculated using PROC GLIMMIX and the PDIFF option in SAS 9.3 (2011). The results for this portion of the study can be seen in Table 7.

BB Vegetarian produced the highest mean (3.3) and was significantly different ($p < 0.05$) from the lowest mean (2.5) group Applesauce 1, Applesauce 2 combined, and Cantaloupe. The flavor profile for BB Vegetarian is much more complex than that of the lower mean group, which could account for the difference in perceived difficulty between the two methods. While some separation was seen, all means fell within the 2.5 to 3.3 range, meaning the tests were perceived as “About the same” in terms of difficulty.

In addition to comparing the difficulty level of the tetrad method to the triangle method for each experiment on a fixed scale, panelists were also given an open-ended question to provide qualitative data to the study. Panelists were encouraged to describe whether they thought performing the tetrad was easier than, harder than, or about the same as the triangle using their own words. Representative comments from panelists are included in Table 8.

Table 7. MEANS ASSIGNED TO DIFFICULTY LEVEL OF TETRAD TEST METHOD WHEN COMPARED TO TRIANGLE METHOD USING FIXED INTERVAL SCALING[a]

Product	N	Mean	Standard error	Mean separation
Black beans	84	3.1	0.10	ABCDEF
Kidney beans	60	3.1	0.12	ABCDEF
Chili beans	54	2.9	0.13	CDEFGH
Pinto beans	54	3.1	0.13	ABCDEF
Baked beans (BB)	60	3.1	0.12	ABCDEF
BB Smoky 1	54	3.2	0.13	ABC
BB Smoky 2	54	3.1	0.13	ABCDEF
BB Vegetarian	54	3.3	0.13	A
BB Molasses	54	3.1	0.13	ABCDEF
BB + liquid smoke	54	3.1	0.13	ABCDE
BB + brown sugar	54	2.8	0.13	DEFGHIJ
BB + BBQ sauce	54	3.0	0.13	ABCDEFG
Lemon-Lime soda	72	3.1	0.12	ABCDEF
Apple juice	150	2.9	0.08	CDEFG
Apple juice[b]	29	3.3	0.18	ABC
Apple juice combined	179	3.0	0.07	BCDEF
Milk	72	3.1	0.11	ABCD
Milk with color	90	3.0	0.10	ABCDEF
Applesauce 1	78	2.5	0.11	J
Applesauce 2	78	2.6	0.11	HIJ
Applesauce 2[b]	31	2.5	0.17	IJ
Applesauce 2 combined	109	2.6	0.09	J
Tomato sauce	72	2.8	0.11	DEFGHI
Tomato sauce w/ carrier	72	2.8	0.11	EFGHIJ
Cantaloupe	54	2.5	0.13	J
Cheese crackers	78	3.2	0.11	AB
Wheat crackers	78	2.8	0.11	FGHIJ
Oat cereal	78	2.7	0.11	HIJ
Lunch meat 1	78	2.7	0.11	GHIJ
Lunch meat 2	54	2.8	0.13	DEFGHIJ

[a]Means based on scale: Much easier= 1, Slightly easier= 2, About the same= 3, Slightly more difficult= 4, Much more difficult= 5.

[b] Tests using naïve panelists only.

Means followed by like letters do not differ ($p>0.05$).

Table 8. REPRESENTATIVE VERBATIM PANELIST COMMENTS WHEN ASKED TO DESCRIBE DIFFICULTY OF TETRAD TESTING METHOD COMPARED TO TRIANGLE METHOD

