Food Quality and Preference 68 (2018) 105–112

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Comparison of triangle and tetrad discrimination methodology in an applied manner

Sara L. Burns[a], Marjorie P. Penfield[a], Arnold M. Saxton[b], Curtis R. Luckett[a,*]

[a] Department of Food Science, Institute of Agriculture, University of Tennessee, 2510 River Drive, Knoxville, TN 37996, United States
[b] Department of Animal Science, Institute of Agriculture, University of Tennessee, 2506 River Drive, Knoxville, TN 37996, United States

ARTICLE INFO

Keywords:
Tetrad
Triangle
Discrimination

ABSTRACT

Choosing a discrimination test can involve numerous factors, one of which is statistical power. The tetrad test has been shown to possess statistical advantages over the more traditional triangle method. However, these statistical advantages may not be present when the effect size decreases by more than 1/3. The frequency of large decreases in effect size, as well as other measures of test performance has not been characterized. In this study, over thirty products were tested using both triangles and tetrads in order to compare the two methods. The products tested ranged from canned vegetables and fresh fruits to deli meats and baked goods. After testing, inconsistencies were found within and across product categories. Only 16% of the tests were found to end with a different conclusion regarding a statistical difference ($p < 0.05$). In six of the experiments, the triangle test showed a significantly higher effect size than the tetrad. In eight of the experiments, the reduction in effect size for the tetrad led to no power advantage of the tetrad test. Participants also noted that the product being tested affected their impression of test difficulty in multiple experiments. This study creates a functional comparison of tetrad and triangle testing and quantifies the frequency in which the tetrad method effect size decreases by more than 1/3, leading to decreased statistical power.

1. Introduction

Discrimination testing is used to determine if a difference exists between products, and implemented 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 used may depend on the complexity of the product, test sensitivity, and panelists to be used.

Triangle and tetrad are two of the most common discrimination testing methods currently used. In the triangle method, panelists are presented with three samples simultaneously. Of these three, two are alike and one is different or "odd". The probability of correctly guessing the odd sample is 1 out of 3 (or 33.3%). In tetrad testing, four samples are presented simultaneously – two from one group and two from another. Instead of choosing one sample, panelists are asked to sort the samples into groups based on similarity. The probability of correctly sorting the groups by chance is also 1 out of 3 (Ennis, 2012). Both methods employ Thurstonian discriminal difference modeling and use binomial distribution to determine if a difference exists. Power for each test can be calculated when effect size (d'), chosen alpha (0.05), and number of panelists used are taken into account. The tetrad method is said to be more powerful than the triangle as long as the effect size does not decrease by more than 1/3 and perceptual noise increases by less than 50% (Ennis, 2012). The d' value should theoretically always decrease for the tetrad because the additional sample inherently adds noise to the test. However, it is not clear how frequently the d' does not decrease by 1/3 and the perceptual noise increases by less than 50%. Therefore, there is not a clear understanding of the how the theoretical power increases associated with tetrads manifest themselves in real situations.

The tetrad method has been described as a more sensitive testing method than the more traditional triangle that could save companies money by reducing the number of panelists and number of samples required because of its increased test power (ASTM, 2015; Ennis, 2012; Ennis, 2013). This reduction in required resources could allow companies to test products in-house and receive results immediately, saving both time and money involved in outsourcing tests. Larger proportion of correct responses and smaller variations in d' have also been seen when using the unspecified tetrad method, in which the differing attribute is not identified, than the triangle method (Bi & O'Mahony,

* Corresponding author.
E-mail address: cluckett@utk.edu (C.R. Luckett).

https://doi.org/10.1016/j.foodqual.2018.02.009
Received 15 December 2017; Received in revised form 18 February 2018; Accepted 22 February 2018
Available online 23 February 2018
0950-3293/

2013). This would be very advantageous in instances where equivalence or intensity-related testing is needed. Many concerns surrounding the tetrad methodology, however, have been presented. The addition of a fourth sample could lead to panelist fatigue and a reduction in sensitivity to the stimulus (Ennis, 2012; Bi & O'Mahony, 2013). 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). A study by Ishii, O'Mahony, and Rousseau (2014), using fruit juices, acknowledged this concern, stating that even though they found the tetrad to be a more powerful alternative to the triangle, more "fatiguing" products might produce different results.

