Journal of Food Science WILEY

On the validity of longitudinal comparisons of central location consumer testing results prior to COVID-19 versus home use testing data during the pandemic

Menghan Shi | **Alina Stelick** [ORCID] | **Susan Licker** | **Robin Dando** [ORCID]

Department of Food Science, Cornell University, Ithaca, New York, USA

Correspondence
Robin Dando, Department of Science, Cornell University, Ithaca, NY 14853, USA.
Email: robin.dando@cornell.edu

Abstract: Consumer testing assays a panel's liking of a food or other sensory stimulus. However, liking can be influenced by mood, with people feeling more uncomfortable, or more unhappy reporting lower liking ratings than those in a higher affect. Though consumer testing typically takes place as a central location test (CLT, usually in a set of standardized sensory booths), the COVID-19 pandemic has resulted in a global pivot to home use tests (HUTs), where panelists can taste and smell samples unmasked more safely while in their own homes. Unfortunately, as this situation differs in many ways to a central location test, this puts the validity of longitudinal comparisons of liking scores under question. Further, as people across the globe report feelings of worry, unease, and stress during the pandemic, this may present a second source of variation in affect with previous years. We tested a set of snack bar samples both at home and in a central location, in repeated measures with the same panel, to test the validity of comparisons across locations. We further compared CLT results to those when testing the same samples in a previous year. Finally, we performed a meta-analysis of existing data on this subject. While liking behavior in CLTs did not differ between years, panelists rated some samples higher when in their own homes, in line with results from the meta-analysis of previous reports. Interestingly, panelists in the study also assigned fewer penalties in the HUT, implying a less analytical mindset when in the home. Results suggest that care should be taken when comparing results taken at home during the COVID-19 pandemic to those taken previously in a central location.

Practical Application: Consumer testing is applied in the food industry to evaluate a panelist's liking for a food product or stimulus. However, liking is also dependent on factors extrinsic to the samples tested. Thus, with the switch to in-home testing due to COVID-19, we compared liking scores from in-home and central locations testing, with higher scores common in HUTs.

KEYWORDS
affect, consumer testing, COVID-19, sensory liking

1 | INTRODUCTION

The COVID-19 pandemic has changed practices throughout the food system. One of these is the broad pivot in sensory testing to a lower exposure model primarily using home use tests (HUTs), instead of the standard central location tests (CLTs), which require panelists to convene in a single place, usually a sensory lab. The context in which a sample is evaluated can be vital to its perceived sensory characteristics, or hedonic properties (reviewed by Delarue & Lageat, 2019). Despite these testing models having fundamental differences, only a small amount of research has gone into assuring that valid comparisons can be made between results from HUTs with CLTs. When evaluating a product at home, much of the control inherent in a sensory booth is absent (ASTM, 1986). Conversely, as people are generally more comfortable in their own home, this may also alter how products are evaluated for their hedonic properties. Historically, not all researchers agree over whether hedonic ratings vary between CLTs and HUTs, put simply, some report they do (Karin et al., 2015; Sveinsdottir et al., 2010), some they do not (Murphy et al., 1958; Pound et al., 2000), and some suggesting "it depends" (Boutrolle et al., 2007; Daillant-Spinnler & Issanchou, 1995). Likewise, CLTs may exhibit better capabilities in differentiating between the specific sensory features of one sample versus another (Sveinsdottir et al., 2010), although again this is not true of all reports (Zhang et al., 2020). While it is certainly not considered best practice to directly compare scores from one consumer test with another, it is also not uncommon within the food industry to examine trends in liking over time, or hold products up against established liking benchmarks, or an arbitrary liking cutoff as an indicator of potential performance in the marketplace.

Additionally, the COVID-19 pandemic has induced a period of greatly increased stress (Qiu et al., 2020), depression, anxiety or other mental health-related issues (Huang & Zhao, 2020; Wang et al, 2020), and altered sleep patterns (Gupta et al., 2020; Wright Jr et al., 2020). As any of these factors can alter sensory responses or hedonic evaluation (Amsterdam et al., 1987; Lv et al., 2018; Nakagawa et al., 1996), we also sought to understand if identical products are viewed less (or indeed more) favorably during the COVID-19 pandemic versus before.