Perceived difficulty	Product	Comment
Tetrad easier	Pinto beans	Triangle tests are slightly harder because it can be difficult to find an odd sample. In a tetrad test, it is easier to match products together based on characteristics picked up. tetrad test is a method for panelists to 'double check' the differences
	Baked beans (BB)	I suppose it gives more reassurance being able to taste a pair different than having the thought that one may or may not be different.
		seems a little easier since you have a extra sample to make comparisons and confirm your observations
	BB Smoky 2	It varies depending on what I'm testing, but for this test, having a second sample to corroborate the differences I thought I'd noticed the first time is a good verification for me. It helps my confidence in making a choice.
	BB Molasses	I thought it was easier because by the third one I was really confused but the fourth one really sealed the deal.
		When the differences are as significant as today it is not hard to do a tetrad but otherwise it is easier to do a triangle test.
	BB + brown sugar	It seemed easier to pick the two that were most alike and two that were different, than it would be to pick the lone different sample in a triangle. Could just be psychological too.
	BB + BBQ sauce	tetrad tests are easier because they allow matching of two samples, whereas in a triangle test panelists must find the odd sample out. tetrad is a easy way of 'double checking' what is perceived
	Apple juice	I like having the slight edge of four samples to just three. Makes you think more about the flavors.
		Having another sample to compare the "odd" sample to made it easier to group them.
	Milk	Since I am grouping samples, I am more focused on determining distinguishing features between samples. I must group them so I focus on similarities and differences.
	Applesauce 1	The tetrad test seemed easier to me because I was more certain about the difference that I detected because the other samples confirmed it in my mind.
	Applesauce 2	They were relatively similar, but the tetrad test seemed easier to me, since the difference could be verified by that second sample. It made my decision more confident.
	Tomato Sauce	I was pretty confident I picked the 'right' ones in the 4 test, not very confident in my 3 test.
	Oat cereal	For some reason today the 4 sample test was easier to draw a visual difference. The texture was also easier to narrow down.
About the same	Pinto beans	These samples were pretty different so both tests were fairly easy. I usually find the 3-sample to be just slightly easier.
		I don't find it more difficult, just samples may need to be retested in this type of sampling to ensure you are getting the correct flavors down.
	BB Vegetarian	it is about the same, since very little flavor is carried from sample to sample

Table 8 (continued).

Perceived difficulty	Product	Comment
About the same	BB Molasses	When the samples are as easy to tell apart as they were today, it's the same difficulty in a 3- vs 4- test
	BB + liquid smoke	When the difference is pretty obvious, there's not difference in difficulty between triangle and tetrad.
Tetrad harder	Black beans	I prefer the triangle test. Four samples - even though two match - is a bit much to process.
	Kidney beans	I had to taste each sample twice to remind myself which flavor went with which sample
	BB Smoky 2	it was difficult in the fact that I thought the difference was so slight and the aftertaste is strong enough to impact the next sample
		It made it a little more difficult because once you found one that was different you had to match it to another one. It involved 2 steps as opposed to a triangle test.
	BB Molasses	I was able to determine one member of a group, but had difficulty determining its mate. This sample was much too similar to determine groupings.
		Comparing 4 samples that are only slightly different is always more difficult with doing 4 samples compared to three. Longer tasting time between 1st to 4th and trying to remember each taste
	BB + liquid smoke	This tetrad required me to resample previous samples due to me not really tasting a huge difference between products
		It is easier for me to pick out one sample that is different and not two samples that are different. I have to taste the samples multiple times whereas with the three I can usually guess the first time which one is different.
	BB + BBQ sauce	It was more difficult than the three sample test because I was not only looking for differences, but similarities between different samples. Once you get to the fourth sample, it is difficult to remember how the first one tastes.
	Apple Juice	The triangle was more intuitive and was easier
		Even though I didn't detect a difference in the triangle test samples, the tetrad set up felt harder because I had to remember all four tastes. As compared with the triangle test where I only had to remember two tastes, and then compare/contrast the third
	Milk	I did not have a problem differentiating between the first two samples but by the time I got to the fourth I started to become unsure of myself.
	Applesauce 1	It's much easier to find one out of three that is different vs finding a pair in four. The more I tasted in the tetrad to find a pair, the more the samples tasted so much alike. It's harder to find a pair in the tetrad and much easier to find one that differs
	Tomato Sauce	It was easier to pick one odd sample than to group samples by twos. The triangle can be completed with one tasting; the tetrad required retasting to corroborate group choice, which led to taste fatigue and second guessing.
	Cantaloupe	The 4 sample test was more difficult as the attributes began to mesh together making the separation more difficult

Many of the panelists who cited the tetrad as being easier to perform than the tetrad mentioned increased confidence in their group selection. The addition of the fourth sample helped confirm their choice for the “odd” sample. Those who reported the tetrad as being harder to perform than the triangle mentioned taster fatigue and having to re-taste samples to remind themselves of what each sample tasted like. This concern was expressed previously in the literature (Ennis 2012). A few panelists also noted that the tetrad method in general took longer to complete than the triangle method.

