



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Food Quality and Preference

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

Variations in consumer acceptance, sensory engagement and method practicality across three remote consumer-testing modalities

Marta Albiol Tapia, Soo-Yeun Lee*

Department of Food Science and Human Nutrition, University of Illinois Urbana-Champaign, 905 S. Goodwin Ave, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:

Engagement
Remote testing
Cookies
Live Zoom
Video-guided
Written protocol

ABSTRACT

In the context of the COVID-19 pandemic, it has become challenging for sensory scientists to conduct in-person sensory tests, particularly large central location tests. Sensory literature comparing central location and home use tests shows no clear consensus about how each methodology affects sample ratings and panelist engagement. Research on instructional delivery suggests that the most effective method of increasing engagement involves interactive video conferencing. The objective of this study was to compare three methods of remote consumer testing regarding sample acceptance, sensory engagement, and method practicality. Eighty-four participants rated five chocolate-chip cookie products on a 9-pt hedonic scale in each of three methods: 1) a live (synchronous) Zoom session, 2) an asynchronous video-guided session, and 3) a fully written protocol session. Results showed no significant differences in sample liking pattern across the methods used. Engagement scores approached the limit of significance for the Active Involvement dimension, indicating panelists were least likely to feel distracted, zoned out or lose interest in the written protocol method. There were no significant differences in the time spent on the test by the panelists across the three methods. Asynchronous methods showed to be most suitable in terms of the convenience of the time of day at which the tests were completed, but showed no significant differences in other aspects of method practicality. Overall, a written protocol method of remote consumer testing is recommended, as it is less time-consuming for researchers while providing similar acceptance and engagement as other methods.

1. Introduction

In the context of the COVID-19 pandemic, it has become challenging for sensory scientists to conduct in-person sensory tests, particularly large central location tests. The COVID-19 virus is thought to spread through respiratory droplets between people who are within 6 feet from each other (Centers for Disease Control and Prevention, 2020), even in people who are asymptomatic. The use of masks and social distancing have been enforced in public settings, and consequently, many businesses have transferred to remote/online platforms. Given the fact that sensory tests usually involve panelists smelling and tasting food samples, the food industry has had to quickly overcome many challenges to be able to conduct sensory tests. Home-use tests (HUT) offer a good alternative to in-person testing; however, they present numerous challenges including the inability to control experimental conditions as well as the limitations in terms of shelf life and packaging.

Many research studies have compared sensory perception and consumer acceptance of different samples in central location test (CLT) and

HUT. Some of these studies show no effect of testing location (Lee & Lee, 2021; Schouteten, Gellynck, & Slabbinck, 2021), whereas others have observed a significant effect (Schouteten, Gellynck, & Slabbinck, 2019; Sveinsdóttir et al., 2010; Zhang, Jo, Lopetcharat, & Drake, 2020). Inconsistencies shown in previous literature may be due to the type of products tested and also to the level at which a given food or beverage product is associated to a specific context in consumer's minds. A previous study comparing HUT, CLT, and laboratory test found no difference in attribute liking, but did find differences in the perceived intensity of certain attributes (Pound, Duizer, & McDowell, 2000). Another study that compared CLT and HUT found different results depending on the product type and its usual context of consumption (Boutrolle, Delarue, Arranz, Rogeaux, & Köster, 2007). Previous studies that compare these two methods show no consensus about how each methodology affects sample ratings. Some studies have even shown that consumers display a higher discrimination (McDaniel & Sawyer, 1981) and are more critical of certain attributes (Miller et al., 1995) in home panels compared to laboratory panels. Another study showed that CLT

* Corresponding author.

E-mail address: soolee@illinois.edu (S.-Y. Lee).

<https://doi.org/10.1016/j.foodqual.2022.104616>

Received 20 December 2021; Received in revised form 14 April 2022; Accepted 21 April 2022

Available online 25 April 2022

0950-3293/© 2022 Elsevier Ltd. All rights reserved.

yield more robust results than HUT, and lower average liking scores (Boutrolle, Arranz, Rogeaux, & Delarue, 2005). Recent studies that have adopted the approach of online/remote testing have focused on viewing photographs instead of tasting samples (Oliveira e Silva, do Carmo Rouxinol, & da Silva Coutinho Patarata, 2020), so comparison to previous literature is limited.

This scientific discipline has had to find a way to suddenly shift its standard practices, so it is valuable for the food industry to have a method comparison between supervised/synchronous and unsupervised/asynchronous online consumer tests. It is essential to collect data on sensory engagement and consumer behavior in each of the studied methods to evaluate which type of treatment assists panelists best in their evaluation of samples. This involves panelists not getting distracted or losing interest in the sample evaluation, as well as finding the task meaningful and enjoyable. Previous research on instructional delivery suggests that the most effective method of increasing engagement involves instructor visibility through interactive video conferencing (Carr, 2014). Sensory protocol also determines that testing conditions and execution can be a source of variability if there is a lack of control (Moskowitz, Munoz, & Gacula, Maximo, 2003).

The objective of this study was to compare three methods of online consumer testing in terms of sample perception, sensory engagement, and method practicality. It was hypothesized that a live, facilitated method of remote consumer testing will result in: 1) less variability across panelists, due to testing conditions being monitored live by the researcher; 2) a higher degree of sensory engagement due to a lower likelihood of losing interest during the task, and 3) a lower degree of practicality overall, due to the inconvenience of being scheduled for a testing session, in comparison to asynchronous written and video-guided methods of remote testing.

2. Material and methods

2.1. Samples and subjects

Five commercially available chocolate-chip cookie products were selected as test samples. This product category was selected due to its popularity and wide product variation. All five cookie products contained semi-sweet chocolate chips. More sample information can be seen in Table 1.