Emotions in their simplest state are often considered in terms of valence (how happy or sad one feels), and of arousal (the degree of excitement or calmness one is experiencing; Watson et al., 1999). When required to rate the liking of a stimulus, feelings of valence or arousal can therefore influence such ratings, for example, coping with stress (Steptoe et al., 1995). If one is in a better mood, or in a more relaxing environment, hedonic ratings may therefore not reflect those when this is not the case. Stress, mood,

and affect are well documented to alter responses to basic taste stimuli (Ileri-Gurel et al., 2013; Nakagawa et al., 1996; Noel & Dando, 2015; Platte et al., 2013) meaning what we sense is different, and not just how we hedonically respond to it, in times of negative affect. Taste-induced hedonic capacity is reduced in a bout of acute depression (Willner & Healy., 1994). There are also many reports of stress-induced preference for more energy dense foods (Oliver et al., 2000), more "snack" type foods over more recognized "meal" foods (Oliver & Wardle, 1999), as well as foods that are generally found more highly palatable (see review by Singh, 2014).

In this study, we sought to compare the results of the same panelists, testing the same products, when in a HUT versus in a CLT, in a consumer test of liking. Our goal was to test whether liking patterns varied when testing in HUTs, as many sensory labs are currently choosing to run their testing. We also sought to gather together previous results comparing CLTs versus HUTs in a meta-analysis, and contrast to our own. As a secondary hypothesis, we compared the results of a previous year's consumer test of the same products, using the same questionnaire, in the same testing location as the CLT session, to provide a comparison with CLT results taken before COVID-19 versus during, to assay whether hedonic responses were altered in general this year.

2 | MATERIALS AND METHODS

2.1 | Panelists

All parts of the study concerning human subject research were reviewed and approved by the Cornell University Institutional Review Board for Human Subject Research, and all work was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Seventy nine panelists between the age of 18 and 68 took part in both tests, with identical questions delivered at home and in a central location, the same panelists participating in both central location and at-home testing, and the order of testing location counterbalanced. Panelists had at least a week between tests, so that samples would not be remembered, and the true purpose of the test could be obscured. Panelists were compensated with $5 for participating in each session. Panelists were recruited from the Cornell University Sensory Evaluation Center through a series of mailing lists, and were self-stated consumers of cereal bars, with no food allergies. All panelists provided informed consent. In both sessions, a questionnaire link was provided on the instruction sheet, where panelists could follow the testing instruction and questions on an iPad during the CLT (Apple Inc., Cupertino, CA, USA) or

on their own device during the HUT session (full questionnaire given in supplementary document 1). Each test lasted ≈20 min, with data collected using RedJade sensory evaluation software (RedJade Consumer & Sensory Software, Martinez, CA, USA). For the test in 2018, 102 panelists participated in testing, recruited in the same manner, and with the same consent procedure and questionnaire.

2.2 | Samples

Three commercially available snack bars (chocolate, cinnamon, and honey flavor) were purchased from local vendors in the Ithaca, NY area, and served with identifiers removed, in small plastic containers marked with three-digit codes, presented in a counterbalanced order during the evaluation. Sample bars were all from large multinational brands, where production methods were presumed to be consistent from year-to-year.

2.3 | Central location test

During the CLT sessions, testing was conducted in the Sensory Evaluation Center at Cornell University. Panelists were provided with unsalted crackers and water for palate cleansing, as well as napkins and a spit cup in case a sample was highly disliked. In the 2020/21 sessions, additional cleaning supplies such as hand sanitizers and paper towels were placed inside the room for panelists' use. Due to COVID-19-related density issues, only one panelist was permitted into the sensory center at a time, thus the test took place over multiple days for the full cohort to test. Panelists were instructed to cleanse their palate with unsalted cracker and rinse their mouth with water between each sample. The product evaluation questionnaire presented to the panelists consisted of 9-point hedonic scales, bipolar JAR scales, and visual analog questions, with product usage and demographic questions given at the end of the session. Additional COVID-19-related questions, which asked about consumers' emotions during sample evaluation and general feelings during COVID-19, were also presented to the panelists at the end of the 2020/21 session. Table 1 outlines the test design, questions asked, and type of scale used for each product.