Again, product dependencies can be seen between the two methods. Many panelists pointed out that when large differences existed between samples, difficulty levels were not impacted, and the tetrad was just as easy to complete as the triangle. In all three of the categories, panelists pointed out that when an aftertaste was present or differences between control and test samples were slight, the triangle method was easier to complete than the tetrad method. Fewer samples to choose from prevented flavors from muddling together and put less strain on panelists’ cognitive memory load.

CHAPTER V

Summary and Conclusions

Based on findings from this study, the theoretical and statistical advantages of the tetrad discrimination method may not outweigh the concerns surrounding the test. Variations in test performance were seen within and across product categories for many of the testing parameters, especially with the canned beans category. These variations can be seen in the qualitative data collected from participating panelists and much of the quantitative data analysis.

Many panelists participating in this study noted that their perception of method difficulty depended on the product being tested. In experiments where large differences existed between samples, panelists voiced that the difficulty level of the two methods was about the same. Panelists also stated that products with more complex flavor profiles, like many of the canned beans and tomato sauces, were easier to differentiate when the triangle method was used. A number of panelists also stated that the tetrad took longer for them to complete because they had to re-taste samples. The strong flavors and carryover between samples were too taxing on panelists' memory load. Many of the panelists who stated that the tetrad was easier to perform than the triangle said the fourth sample increased confidence in their decision. The additional sample helped confirm their choice of the "odd" sample. When asked to rate the perceived level of method difficulty on a fixed scale, the tests were viewed as "About the same"

with very little separation between experiments. Experiments with blander flavored products, again, were “slightly easier” to complete with the tetrad. The tetrad tended to be closer to “slightly more difficult” to complete than the triangle.

When the degree of difference between samples was measured, a significant difference ($p \leq 0.05$) was found in six experiments. Of these six, the triangle method had a higher degree of difference; meaning panelists were able to perceive a larger degree of difference using the triangle than the tetrad. This result was expected as increasing the number of samples presented, as noted in the panelist comments, decreases taster sensitivity. It is interesting to note that no difference was seen between methods for the coloring experiment where no tasting was done.

Increased test power has been a major selling point for advocates of the tetrad discrimination method. An increase in power would result in a smaller number of panelists needed to find a difference. This could indeed prove advantageous for companies looking to decrease panel cost. When the d′ values were inspected, it was found that the tetrad was not as powerful as the triangle in eight cases as the perceptual noise increased by more than 50% and the d′ for the tetrad decreased by more than 1/3. These findings were further confirmed after test power was calculated. For the products used in these eight experiments (Pinto beans, BB Vegetarian, BB + BBQ sauce, Milk, Applesauce 1, Applesauce 2, Tomato sauce with carrier, and Oat cereal), testing power for the tetrad was

lower than or equal to that of the triangle. The triangle method resulted in higher test power in nine other experiments as well (Chili beans, BB Smoky 1, BB + liquid smoke, Apple juice, Applesauce 2 with naïve panelists, Applesauce 2 combined, Cantaloupe, Cheese crackers, and Lunch meat 1). Companies would be advised to use the triangle method over the tetrad for discrimination tests for these products.

When determining if significant differences exist between samples, p-value calculations are commonly done. For this study, most of the significant differences ($p < 0.05$) found for each test agreed for the comparison experiments. The p-value results did differ for five of the experiments. Baked beans (BB), Applesauce 1, and Tomato sauce with carrier produced significant p-values (< 0.05) when the triangle method was used but not the tetrad. The opposite was true with the smaller Applesauce 2 with naïve panelists and Cheese cracker experiments where a difference was found ($p < 0.05$) with the tetrad but not the triangle. It is interesting to note that when the same products for the smaller Applesauce 2 panel were repeated with a larger group of panelists that were more accustomed to discrimination testing, a difference was found with both methods. Comfort level and experience with methodology could have affected the results of the experiments. This should especially be of interest to companies with discrimination panels already in place.

Based on these findings, it should be evident that many factors play into the outcome of a test. As was seen, flavor profiles, panelists' perception, and type of data analysis completed all had an effect on which test was advantageous. Because of this, it is strongly recommended that companies have a very thorough understanding of their products and existing discrimination panels before deciding to switch from the triangle method to the tetrad. Testing their own products with both methods may prove to be beneficial to companies considering making the switch.