Eighty-four consumers were recruited through email, social media, and university campus newsletters. The participation requirements were: 1) be 18 years of age or older, 2) have no known food/beverage allergies, 3) have no dietary restrictions, 4) not be pregnant or nursing, 5) not have any pre-existing or new smell/taste disorders, 6) have access to the Internet, a computer with a working microphone and camera, and a Zoom account, 7) be a frequent consumer of cookies, which was defined as consuming cookies more than once a month on average, 8) be on or near the University of Illinois Urbana-Champaign (UIUC) campus to pick up sample kits, and 9) be available for four remote sessions. Requirement #5 (not have pre-existing or new smell/taste disorders), was included as a means to screen out any consumers that may have been infected with COVID-19 or may still be recovering from the infection. Recent studies have demonstrated that loss of taste and smell are strong predictors of COVID-19 infection (Dawson et al., 2021; Menni

Table 1
Product information for the five cookie products tested.

Sample	Weight per cookie (g)	Calories per cookie	Type of brand	Price per 357–397 g of cookies (\$)
C1	14	70	National	9.58
C2	11	53.3	Generic	2.29
C3	14	70	National	6.58
C4	15	80	National	3.49
C5	11	53.3	National	2.89

et al., 2020; Printza & Constantinidis, 2020). In addition to meeting the screening requirements, panelists had to show proof of a recent negative COVID test in order to enter the building where they were instructed to pick up their sample kit. Due to the number of strict participation requirements, the researchers were only able to recruit eighty-four panelists. However, there is a certain range for subject number in published literature. Depending on the experimental design and nature of each consumer test, the number of participants varies from around eighty (Giacalone et al., 2019; Hall, Hampl, Stanton, & Monte, 2003) to several hundred (Cardello, Llobell, Giacalone, Roigard, & Jaeger, 2022; Crisosto, Crisosto, & Metheny, 2003). The demographic characteristics of this study's participants can be seen in Table 2.

Panelists received a gift card payment for their participation in the study. This research was reviewed and approved by the Institutional Review Board at the University of Illinois Urbana-Champaign, and informed consent was obtained from each subject prior to their participation in the study.

2.2. Test variables

The study consisted of four remote sessions in a time span of two weeks. The first session did not involve tasting; it was an introductory session aimed at providing an overview of the testing sessions and ensuring the panelists' cameras were in working condition. Additionally, the researcher commented on general instructions and expectations of the study and provided details on how to pick up the sample kits. Lastly, each panelist signed a consent form and completed a demographic questionnaire. This was the first session for all panelists, and

Table 2
Demographic information of subjects.

		Count (n = 84)	Percentage	
Gender	Female	64	76.2	
	Male	19	22.6	
	Non-binary / third gender	1	1.2	
Age	Prefer not to say	0	0.0	
	18–25 years old	16	19.0	
	26–35 years old	23	27.4	
	36–45 years old	25	29.8	
	46–55 years old	14	16.7	
	56–65 years old	4	4.8	
Ethnicity	66–75 years old	2	2.4	
	76 years old or older	0	0.0	
	Hispanic or Latino	3	3.6	
	Not Hispanic or Latino	81	96.4	
	Racial background	I don't know / I prefer not to say	0	0.0
		American Indian or Alaskan Native	0	0.0
Asian or Asian American		17	20.2	
Black or African American		1	1.2	
Native Hawaiian or Pacific Islander		0	0.0	
White or Caucasian		62	73.8	
Frequency of consumption of cookies (any type)	Mixed or Other	2	2.4	
	I don't know / I prefer not to say	2	2.4	
	Daily	10	11.9	
	Weekly	47	56.0	
	Monthly	27	32.1	
	Frequency of consumption of chocolate-chip cookies	Daily	3	3.6
Weekly		26	31.0	
Monthly		45	53.6	
Every four to six months		10	11.9	
Once a year or less		0	0.0	
Never		0	0.0	

it was held on the video-conferencing platform Zoom (Zoom Video Communications, Inc., San Jose, CA, USA). All electronic questionnaires were completed on Qualtrics (Qualtrics, Provo, UT, USA), an online survey software.

The other three sessions involved the sensory evaluation of the same five cookie samples, with different instructional delivery methods. The sessions were randomized across panelists, and samples were labeled with different three-digit codes in each session. The first type of testing session was a synchronous, facilitated session through Zoom. In this session, the researcher provided live, step-by-step instructions to all panelists on completing the sensory evaluation of the samples. This included a brief welcome, an explanation of how to follow the rinse protocol, a description of the hedonic scale, and a short introduction to each evaluation question. The researcher tried adapting the pace of the test to all panelists. Given the fact that it was required that all subjects had their cameras on during the entire session, it was possible to know when most people had finished tasting each sample. Participants were free to ask questions at any time during the session. Panelists were scheduled for one of nine live testing sessions according to their availability, which means there were between seven and eleven panelists in each live Zoom session. The researcher followed the same script in all sessions.

The second session type was an asynchronous video-guided session. Panelists received their questionnaire link by email and were given 24 h to complete the sensory test at their preferred time of day. This questionnaire had short, embedded videos that were meant to walk the panelists through the sensory evaluation process. The researcher was the same as for the live sessions, and the videos were recorded in the same background as the one used for the live Zoom sessions. The researcher was also wearing a lab coat and gloves in all videos and all live sessions, to maintain consistency and to ensure that the delivery method was the only variable.

The third and last type of delivery method was an asynchronous, fully written protocol. The questionnaire was the same as the other two sessions, but this time the instructions were typed in detail, with no visual aids such as videos or interactive conference calls. The typed instructions were the same as the ones provided verbally by the researcher in the live session, and taped in the video-guided session.

2.3. Test protocol

Cookie products were purchased from nearby grocery stores and opened the same day the sample kits were prepared. One cookie from each one of the five brands used was placed in 163 mL clear plastic cups with lids (Dart Container Corporation, Mason, MI, USA), labeled with random three-digit codes. Five unsalted crackers were placed in a separate cup. The six cups were placed in a gallon resealable plastic bag (Ziploc, S.C. Johnson & Sons Inc., Racine, WI, USA) together with a 207 mL wax-coated paper cup (Solo Cup Company, Lake Forest, IL, USA) and a 9-cm × 12-cm white dispenser napkin. One bag was prepared for each panelist two days before each one of their testing sessions. The day after the sample bags were prepared, panelists picked them up from a designated room in a building on the university campus. Panelists were given a 10-hour window to pick up their sample bags on each one of the three designated pick-up days. The next day, they would have one of their three testing sessions (live, video-guided, or written protocol). The process was repeated until all panelists had completed all three session types, which were randomized across the panelists.

