Food Quality and Preference 56 (2017) 119–125

Contents lists available at ScienceDirect

Food Quality and Preference

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

Effect of monetary reward and food type on accuracy and assessment time of untrained sensory panelists in triangle tests

Jessilee N. Loucks [a], Dennis L. Eggett [b], Michael L. Dunn [a], Frost M. Steele [a], Laura K. Jefferies [a,*]

[a] Department of Nutrition, Dietetics and Food Science, Brigham Young University, S-221 ESC, Provo, UT 84602, USA
[b] Department of Statistics, Brigham Young University, 223 TMCB, Provo, UT 84602, USA

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 20 September 2016
Accepted 21 September 2016
Available online 22 September 2016

Keywords:
Monetary
Food type
Incentive
Triangle test
Accuracy
Assessment time

ABSTRACT

The effect of monetary reward amount and test food type as factors of triangle test accuracy and assessment time for untrained panelists was explored. Monetary compensation is commonly used to reward panelists for their time and effort. While studies have documented that paying panelists can influence hedonic ratings, research on its possible influence on triangle test accuracy or time to assess products is lacking. Relatedly, some studies suggest that assessment time influences accuracy. Furthermore, little research has been conducted on the effect of the test food itself, and its general likeability, on biasing panelist accuracy or assessment time. Pairs of two liked foods – chocolate chip cookies and cheddar cheese, and two not-as-liked-foods – green olives and lima beans, were tested. In addition to correlating the two response variables with the main effects, interactions of overall expected hedonic liking of the test foods, panelist age, gender, time of day of the test, and day of the week the test were analyzed. Results indicate monetary compensation amount did not influence panelist accuracy or assessment time. However, accuracy increased significantly with longer assessment times; the effect of the "liked" vs. "not-as-well-liked" food categorizations were inconclusive. Expected hedonic liking, gender, age, and time of day of the test, were significant, suggesting that test accuracy and assessment time are likely influenced by multiple intrinsic and extrinsic factors. Overall results suggest that the use of broad demographics of untrained consumers for triangle tests results in data not consistently or strongly biased by payment amount or food type.

1. Introduction

Three unique observations have been made about using humans as instruments in sensory panels: first, panelists change over time; second, panelists as a group are inconsistent; and third, bias greatly affects human panelists (Meilgaard, Civille, & Carr, 2007). A bias in sensory evaluation is defined as anything that influences a panelist in such a way that their scores do not represent the actual sensory experience (Lawless & Heymann, 2010). Researchers minimize biases by using techniques such as, but not limited to, randomization of samples and labeling them with 3-digit blinding codes. While sensory researchers make great efforts to ensure that widely recognized biases are minimized as much as possible, the actual impact of some lesser recognized biases is unknown. Potential sources of bias, such as the amount of panelist reward, either directly through monetary compensation or indirectly through the food product being evaluated, and their possible influencing effect on triangle test accuracy and assessment time has not been studied in detail.

Although monetary payment is commonly used to incentivize panelists to participate in panels and has also been found to effect resulting data, how these results should guide product developers is uncertain. Bell (1993), for example, found that monetarily rewarded and unrewarded panelist groups scored significantly different from each other during hedonic testing when given the exact same foods. He concluded that the presence or lack of monetary incentive influenced consumer scoring; however, not only were conclusions about which scores more accurately represented consumer liking impossible to determine, he noted that even if more accurate results could be obtained from panelists that are not monetarily rewarded, that such payment is an incentive for their participation. Additionally, two other research groups found that in a performance-based incentivized double triangle test, panelists who were paid had greater accuracy in selecting the odd sample, for some foods (Berglund, Lau, & Holm, 1993; Lau, Post, & Kagan, 1995). With regard to how much sensory panelists should be compensated, Lawless and Heymann (2010) warn that

* Corresponding author.
E-mail address: laura_jefferies@byu.edu (L.K. Jefferies).

http://dx.doi.org/10.1016/j.foodqual.2016.09.007
0950-3293/

monetary incentives should only be enough to elicit panelist participation in evaluation, and not so extreme as to make the incentive the only motivation for participation.

Little research has been done to explore the possible effects of the degree of liking or disliking of the food being evaluated on panelist accuracy and assessment time in sensory testing. Even though consumer evaluation is primarily conducted using those who are 'likers' or consumers of a product, not all foods are liked equally among and between individuals. Consequently, the degree of liking of the test food may influence panelists' willingness to evaluate foods conscientiously (Meilgaard et al., 2007).

Much of the literature on identifying general incentivizing factors for sensory panelists or survey participants either fails to conclude the extent of their effects, has explored the role of payment based on panel performance rather than participation alone, or simply does not exist (Bell, 1993; Berglund et al., 1993; Lau et al., 1995; Stone, Bleibaum, & Thomas, 2012). The marked difference in results for rewarded and unrewarded panelists in Bell's study leads to the question of whether increasing levels of monetary reward, without the pressure of "performance" might lead to greater panelist attentiveness to sample differences, especially in instances when not-as-well-liked foods are evaluated, even by willing participants.