LIST OF REFERENCES

AMERINE, M.A., PANGBORN, R.M., and ROESSLER E.B. 1965. *Principles of Sensory Evaluation of Food*. New York: Academic Press Inc.

ASTM Standard E2262-03. 2009. Standard practice for estimating Thurstonian discriminal distances. ASTM International. West Conshohocken, PA. 2010. DOI: 10.1520/E2262-03R09, <www.astm.org>.

ASTM WK32980. 2011. New test methods for sensory analysis- tetrad test. ASTM International. West Conshohocken, PA. <www.astm.org>.

BI, J., ENNIS, D.M., and O'MAHONY, M. 1997. How to estimate and use the variance of d′ from difference tests. J. Sensory Studies 12, 87-104.

BIOSYSTEMES. 2009. FIZZ software solutions for sensory analysis and consumer tests version 2.4. Couteron, France.

CARR, T.. 1998. Discrimination test analysis tool. Carr Consulting. Wilmette, IL.

DELWICHE, J. and O'MAHONY M. 1996. Flavor discrimination – an extension of the Thurstonian paradoxes to the tetrad method. Food Qual. Prefer. 7, 1-5.

ENNIS, D.M. 1993. The power of sensory discrimination methods. J. Sensory Studies 8, 353-370.

ENNIS, D.M., MULLEN, K., and FRIJTERS, J.E.R. 1988. Variants of the method of triads: unidimensional Thurstonian models. Br. J. Math. Stat. Psychol. 41, 25-36.

ENNIS, D.M., ROUSSEAU, B., and ENNIS, J.M. 2011. Tables for product testing methods. In: *Short stories in Sensory and Consumer Science*. 1st ed. revised. IFPress. Richmond, VA.

ENNIS, J.M. 2012. Guiding the switch from triangle testing to tetrad testing. J. Sensory Studies 27, 223-231.

ENNIS, J.M. 2013. The year of the tetrad test. J. Sensory Studies 28(4), 257-258.

ENNIS, J.M., ENNIS, D.M., YIP, D., and O'MAHONY, M. 1998. Thurstonian models for variants of the method of tetrads. Br. J. Math. Stat. Psychol. 31, 205-215.

ENNIS, J.M. and JESIONKA, V. 2011. The power of sensory discrimination methods revisited. J. Sensory Studies 26, 371-382.

FOOD SAFETY INTERNATIONAL NETWORK. 2012. Internal sensory testing: tetrad test, power and consumer relevance course. FOOD SAFETY INTERNATIONAL NETWORK, Inc. Los Angeles, CA. <http://www.safefoodnetwork.com/english/english/news/learning-events/3150-techniques-advanced-for-sensory-evaluation-of-foods-workshop.html> (May 24, 2012).

GARCIA, K., ENNIS, J.M., and PRINYAWIWATKUL, W. 2012. A large-scale experimental comparison of the tetrad and triangle tests in children. J. Sensory Studies 27, 217-222.

GARCIA, K., ENNIS, J.M., and PRINYAWIWATKUL, W. 2013. Reconsidering the specified tetrad test. 2013. J. Sensory Studies 28, 445-449.

GELSKI, J. 2013. Switching sensory test protocol benefits General Mills. Food Business News. <http://www.foodbusinessnews.net> (Feb 28, 2013).

ISHII, R., O'MAHONY, M., and ROUSSEAU, B. 2014. Triangle and tetrad protocols: small sensory differences, resampling and consumer relevance. Food Qual. Prefer. 31, 49-55.

LAWLESS, H.T. and HEYMANN, H. 1998. *Sensory evaluation of food: principles and practices.* Chapman & Hall. NY.

LAWLESS, H.T. 2013. *Quantitative sensory analysis: psychophysics, models, and intelligent design.* John Wiley & Sons. Somerset, NJ.

LEATHERHEAD FOOD RESEARCH. 2014. Sensory difference testing - overview & new developments. Leatherhead Food International Limited. Surrey, UK. < http://www.leatherheadfood.com/sensory-difference-testing>.

MASUOKA, S., HATJOPOULOS, D., and O'MAHONY, M. 1995. Beer bitterness detection: testing Thurstonian and sequential analysis models for triad and tetrad methods. J. Sensory Studies 10(3), 295-306.

MEILGAARD, M.C., CARR, B.T., and CIVILLE, C.G. 2006. *Sensory evaluation techniques*, 4th ed. CRC Press. Boca Raton, FL.