2.4. Sensory evaluation procedure

Panelists were instructed to complete each test in one sitting, so they were asked not to begin the test until they were ready to complete it. They were asked to have all the necessary materials in front of them before beginning the test, and, if possible, to choose a quiet, distraction-free environment to complete the session. They were also reminded of

the importance of matching the code in the screen to the code in the cup for each question.

Panelists were also instructed to follow the rinse protocol before tasting each cookie sample. The rinse protocol consisted of biting half a cracker, chewing and swallowing, and then taking a sip of room temperature water. They were also instructed to not consume the whole sample, but to leave approximately half for re-tasting. After typing in their start time, panelists would rate each of the five samples on overall liking on a modified 9-point hedonic scale, with 1 being 'dislike extremely', 9 being 'like extremely', and 5 being 'neither like nor dislike'. The order of appearance of the overall liking questions was randomized among panelists. After completing this set of questions, panelists moved on to evaluating each sample on liking of aroma, taste, and texture, also on a 9-point hedonic scale. To do this, panelists re-tasted each sample. After typing in their end time, panelists completed a validated engagement questionnaire (Hannum & Simons, 2020), where they rated their level of agreement or disagreement with a series of ten statements measuring three different dimensions of sensory engagement on a 7-point scale. Panelists also answered questions about their impressions of the practicality of each one of the methods used. Lastly, panelists were given the opportunity to type in any comments about the test in a text box. The full research questionnaire is available as supplementary material.

2.5. Data analysis

XLSTAT Sensory version 2019.1.1 (Addinsoft, NY, USA) was the data analysis software used to identify significant differences across samples evaluated using different remote testing methods. Analysis of Variance (ANOVA) was used to analyze consumer ratings with 'panelist' as a random factor and 'testing method', and 'sample' as fixed factors. Interaction among test types and samples was explored for significant impact on liking scores. Post hoc tests were conducted by the Fisher Least Significant Difference (LSD) method, with a confidence interval of 95%.

To group subjects into statistically similar clusters according to their overall liking ratings of each product, Agglomerative Hierarchical Clustering (AHC) was used with the Euclidean distance for the dissimilarity scale by Ward's method. The cluster truncation was automatic based on entropy. Data from the engagement questionnaire were averaged to generate three types of dimensional scores for each type of testing method: 1) Active Involvement, 2) Purposeful Intent, and 3) Affective Value. The questions that were part of the Active Involvement dimension were reverse coded so that high values would indicate an equal response for each item. A two-way ANOVA with 'panelist' and 'testing method' as main effects was conducted for each dimension to determine any significant effects of the testing method on sensory engagement. Agglomerative Hierarchical Clustering was also conducted on engagement scores to categorize panelists according to their level of overall engagement in each one of the three testing methods. The scores for each engagement dimension were averaged to calculate an average factor score. Analysis of Variance was run on the data extracted from each cluster, and post-hoc Fisher LSD tests were conducted to identify significant differences within (not across) each cluster. Clusters formed by a single panelist were combined in a logical way following the corresponding dendrogram. Ratings from the method practicality questionnaire were analyzed using ANOVA, with 'panelist' as a random factor, and 'method' and 'time of day' as fixed factors. 'Time of day' was included as a factor in order to determine any differences in liking according to the time chosen by panelists to complete the test. Fisher LSD was conducted to resolve the direction of any significant effects and was set to a significance level of 0.05.

3. Results

3.1. Product acceptance

Based on the Analysis of Variance shown in Table 3, the variable 'testing method' did not bring significant information to explain the variability of the dependent variables of overall liking, or liking of any of the specific attributes tested. There was also no significant interaction of 'testing method' with 'sample', as can be seen in Fig. 1, which displays acceptance scores of each sample in the different modalities used. The average acceptance scores for each sample can be seen in Fig. 2, where on average, sample C3 was liked the most, and samples C2 and C4 were liked the least.

Panelists were clustered according to their overall liking ratings across testing methods. The results of the AHC analysis (Fig. 3) identified three groups of panelists from their taste preferences, which varied slightly across testing methods. The major cluster in every testing method (Cluster 3) had a similar level of acceptance for all samples, where the average overall liking rating was approximately between 6 and 7.5. The major cluster represented 39% of panelists in the live Zoom method, 54% in the video-guided method and 56% in the written protocol method. Cluster 2 was also similar in all methods, where approximately a third of the panelists rated samples C1 and C3 higher than the rest of the samples. The testing methods varied primarily in their smallest cluster (Cluster 1). In the written protocol session, panelists in Cluster 1 displayed a high degree of hedonic discrimination across samples, with C1 being the least liked and C5 being the most liked. The smallest cluster in the live Zoom testing session exhibited a low degree of hedonic discrimination among samples C2, C3 and C4, with C1 remaining the least liked and C5 the most liked. Lastly, the smallest cluster in the video-guided session was characterized by panelists who liked sample C3 the most and sample C1 the least. Therefore, Cluster 1 in each of the analyses represented a percentage of consumers who did change their preferences based on the testing modality, which ranged from 11% in the written protocol method to 27% in the live Zoom method.

3.2. Sensory engagement

Results of the Engagement Questionnaire (EQ) indicated that sensory engagement was not significantly affected by the testing method, as shown in Table 4. The 'Active Involvement' dimension closely approached the limit of significance ($p = 0.069$), where the written protocol method received higher engagement scores (mean of 6.59) than the other methods (mean of 6.51 for live Zoom and 6.31 for video-guided).

When panelists were clustered according to their level of overall engagement in each testing method, three clusters appeared (Fig. 4). The largest cluster (Cluster 3) was characterized by panelists who displayed a high level of engagement in all methods, with a mean of 6.7 out of 7. In the second to largest cluster (Cluster 2), panelists also displayed a similar degree of engagement in all three testing methods, with the average score being lower than the largest cluster (5.9 out of 7). The last cluster (Cluster 1), comprised of only seven panelists, displayed a higher

Table 3
Analysis of Variance (ANOVA) F-values of overall liking and specific attribute scores.