In addition to measuring respondent accuracy, some researchers suggest that panelist motivation can be estimated and easily compared by measuring the relative time required to make a judgement (Brüggen & Dholakia, 2010; Zagorsky & Rhoton, 2008). A study of general survey participants found that the highest payment of $40 increased the length of interview time for interview-type surveys and number of items answered for mail surveys compared to a $20 incentive and no incentive (Zagorsky & Rhoton, 2008). Research conducted with web-based surveys also found that motivational rewards increased the number of completed surveys, the words per survey, comments per survey, and the time spent on the survey (Brüggen & Dholakia, 2010). These results suggest a positive relationship between payment amount and data quality, in part due to the increased time taken to complete the surveys, in these contexts. Trained sensory panelists report that among the most influential motivational factors for participation is extra income (Lund, Jones, & Spanitz, 2009), but how payment affects the amount of time they take during testing, or their accuracy rate with discrimination tests, is unknown.

The objective of this study was to evaluate the effects of potential incentivizing biases --stemming from monetary incentive amount and food type and their potential interaction on untrained panelists, by measuring triangle test accuracy and panelist assessment time. The effects of panelist age and gender, overall hedonic scores of the test foods, time of day of the test, and day of the week in which the test was conducted, were also studied. It was hypothesized that increased payment amount and evaluation of well-liked foods would lead to increased accuracy and assessment time.

2. Methods

Three preliminary steps were conducted in preparation for the main study triangle tests. The first was to survey a general consumer populace regarding their past acceptance of a wide variety of foods of expected differences in liking, using an online survey. From these results, two liked and two not-as-liked foods were selected. Then, two samples were selected for each food with intent that they differ. Each pair was validated for difference using a two-step process by presenting them each in their own initial triangle and a difference-from-control tests. After validation, the main study, comprised of a series of triangle tests, was conducted where panelists (in each monetary incentive group) received the same payment for evaluating each food pair. All tests were performed at the Brigham Young University Sensory Lab (BYUSL) (Provo, Utah, U.S.A.); sensory data was collected using Compusense *five*® software (Guelph, Ontario, Canada). The Brigham Young University Institutional Review Board approved all tests.

2.1. Preliminary test 1: determination of test foods

In order to determine two well-liked and two not-as-well-liked foods, two online surveys were completed, via convenience sampling, by members of the BYUSL panelist database. Panelists were asked the following question (without the food to taste) for a variety of food products, "Considering your overall impression of (insert food name here), how much do you like or dislike it/them?" Responses were scored on a discrete 9-point hedonic scale, including a corresponding numerical value (9 = like extremely, 1 = dislike extremely). The average numerical score for each surveyed food was calculated. The first survey asked panelists (n = 1004) to report their degree of liking/disliking for 29 different foods; a subsequent survey (n = 1102) expanded the list of foods to 48. Foods surveyed were presented in a randomized order. Data were collected using Qualtrics, LLC survey software (Provo, Utah, U.S.A.). Panelists were not compensated for survey participation.

Foods that scored a mean of $\geqslant 8$ were categorized as well-liked. Foods with a mean >5 to $\leqslant 6.5$ were considered foods that are generally not-as-well-liked, but not disliked. From this survey, four foods were selected for subsequent study: well-liked – chocolate chip cookies and cheddar cheese, and not-as-well-liked – lima beans and green olives with pimentos. Then, two samples were selected for each food category with intent that they differ. The pair was validated for difference using a two-step process described in s 2.2 and 2.3. Table 1 describes the product treatment differences for each food pair.

2.2. Preliminary test 2: triangle test treatment validation

Treatments for each food pair to be used in the main study were tested in a preliminary step to validate that such differences were enough to be detected by panelists, but were not too obvious. The threshold for validation for all of the food pairs was set at a guideline of $\leqslant 15\%$ proportion of distinguishers (% pd). Since use of % pd can be product-and situation-dependent, this value was selected in consideration of the wide variety of test foods and their individual characteristics; the goal was that samples within each food pair fell within acceptable consumer expectations, and that differences between the pairs were not too obvious (Lawless & Heymann, 2010). Table 1 shows the number of correct responses/number of panelists, and% pd for each food. Triangle tests were conducted as described by Pilgrim and Peryam (1958) with additional specifications.

For each test, 30–35 untrained panelists were selected from the BYUSL panelist database based on their willingness to try the food (Stone et al., 2012) and an absence of allergenicity to it. Untrained panelists were used as trained panelists may have recognized specific or similar training foods as those used in the study, the challenge of finding enough panelists with the breadth of skill required to become trained in four very different food types as each payment group evaluated all foods, and that trained panelists would likely question why payment amounts received in this study may have differed from payment during their training or participation in other panels.

Panelists, who ranged from frequent attendees to brand new, were also recruited based on balanced gender and age groups of 18–29, 30–39, 40–49, 50–59, and 60 years and older. Prior to evaluating the samples, panelists successfully completed a practice triangle test on paper in order to demonstrate that they understood