O'MAHONY, M. 2013. The tetrad test: looking back, looking forward. J. Sensory Studies 28(4), 259-263.

O'MAHONY, M. and ROUSSEAU, B. 2002. Discrimination testing: a few ideas, old and new. Food. Qual. Prefer. 14, 157-164.

ROUSSEAU, B. and ENNIS, J.M. 2013. Importance of correct instructions in the tetrad test. J. Sensory Studies 28, 264-269.

SAS INSTITUTE, INC. 2011. SAS/Stat Software. The SAS systems for Windows release 9.3. Cary, NC: SAS Institute, Inc.

SENSORY DIMENSIONS. 2013. Sixth sense: introducing the tetrad test. Sensory Dimensions. Reading, UK. <http://sensorydimensions.co.uk/Documents/e032_December_2013.htm> (Dec 2013).

TEIXEIRA, A., ALVARO, R., and CALAPEZ, T., IBS – ISCTE Business School (Lisbon). 2009. Statistical power analysis with Microsoft Excel: normal tests for one or two means as a prelude to using non-central distributions to calculate power. JSE 17: < www.amstat.org/publications/jse/v17n1/teixeira.html> (Nov 1, 2009).

YIP, D. 1996. Triadic and tetradic taste discrimination testing: Thurstonian and sequential effects. MS Thesis, University of California, Davis, 110pp. As cited by: O'MAHONY, M. 2013. The tetrad test: looking back, looking forward. J. Sensory Studies.

APPENDICES

APPENDIX A

Demographic Information

PANELIST AGE DISTRIBUTIONS FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS

	Tetrad tests							Triangle Tests						
Product	*N*	*18-24*	*25-34*	*35-44*	*45-54*	*55-64*	*Over 65*	*N*	*18-24*	*25-34*	*35-44*	*45-54*	*55-64*	*Over 65*
Black beans	84	16	24	9	25	10	0	78	18	21	9	20	9	1
Kidney beans	60	16	16	11	9	8	0	54	22	13	3	8	8	0
Chili beans	54	21	15	5	4	8	1	54	18	18	5	7	5	1
Pinto beans	54	19	20	3	5	5	2	54	11	19	7	8	8	1
Baked beans (BB)	60	15	22	8	8	7	0	60	11	17	13	12	6	1
BB Smoky 1	54	13	15	8	11	6	1	54	12	15	6	13	7	1
BB Smoky 2	54	10	20	6	12	5	1	54	14	12	7	14	6	1
BB Vegetarian	54	20	14	6	8	3	3	54	17	14	5	10	7	1
BB Molasses	54	19	18	4	4	8	1	54	13	18	6	8	8	1
BB + liquid smoke	54	15	13	6	8	10	2	54	14	18	5	7	9	1
BB + brown sugar	54	12	18	7	10	7	0	54	16	18	6	6	7	1
BB + BBQ sauce	54	24	13	6	5	5	1	54	17	16	5	8	7	1
Lemon-Lime soda	72	22	24	10	10	6	0	72	25	22	8	12	5	0
Apple juice	150	57	42	15	20	15	1	150	57	42	15	20	15	1
Apple juice[a]	29	27	2	0	0	0	0	29	27	2	0	0	0	0
Apple juice combined	179	84	44	15	20	15	1	179	84	44	15	20	15	1
Milk	72	20	22	7	15	8	0	72	20	24	10	9	9	0
Milk with color 1	90	38	23	9	10	9	1	90	38	23	9	10	9	1
Milk with color 2	90	38	23	9	10	9	1	90	38	23	9	10	9	1
Applesauce 1	78	15	28	11	14	10	0	78	15	28	11	14	10	0
Applesauce 2	78	20	25	9	13	10	1	78	20	25	9	13	10	1
Applesauce 2[a]	31	29	2	0	0	0	0	31	29	2	0	0	0	0
Applesauce 2 combined	109	49	27	9	13	10	1	109	49	27	9	13	10	1
Tomato sauce	72	17	24	11	12	7	1	72	17	24	11	12	7	1
Tomato sauce w/ carrier	72	25	21	8	13	5	0	72	25	21	8	13	5	0

[a] Tests using naïve panelists only.

PANELIST AGE DISTRIBUTIONS FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS (continued).