2.4 | Home use test

During the HUT session, panelists were instructed to pickup a sample bag prepared for the HUT containing identical samples to complete the HUT, however panelists were unaware that samples were the same as the alternate

TABLE 1 Summary of testing questions

Question	Type of scale	Anchors
Overall liking	9-point hedonic	Dislike extremely–like extremely
Appearance liking	9-point hedonic	Dislike extremely–like extremely
Color intensity	JAR bipolar line scale	Light–dark
Gloss/sheen	JAR bipolar line scale	Dull–glossy
Particle uniformity	JAR bipolar line scale	Non-uniform/heterogeneous–uniform/homogenous
Flavor liking	9-point hedonic	Like extremely–dislike extremely
Sweetness intensity	JAR bipolar line scale	Not at all enough–way too much
Chocolate/honey/cinnamon flavor intensity	JAR bipolar line scale	Not at all enough–way too much
Texture liking	9-point hedonic	Like extremely–dislike extremely
Crunchiness intensity	Line scale	Not at all–extremely
Chewiness intensity	Line scale	Not at all–extremely
Softness intensity	Line scale	Not at all–extremely

session, and the true meaning of the test was hidden. Panelists were instructed to follow the same evaluation procedure as they would in the CLT session. Questions were identical to the CLT (demographics and the non-location specific COVID-19 questions were always asked in the first session, either CLT or HUT), and panelists in the CLT or the HUT completed that day's testing in a single session.

2.5 | Meta-analysis

A meta-analysis is a statistical approach that provides integrated estimated effects from a collection of studies with the purpose of evaluating their heterogeneity, and ultimately produce a synthesized finding on a set of studies (Bakbergenuly et al., 2020). Initially, meta-analysis was a popular method in medical research, but has recently demonstrated its utility in food-related fields, such as food safety, food quality, and consumer perception of food products (Esteves & Aníbal, 2021). In a comparative analysis with continuous subject-level data, the outcome of meta-analysis can be measured in either mean difference (MD) or the standardized mean difference (SMD), depending on whether the testing results from all evaluated studies are obtained from the same scale (Bakbergenuly et al., 2020; Higgins et al., 2021). Responses on the 9-point hedonic scale are usually treated as continuous rather than categorical or discrete data in order to utilize more sensitive parametrical statistics than the non-parametric method (Lim, 2011).

An electronic search was conducted on Web of Science and Science Direct databases following the PRISMA principles (Krop et al., 2018; Moher et al., 2015) through the years 2000 to 2021 using relevant search strategies: (1)" HUT" AND "CLT"; (2)" home use test" AND "central location test"; (3) "HUT" AND "CLT" AND "food" AND "consumer testing"; (4) "home use test" AND "central location test" AND "food" AND" consumer testing". After removing duplicates, a total of 68 study results were obtained. The studies were screened through the title, abstract, and full-text based on the exclusion criteria, excluding (1) non-home contextual environment such as virtual reality (VR) environment, other natural consumption conditions (restaurant, bar, airplane, canteen), modified home-use test, evoked context and more; (2) non-blinded label testing design; (3) did not use 9-point or 10-point hedonic scale or lack of other necessary statistical data for meta-analysis. A summary of the searching and exclusion criteria flow chart is demonstrated in Figure 1 below.

The remaining six articles, detailing 28 subgroup studies were evaluated. Since many articles provided more than one tested food, instead of combining these foods (using the article as unit of analysis), subgroups based on tested foods from the articles were separately entered into the meta-analysis (using the subgroup within the article as unit of analysis; Krop et al., 2018). These subgroups contain the same experiment repeated with different test foods. All studies included were collected on 9-point hedonic scaling, except for (Boutrolle et al., 2007), which used a 10-point hedonic scaling. Since Boutrolle et al. (2005), results were also summarized in Boutrolle et al. (2007); results were only counted once (as Boutrolle et al., 2007). The forest plot from the meta-analysis compared changes in overall liking scores of HUT and CLT from various food products.

2.6 | Data analysis

Data from 9-point hedonic scales were analyzed using a linear mixed model, using SPSS (IBM Corp, Armonk, NY), with a statistical cutoff of $p < 0.05$. Forest plot was generated from random effect meta-analysis using R metafor package (Viechtbauer W, 2010), using previously published data sets and data taken in this study from 2020/21. The various (residual) heterogeneity estimators were set to use the restricted maximum-likelihood estimation (method = "REML") default method when estimating τ^2. Penalty analysis was conducted using XLSTAT (Addinsoft, Paris, FR). COVID-19-related questions were analyzed using a series of paired-t tests. Data were analyzed with a linear mixed model built in SPSS using estimate marginal (EM) means syntax modification adjusted with Bonferroni correction. In testing where the same panelists make multiple ratings, treating the participants as a random effect is considered more reflective of reality when accounting for intra-subject effects (Lahne et al., 2014). Panelist was selected as a random effect, with fixed effects for condition (testing location, HUT vs. CLT) and product (three snack bars tested; chocolate, honey, cinnamon). The interaction between testing conditions and sample types was also evaluated in the model (score = product + condition + product*condition).