Attribute Liking	Panelist (P) df = 83	Testing Method (M) df = 2	Sample (S) df = 4	M*S df = 8
Overall	4.26***	0.87	22.87***	0.39
Aroma	5.92***	2.27	22.69***	0.43
Taste	4.86***	0.46	29.68***	0.64
Texture	6.73***	1.36	16.59***	0.26

*, **, and *** indicate significance at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively; df = degrees of freedom.

engagement in the live Zoom session compared to the other two sessions, though not significantly higher upon Fisher LSD analysis. Cluster 1 displays large error bars given the small number of panelists included in this cluster. It also demonstrates that only a small percentage of the subjects were not as engaged during their participation in this study. The majority of panelists displayed high sensory engagement scores overall.

3.3. Method practicality

The three testing methods proved to be similar regarding overall practicality (Table 5). The only variable that showed a significant difference across methods was the 'convenience of the time of day'. Both asynchronous methods (video-guided and written protocol) were rated significantly higher in convenience than the synchronous method through Zoom. The time of day that showed to be most convenient for panelists to complete the remote consumer test was the evening, with an average convenience rating of 6.8 out of 7.

Fisher LSD analysis showed that the time spent on each session was not significantly different across the three methods, with an average of 13.8 min spent on each questionnaire.

4. Discussion

4.1. Product acceptance

Acceptance of the cookie products did not differ across the three remote testing methods. This outcome was expected, as previous research from this laboratory has shown acceptance to be rather robust and absolute, and not greatly be affected by context effect (Albiol Tapia, Baik, & Lee, 2021), which in this study was created by the different delivery methods. There were also no differences in overall liking of samples across session orders.

Comparison to previous literature is limited, as not many research studies have explored different delivery methods of remote consumer testing with the same samples. Most studies focus on comparing test locations, such as comparing CLT with HUT. Home Use Test, in comparison to CLT, is generally characterized by natural (though uncontrolled) eating conditions, pleasant ambiance, consumption of a larger amount of food, repeated exposures, and a multi-session monadic sequential presentation of products (Boutrolle et al., 2007). Additionally, HUT allow consumers to choose the moment of consumption as well as the way the product is consumed, the possibility of a social experience and the chance of allowing post-ingestional effects before providing an opinion of the food (Boutrolle et al., 2007). In this study, the live Zoom method had aspects of a CLT in that the researcher could ensure that the samples were being tasted and the rinse protocol was being followed, since panelists were performing the sensory evaluation live on camera. In addition, panelists were not completely free to choose the time of day to complete the test, as they were scheduled for a session ten days prior, based on their reported availability in the screening questionnaire. The two other methods (video-guided and written protocol sessions) can be described as closely following the protocols of a HUT, in the sense that the consumers are able to choose an appropriate time of day to consume the food, and the conditions in which they decide to complete the sensory evaluation are completely uncontrolled. Previous studies comparing CLT to HUT have resulted in different outcomes. For products like fermented milk beverages (Boutrolle et al., 2005), salted crackers and sparkling water (Boutrolle et al., 2007), HUT results showed higher acceptance ratings than CLT. Studies testing other snack products showed product-specific results, where chips were liked more in HUT, juice was liked more in CLT, and yogurt showed no contextual differences (Schouteten et al., 2019). In another study, consumers that tested cod also averaged higher liking results in HUT setting, although the scores were comparable when the cooking method was similar to the method used in CLT (Sveinsdóttir et al., 2010).

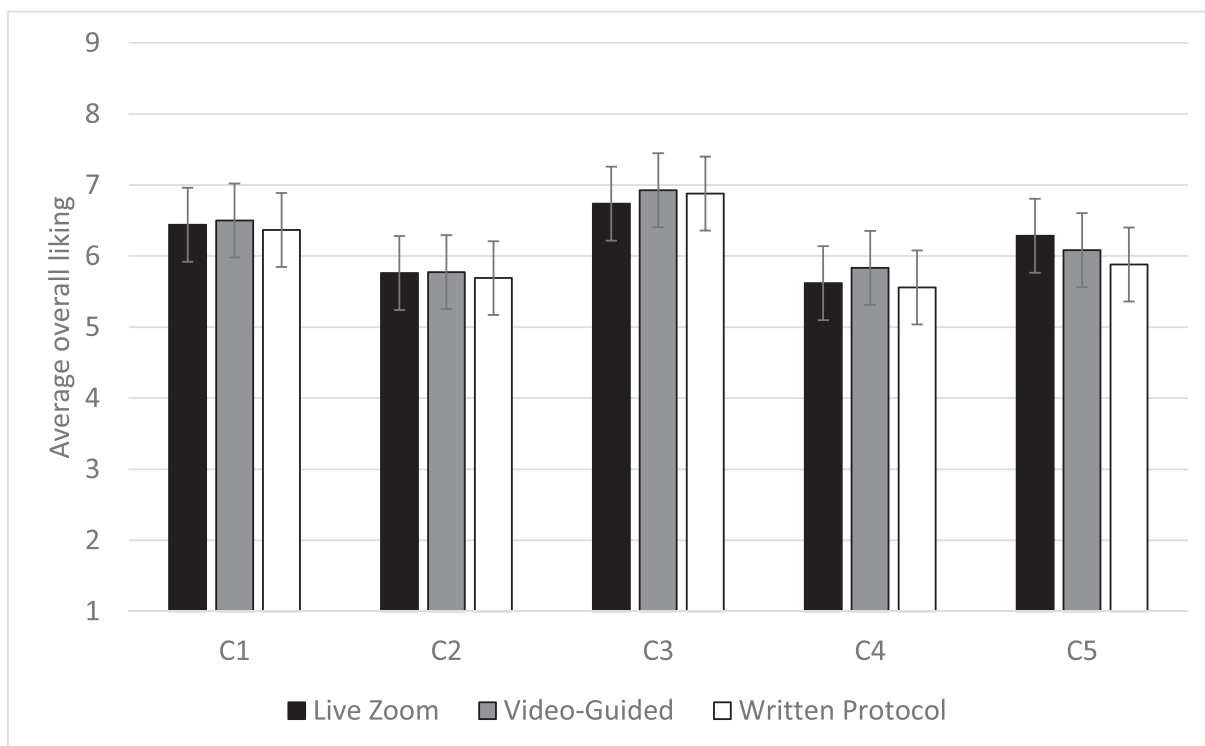


Fig. 1. Average overall liking ratings on 9-pt hedonic scale for each sample tested in each of the three testing modalities. Sample*Method interaction not significant ($p = 0.925$).