There are also other issues dealing with the implementation of tetrad methodology for which there are not consensus. For example, Delwiche and O'Mahony (1996) found a specified tetrad method, in which a specific attribute, like sweetness or bitterness, is addressed, more statistically advantageous than an unspecified version. Conversely, Masuoka, Hatjopoulos, and O'Mahony (1995) found no difference between addressing a specific attribute and performing the tetrad as an overall difference test. Compounding this uncertainty, very few direct comparisons have been conducted between the triangle and tetrad methods. O'Mahony (2013), using Delwiche and O'Mahony (1996) data, did so with conflicting results. When looking at Yip (1996), O'Mahony found the tetrad methodology to have a significantly lower d' ($p = 0.0005$) than the triangle. However, the reduction in d' was less than 1/3 (1.66 for the triangle and 1.17 for the tetrad), indicating that the tetrad was theoretically more powerful than the former. Bi and O'Mahony (2013) compared a number of force-choice methods using the variance in d' to take treatment effects into account. They found smaller variations of d' with the unspecified tetrad inferring greater power but warned the addition of a fourth sample could affect its "operational" power. Xia et al. (2015) utilized a variety of beverages when conducting their comparison of the triangle and tetrad methods. They also concluded that the tetrad was a more powerful alternative but did not see the expected drop in d' for the tetrad.

The purpose of this study is to address these concerns in an applied, industrial approach to compare triangle and tetrad test results. The main objective was to assess the frequency of large decreases in d' and notable increases in perceptual noise, both of which diminish the statistical power of a tetrad test. Additionally, the studies were analyzed to compare the two methods using common measures of discrimination tasks. These tests were completed in a single session to determine which differences and similarities exist. Qualitative data were also gathered to gain insight on panelist perception of the testing methods, which could be helpful to those trying to decide whether to make the switch from triangle to tetrad or not.

2. Materials and methods

All experiments in this study were conducted in the University of Tennessee at Knoxville in individualized booths using FIZZ by Biosystemes (Version 2.4; Couternon, France) for data acquisition. Samples were presented in balanced orders. Randomly chosen 3-digit codes were assigned to products. 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 pronounced visual differences. Demographic information was collected following the discrimination task. 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. Comments were sorted by implied perception of method difficulty into three groups (Tetrad easier, About the same, and Tetrad harder).

2.1. Participants

Participants participating in this study were recruited using the University of Tennessee at Knoxville sensory email database, which includes roughly 600 food consumers. To be included in this database, previous participation in sensory studies at the University of Tennessee is required. These individuals were familiar with testing protocol and considered to be experienced panelists. All participants in the database 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. This information was also posted outside of the lab during testing for individuals who were not receiving emails but still interested in participating. These individuals were considered naïve panelists. In order to participate, panelists must have been 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 conducted according to the Declaration of Helsinki for studies on human subjects and approved by the University of Tennessee IRB review for research involving human subjects.

Approximately 63% of the panelists were 18–34 years of age. The exceptions include the smaller apple juice and applesauce experiments, where 93% of the panelists were 18–34 years of age. Strictly naïve panelists were recruited for these two experiments. 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.

2.2. Products

In order to simulate the practical application of the tetrad method in the food industry, this study was conducted in the same manner a company would approach difference testing. Because of this, each test was done in a single session. A variety of products was used to encompass the many facets of the food industry. Table 1 contains descriptions of the products used for control and test samples in each experiment. Product names used in later results tables and in the discussion, can also be found in the table following the control description. To maintain the proprietary nature of the data for industry partners that provided products, specific brands names are not mentioned. Materials used in the experiments that were not provided by industry partners were purchased from local supermarkets.

Both tests where samples could be prepared in advance occurred on the same day in a balanced fashion, half of the panelists received the tetrad first and half received the triangle first. Order of test presentation was determined to not have an effect on results. If serving was time or temperature 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. Participation in both the triangle and tetrad was not required in these instances.

2.3. 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. Following the completion of both tests for experiments with a balanced design or the tetrad for single day tests, panelists were asked to describe their impression of both testing methods using their own words.