3 | RESULTS AND DISCUSSION

3.1 | Testing location

The model showed (Table 2) a significant product effect ($p < 0.05$) for overall liking, appearance liking, flavor liking, and texture liking, as would be expected for clearly different samples. While testing location (condition) was not a significant factor alone, the model also showed a significant product*condition interaction effect for overall liking ($p = 0.04$), with the chocolate flavored snack bar more liked when tested in HUT than in a CLT.
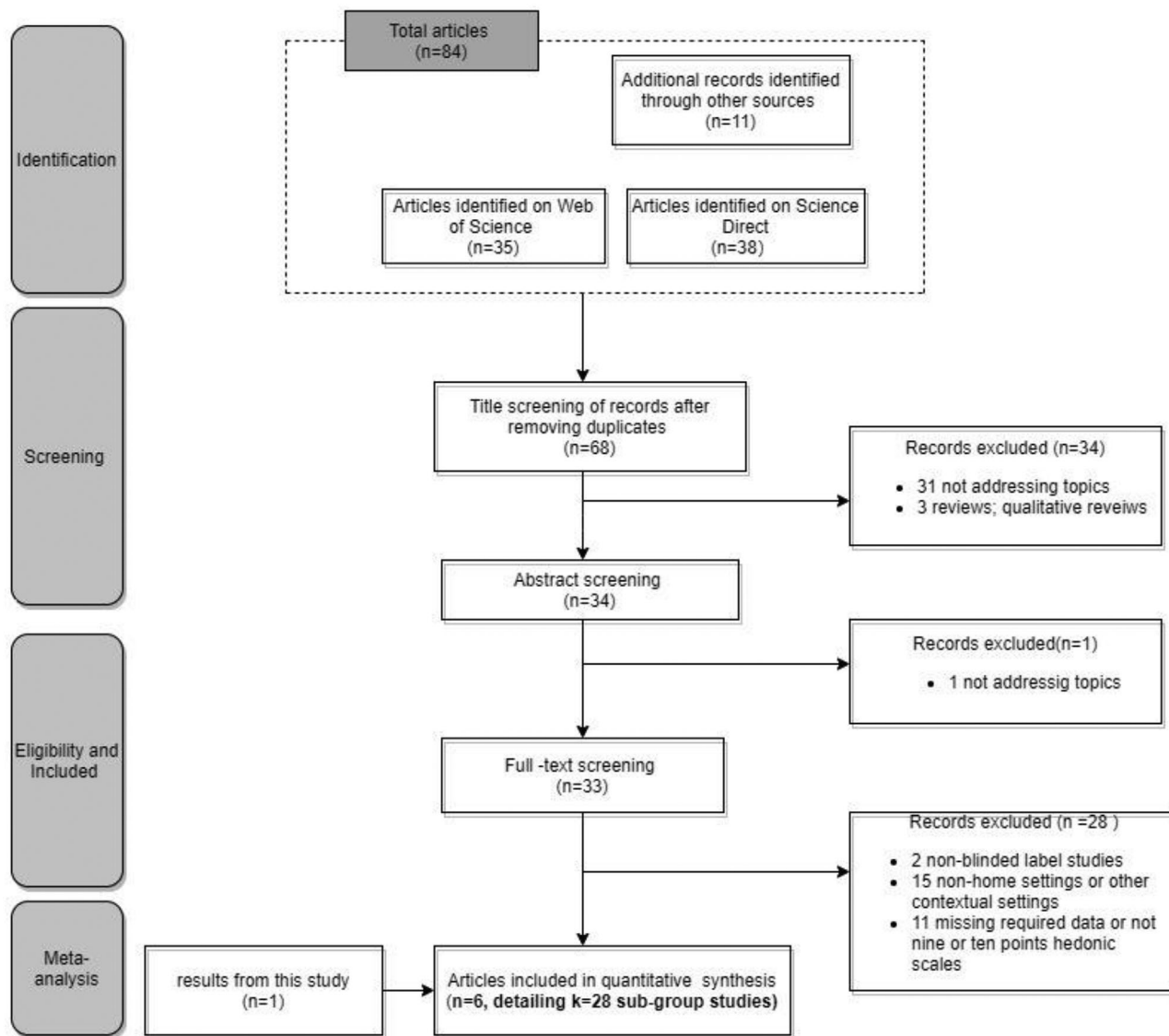
FIGURE 1 PRISMA flow-chart of study selection procedure