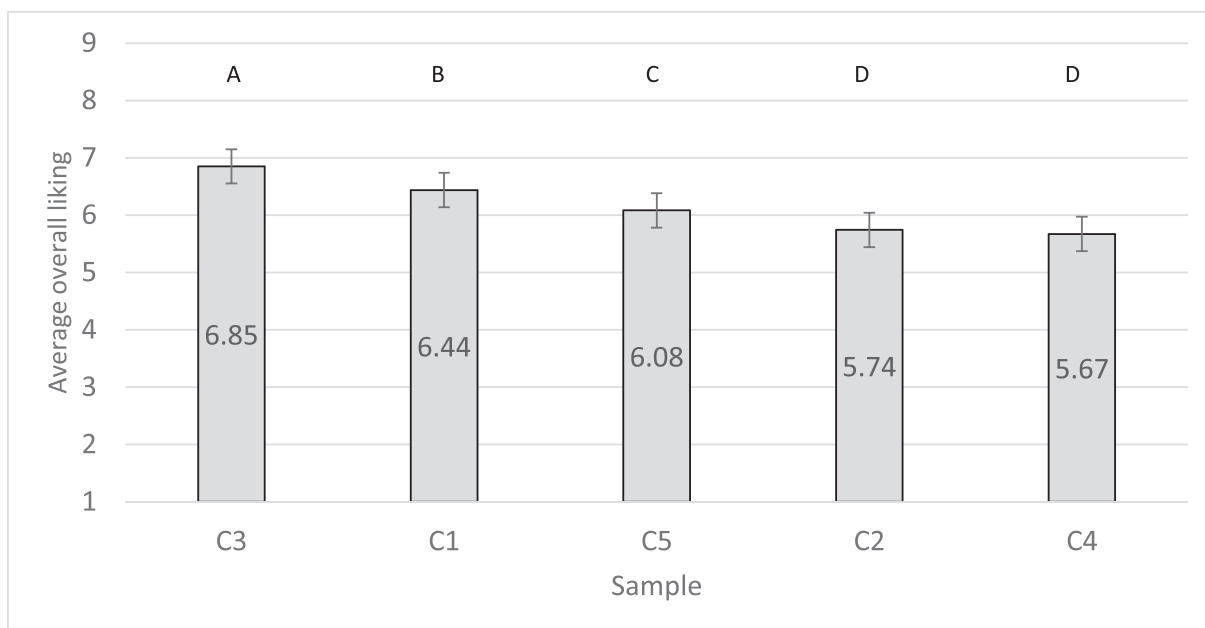


Fig. 2. Average overall liking ratings on 9-pt hedonic scale for each sample tested. Means labeled with the same letter indicate no significant differences in overall liking (Fisher Least Significant Difference).

4.2. Sensory engagement

No significant differences in sensory engagement were found across testing methods. A potential reason for this outcome is that the testing methods were very standardized, with the delivery of instructions being the only difference. This was designed purposefully to test the delivery method as the only variable. The ‘Active Involvement’ dimension, which closely approached the limit of significance, showed that overall,

panelists felt less ‘zoned out’ in the written protocol session. This may be because to receive the instructions, panelists had to actively read text on their screens, as opposed to passively listen in the video-guided and live Zoom formats. Once the panelist was familiar with the sensory evaluation dynamics, it may have been easier to lose focus in the video-guided and live sessions. The ability to skim or even skip instructions is also higher in the written protocol session compared to the video-guided session. Panelists may have opted for reading at a faster pace after