	Tetrad tests							Triangle Test						
Product	*N*	*18-24*	*25-34*	*35-44*	*45-54*	*55-64*	*Over 65*	*N*	*18-24*	*25-34*	*35-44*	*45-54*	*55-64*	*Over 65*
Cantaloupe	54	21	15	2	8	7	1	54	21	15	2	8	7	1
Cheese crackers	78	31	25	6	8	6	2	78	31	25	6	8	6	2
Wheat crackers	78	28	23	7	11	9	0	78	28	23	7	11	9	0
Oat cereal	78	22	25	5	12	12	2	78	22	25	5	12	12	2
Lunch meat 1	78	25	18	9	15	10	1	78	25	18	9	15	10	1
Lunch meat 2	54	17	16	5	8	8	0	54	17	16	5	8	8	0

PANELIST GENDER FREQUENCIES FOR TETRAD AND TRIANGLE COMPARISON EXPERIMENTS

	Tetrad tests			Triangle tests		
Product	*N*	*Males*	*Females*	*N*	*Males*	*Females*
Black beans	84	29	55	78	28	50
Kidney beans	60	15	45	54	11	43
Chili beans	54	13	41	54	16	38
Pinto beans	54	18	36	54	15	39
Baked beans (BB)	60	18	42	60	16	44
BB Smoky 1	54	18	36	54	16	38
BB Smoky 2	54	20	34	54	15	39
BB Vegetarian	54	21	33	54	16	38
BB Molasses	54	16	38	54	18	36
BB + liquid smoke	54	18	36	54	17	37
BB + brown sugar	54	14	40	54	15	39
BB + BBQ sauce	54	20	34	54	18	36
Lemon-Lime soda	72	22	50	72	20	52
Apple juice	150	50	100	150	50	100
Apple juice[a]	29	8	21	29	8	21
Apple juice combined	179	58	121	179	58	121
Milk	72	23	49	72	19	53
Milk with color 1	90	25	65	90	25	65
Milk with color 2	90	25	65	90	25	65
Applesauce 1	78	25	53	78	25	53
Applesauce 2	78	29	49	78	29	49
Applesauce 2[a]	31	8	23	31	8	23
Applesauce 2 combined	109	37	72	109	37	72
Tomato sauce	72	23	49	72	23	49
Tomato sauce w/ carrier	72	21	51	72	21	51
Cantaloupe	54	15	39	54	15	39
Cheese crackers	78	29	49	78	29	49
Wheat crackers	78	23	55	78	23	55
Oat cereal	78	23	55	78	23	55
Lunch meat 1	78	22	56	78	22	56
Lunch meat 2	54	16	38	54	16	38

[a] Tests using naïve panelists only.

APPENDIX B

Additional Formulas

Power calculations for Excel®

Z value

$$Z = (ABS(d'_1 - d'_2)) / (SQRT(Var\ d'_1 + Var\ d'_2))$$

p-value

p-value = 1 - _xlfn.NORM.DIST(ABS(Z),0,1,True))*2

Test power

Power = 1 – NORMDIST((-NORMSINV(0.05/2)),(d′/SQRT(2*Var d′)),1,1) + (NORMDIST(NORMSINV(0.05/2)), (d′/SQRT(2*Var d′)),1,1)

SAS 9.3 Code

```
proc glimmix data=methodease2;
class prod;
model ease = prod/ ddfm=kr ;
lsmeans prod/ pdiff;
 ods  exclude lsmeans diffs;
 ods output lsmeans=mmm diffs=ppp;
 output out=rrr resid=resid;
run;
%pdmix(ppp,mmm);
%include 'a:pdmix800.sas';
%pdmix800(ppp,mmm,alpha=.05,sort=yes);
```

VITA

Sara Lyn Carlisle was born in West Palm Beach, Florida in 1990 to Wm. Terry and Jane Carlisle. Sara has two younger brothers, Dylan and Dakota Carlisle. The family moved to Springfield, Tennessee in 1996 where Sara grew up and attended school. She graduated from Springfield High School in 2008. Sara graduated from the University of Tennessee at Knoxville in the summer of 2012 with a Bachelor of Science in Food Science and Technology and a concentration in Business after completing a sensory internship with Bush Brothers & Company. The following fall Sara began work as a graduate research assistant in the University of Tennessee Sensory Lab while studying toward a Master of Science degree in Food Science and Technology with a minor in Statistics.