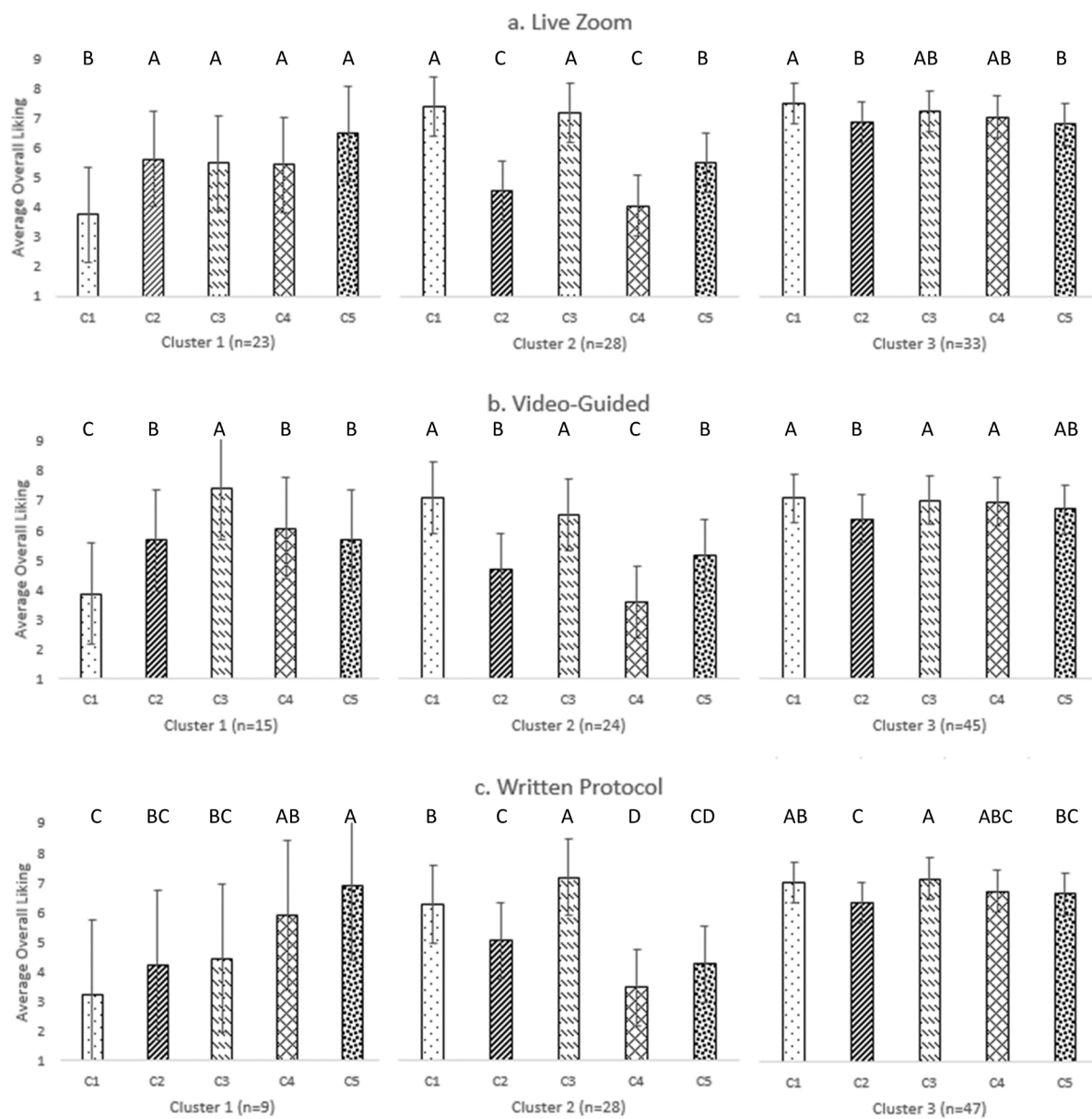


Fig. 3. Agglomerative Hierarchical Clustering of overall liking scores. C1-C5 = Cookie samples 1 through 5 for a) Live Zoom, b) Video-Guided, and c) Written Protocol methods. Means labeled with the same letter indicate no significant differences in overall liking (Fisher Least Significant Difference).

being familiarized with the instructions. When looking solely at engagement scores for panelists' first testing session, we observe lower scores in the 'purposeful intent' dimension, which assesses panelists' perceived personal relevance of the sensory evaluation. This effect was driven by the video-guided session, which received higher engagement scores when it was the second or third session for panelists.

Previous research comparing different types of instructional delivery is mostly focused on student learning, especially in children. Some studies found that, during teacher-directed instruction, children were less likely to attend to instructions and to ask questions. During seat work, however, they were more likely to get disorganized (Stright & Supplee, 2010). In research on adults, our results conflict with two of Richard Mayer's principles of multimedia design, which involve replacing visual text with spoken text (modality principle), and adding visual cues relating elements of a picture to the text (multimedia principle) (Mayer, 1999). These principles have been shown to increase the effectiveness of multimedia instructions regarding less mental effort spent on understanding explanations. A study testing these principles on college students, however, found lower retention and transfer scores

when visual text was replaced by spoken text (Tabbers, Martens, & van Merriënboer, 2004). This study argued that visual text may be more appropriate for presenting procedural information (which in the case of the present study, can apply to the instructions on how to proceed with the sensory test) than spoken test, because the subject has more time to reflect on the information presented. It is likely that a written protocol method accompanied with visual images of the instructions would have resulted in higher engagement than the video-guided method.

Another reason for the lower engagement scores in the live Zoom method could be that panelists disliked being on camera, or were distracted by watching other panelists on camera. Several panelists reported this at the end of their questionnaires, as there was a text box for them to type any comments they had about each session. The video-conferencing software used does not offer an option where the host can view all participants' video, but the participants can only view the host's video. This feature would be ideal to use in remote consumer tests, as it would allow the researcher to monitor test performance while preventing the bias that comes from panelists viewing each other on camera. It was also reported that the videos were somewhat repetitive, and

Table 4
Analysis of Variance (ANOVA) F-values Engagement scores.

Engagement Dimension	Questions included in each dimension	Panelist (P) df = 83	Testing Method (M) df = 2
Active Involvement	I lost interest in the task [‡]	2.38***	2.72
	I was distracted [‡]		
Purposeful Intent	I felt myself zoning out during the task [‡]	5.57***	1.23
	I found the task meaningful		
	I felt dedicated to finish the task		
Affective Value	I wanted to devote my full attention to the task	6.79***	0.88
	My contribution was significant to the outcome of the task		
	I found the task captivating		
	During the task, I was enjoying myself		
	I was motivated to expend extra effort during the task		

*, **, and *** indicate significance at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively; df = degrees of freedom; ‡ indicates a question was reverse coded for analysis.

that a written protocol encouraged the panelists to focus more. Additionally, ‘Zoom fatigue’ could also explain why engagement was not higher in the live session. Videoconferencing technologies have allowed students to continue their education remotely during the pandemic and have also enhanced our ability to connect when in-person gatherings were discouraged. However, the prolonged use of this technology has also reported drawbacks. Examples include a greater need to concentrate, difficulties relaxing into natural conversations, expectations of instant responses, and even increased self-evaluation from continuously staring at video of oneself, which could lead to loss of self-esteem (Bailenson, 2021; Wiederhold, 2020; Williams, 2021).

4.3. Method practicality

Given that the only significant difference found in the analysis of method practicality questions was the convenience of the time of day, the authors recommend an asynchronous method for remote consumer tests of products like cookies, which do not need any form of cooking and can be easily portioned into sample cups. The evening, defined by the researchers as 6 pm to midnight, resulted in the highest convenience ratings, with a mean score of 6.76 out of 7. Table 6 shows panelists’ choices for the time of day to complete each asynchronous testing session. It is notable that, even though more panelists chose the morning to complete the test, the evening received higher convenience ratings.

Even though the focus of this study is to maximize the method

practicality for the subjects, it is worth noting that the practicality for the sensory scientist conducting the research is also much higher in a fully written protocol method. To prepare for this session, the researcher is not required to spend time pre-recording, repeating, editing, uploading and embedding videos in the research questionnaires, nor does the researcher have to schedule panelists according to their availability. In addition, it seems that panelists did not rush the evaluations in the asynchronous sessions, as the average time spent on each session did not vary significantly. Overall, a written protocol session that includes some images supporting the instructions would be an engaging, practical, and effective method of remote consumer testing.

5. Conclusions

The method of remote consumer testing that proved to be the most practical in this study was an asynchronous, fully written protocol

Table 5
Analysis of Variance (ANOVA) F-values of practicality questionnaire.

Practicality Question	Panelist (P) df = 83	Testing Method (M) df = 2	Time of Day (T) df = 3	M*T df = 4
Easiness of following instructions	1.10	0.07	0.74	0.29
Clarity of instructions	1.36	0.11	0.57	1.47
Adequacy of session length	1.07	1.14	0.72	1.37
Convenience of time of day	2.12***	15.09***	3.79*	0.90
Rating of experience	1.90***	0.79	0.93	0.49

*, **, and *** indicate significance at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively; df = degrees of freedom.

Table 6
Choice of time of day for the completion of asynchronous sessions.

Asynchronous Testing Method	Chosen time of day [‡]	Number of panelists	Percentage
Video-Guided	Morning	37	44
	Afternoon	30	36
	Evening	16	19
	Night	1	1
Written Protocol	Morning	32	38
	Afternoon	26	31
	Evening	25	30
	Night	1	1

[‡] Morning: 6 am–12 pm, Afternoon: 12 pm–6 pm, Evening: 6 pm–12 am, Night: 12 am–6 am.

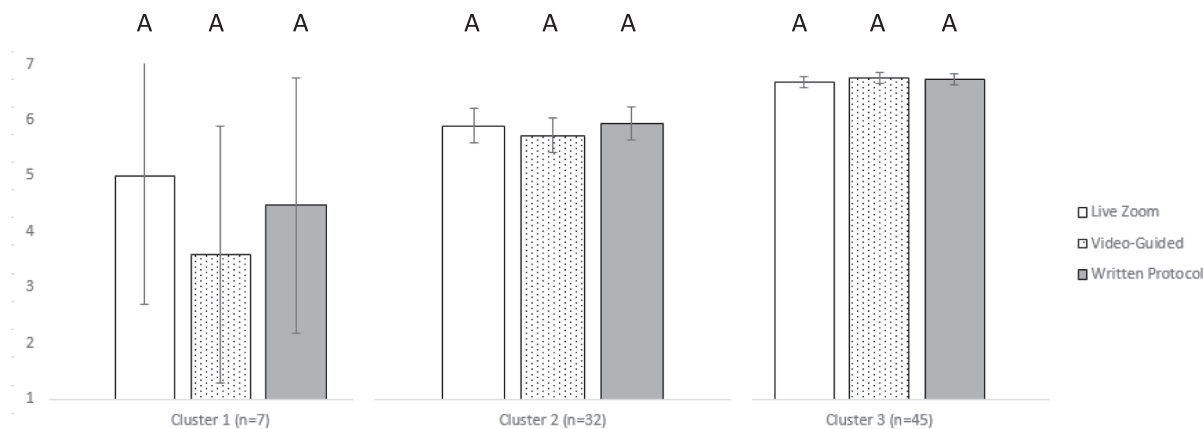


Fig. 4. Agglomerative Hierarchical Clustering of engagement scores (1 = strongly disagree to 7 = strongly agree). Means labeled with the same letter indicate no significant differences in overall liking (Fisher Least Significant Difference).

method. With similar product acceptance ratings as well as sensory engagement scores, this method scored highest in convenience of time of day. A written protocol method also has the most benefits for the sensory scientist, who will not need to schedule panelists for live videoconferencing or pre-record instructional videos to embed them in the research questionnaires. These findings allow sensory scientists in academia and industry to justify the use of asynchronous methods of remote consumer testing during the pandemic. This study also proves that safety can be prioritized over controlled conditions of testing, and that remote testing using a detailed, written protocol can be considered an alternative until in-person testing is deemed safe again.

The main limitation of this study is that results were not compared to different methods of in-person testing, so it is unclear whether acceptance, engagement and practicality would be comparable when tasting chocolate-chip cookies in an in-person setting. In addition, the practicality questions that were used in this study are not part of a validated questionnaire, so it is unclear whether the questions chosen by the researchers were the most appropriate to address subjects' practicality experiences. Lastly, the size of the population used in this research study is also limited ($n = 84$). The objective of this research study is methodological in nature, and as such not meant to be representative of the general population. Additionally, published literature ranges widely in participant number for consumer tests from around 80 to several hundred, so our sample size fits within the published range. Future research can include an additional method consisting of a written protocol with visual images (still or animated) that aid in the description of the instructions. Ultimately, a comparison of scores of in-person and remote testing will be key in determining whether these methods can be comparable for different food products.

CRedit authorship contribution statement

Marta Albiol Tapia: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Visualization, Writing – original draft. **Soo-Yeun Lee:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks to Taylor Wen for her assistance in the study execution. The authors thank Youngsoo Lee, M. Yanina Pepino, and Shelly J. Schmidt for their advice on the study design.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodqual.2022.104616>.

References

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, and Behavior*, 2(1). <https://doi.org/10.1037/TMB0000030>

Boutrolle, I., Arranz, D., Rogeaux, M., & Delarue, J. (2005). Comparing central location test and home use test results: Application of a new criterion. *Food Quality and Preference*, 16, 704–713. <https://doi.org/10.1016/j.foodqual.2005.03.015>

Boutrolle, I., Delarue, J., Arranz, D., Rogeaux, M., & Köster, E. P. (2007). Central location test vs. home use test: Contrasting results depending on product type. *Food Quality and Preference*, 18, 490–499. <https://doi.org/10.1016/j.foodqual.2006.06.003>

Cardello, A. V., Llobell, F., Giacalone, D., Roigard, C. M., & Jaeger, S. R. (2022). Plant-based alternatives vs dairy milk: Consumer segments and their sensory, emotional, cognitive and situational use responses to tasted products. *Food Quality and Preference*, 100(March), Article 104599. <https://doi.org/10.1016/j.foodqual.2022.104599>

Carr, M. (2014). The online university classroom : One perspective for effective student engagement and teaching in an online environment. *The Journal of Effective Teaching*, 14(1), 99–110.

Centers for Disease Control and Prevention. (2020). How to Protect Yourself & Others. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>.

Crisosto, C. H., Crisosto, G. M., & Metheny, P. (2003). Consumer acceptance of “Brooks” and “Bing” cherries is mainly dependent on fruit SSC and visual skin color. *Postharvest Biology and Technology*, 28(1), 159–167. [https://doi.org/10.1016/S0925-5214\(02\)00173-4](https://doi.org/10.1016/S0925-5214(02)00173-4)

Dawson, P., Rabold, E. M., Laws, R. L., Connors, E. E., Gharpure, R., Yin, S., ... Kirking, H. L. (2021). Loss of taste and smell as distinguishing symptoms of coronavirus disease 2019. *Clinical Infectious Diseases*, 48, 18–21. <https://doi.org/10.1093/cid/ciaa799>

Giacalone, D., Degn, T. K., Yang, N., Liu, C., Fisk, I., & Münchow, M. (2019). Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception. *Food Quality and Preference*, 71(July 2017), 463–474. [10.1016/j.foodqual.2018.03.009](https://doi.org/10.1016/j.foodqual.2018.03.009).

Hall, R., Hampf, J. S., Stanton, J. V., & Monte, W. C. (2003). Acceptability of mesquite as a flavoring agent among Native Americans in Arizona. *Journal of Sensory Studies*, 18(4), 269–275. <https://doi.org/10.1111/j.1745-459X.2003.tb00389.x>

Hannum, M. E., & Simons, C. T. (2020). Development of the engagement questionnaire (EQ): A tool to measure panelist engagement during sensory and consumer evaluations. *Food Quality and Preference*, 81(November 2019), 103840. [10.1016/j.foodqual.2019.103840](https://doi.org/10.1016/j.foodqual.2019.103840).

Lee, N., & Lee, J. (2021). Comparison of home use tests with differing time and order controls. *Foods*, 10(6). <https://doi.org/10.3390/foods10061275>

Mayer, R. E. (1999). Research-based principles for the design of instructional messages. *Document Design*, 1(1), 7–19. <https://doi.org/10.1075/dd.1.1.02may>

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

Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., ... Spector, T. D. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*, 26, 1037–1040.

Miller, M. F., Hoover, L. C., Cook, K. D., Guerra, A. L., Huffman, K. L., Tinney, K. S., ... Huffman, L. M. (1995). Consumer acceptability of beef steak tenderness in the home and restaurant. *Journal of Food Science*, 60(5), 963–965.

Moskowitz, H. R., Munoz, A. M., & Gacula, M. C. J. (2003). *Viewpoints and controversies in sensory science and consumer product testing*. Trumbull, Connecticut, USA: Food & Nutrition Press INC.

Oliveira e Silva, R., do Carmo Rouxinol, M. I. F., & da Silva Coutinho Patarata, L. A. (2020). The use of photography to perform an online consumer test on the freshness of chicken breast and the extension of shelf life. *Journal of Sensory Studies*, 35(3), 0887–8250. <https://doi.org/10.1111/joss.12565>

Pound, C., Duizer, L., & McDowell, K. (2000). Improved consumer product development. Part one: Is a laboratory necessary to assess consumer opinion? *British Food Journal*, 102(11), 810–820.

Printza, A., & Constantinidis, J. (2020). The role of self-reported smell and taste disorders in suspected COVID-19. *European Archives of Oto-Rhino-Laryngology*, 277(9), 2625–2630. <https://doi.org/10.1007/s00405-020-06069-6>

Schouteten, J. J., Gellynck, X., & Slabbinck, H. (2019). Influence of organic labels on consumer's flavor perception and emotional profiling: Comparison between a central location test and home-use-test. *Food Research International*, 116(February 2019), 1000–1009. [10.1016/j.foodres.2018.09.038](https://doi.org/10.1016/j.foodres.2018.09.038).

Schouteten, J. J., Gellynck, X., & Slabbinck, H. (2021). Do fair trade labels bias consumers' perceptions of food products? A comparison between a central location test and home-use test. *Sustainability (Switzerland)*, 13(3), 1–17. <https://doi.org/10.3390/su13031384>

Stright, A. D., & Supplee, L. H. (2010). Children's self-regulatory behaviors during teacher-directed, seat-work, and small-group instructional contexts. *The Journal of Educational Research*, 95(4), 235–244. <https://doi.org/10.1080/00220670209596596>

Sveinsdóttir, K., Martinsdóttir, E., Thórsdóttir, F., Schelvis, R., Kole, A., & Thórsdóttir, I. (2010). Evaluation of farmed cod products by a trained sensory panel and consumers in different test settings. *Journal of Sensory Studies*, 25(2), 280–293. <https://doi.org/10.1111/j.1745-459X.2009.00257.x>

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74(1), 71–81. <https://doi.org/10.1348/000709904322848824>

Wiederhold, B. K. (2020). Connecting through technology during the coronavirus disease 2019 pandemic: avoiding “zoom fatigue”. *Cyberpsychology, Behavior, and Social Networking*, 23(7), 437–438. <https://doi.org/10.1089/cyber.2020.29188.bkw>

Williams, N. (2021). Working through COVID-19: ‘Zoom’ gloom and ‘Zoom’ fatigue. *Occupational Medicine*, 71(3), 164. <https://doi.org/10.1093/OCMED/KQAB041>

Zhang, M. T., Jo, Y., Lopetcharat, K., & Drake, M. A. (2020). Comparison of a central location test versus a home usage test for consumer perception of ready-to-mix

protein beverages. *Journal of Dairy Science*, 103(4), 3107–3124. <https://doi.org/10.3168/jds.2019-17260>