TABLE 2 Linear mixed modelparameters, with degrees of freedom (df), F-statistics (F), and significance (Sig)

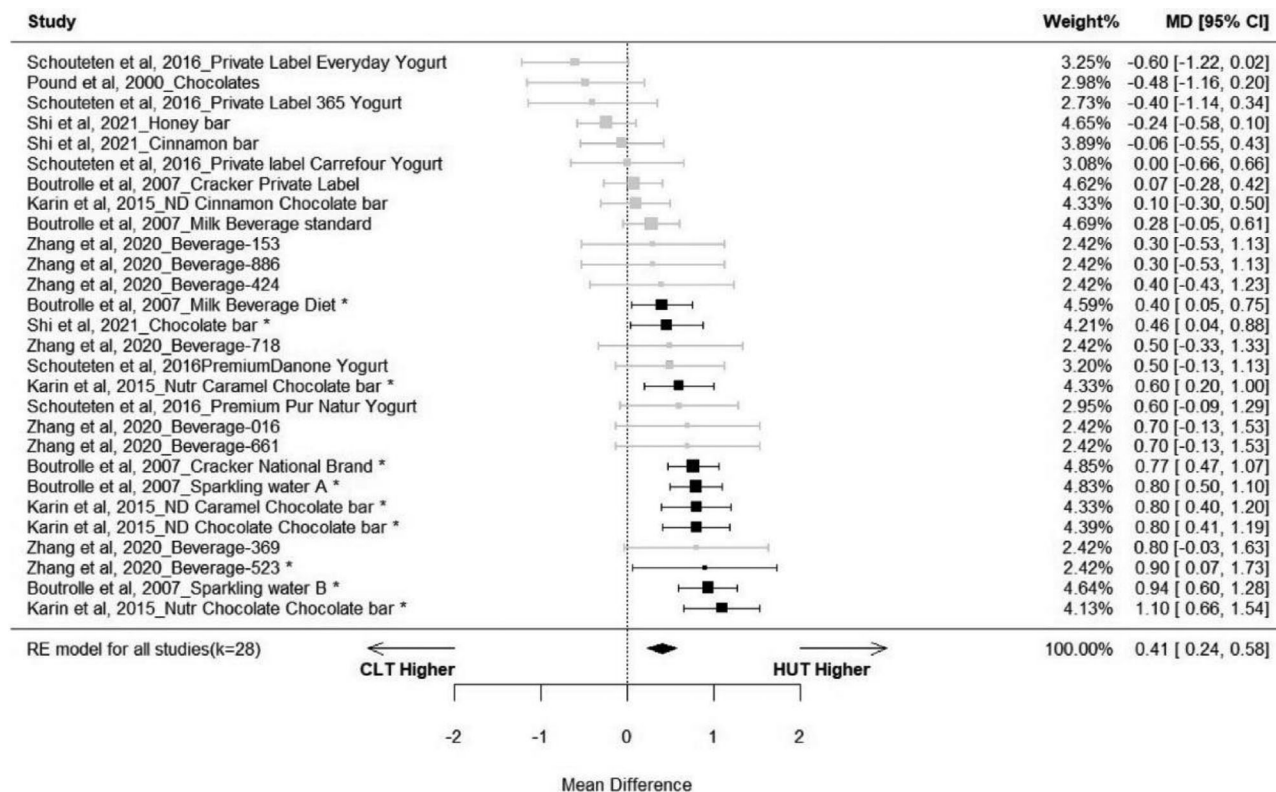| Measure	Source	Numerator df	Denominator df	F	Sig.
	Product	2	390	10.953	0.000
Overall liking	condition	1	390	0.190	0.663
	Product *condition	2	390	3.242	0.040
	Product	2	390	7.565	0.001
Appearance liking	condition	1	390	0.014	0.905
	Product *condition	2	390	1.209	0.300
	Product	2	390	12.727	0.000
Flavor liking	condition	1	390	0.459	0.498
	Product *condition	2	390	1.055	0.349
	Product	2	390	7.204	0.001
Texture liking	condition	1	390	0.008	0.931
	Product *condition	2	390	2.202	0.112
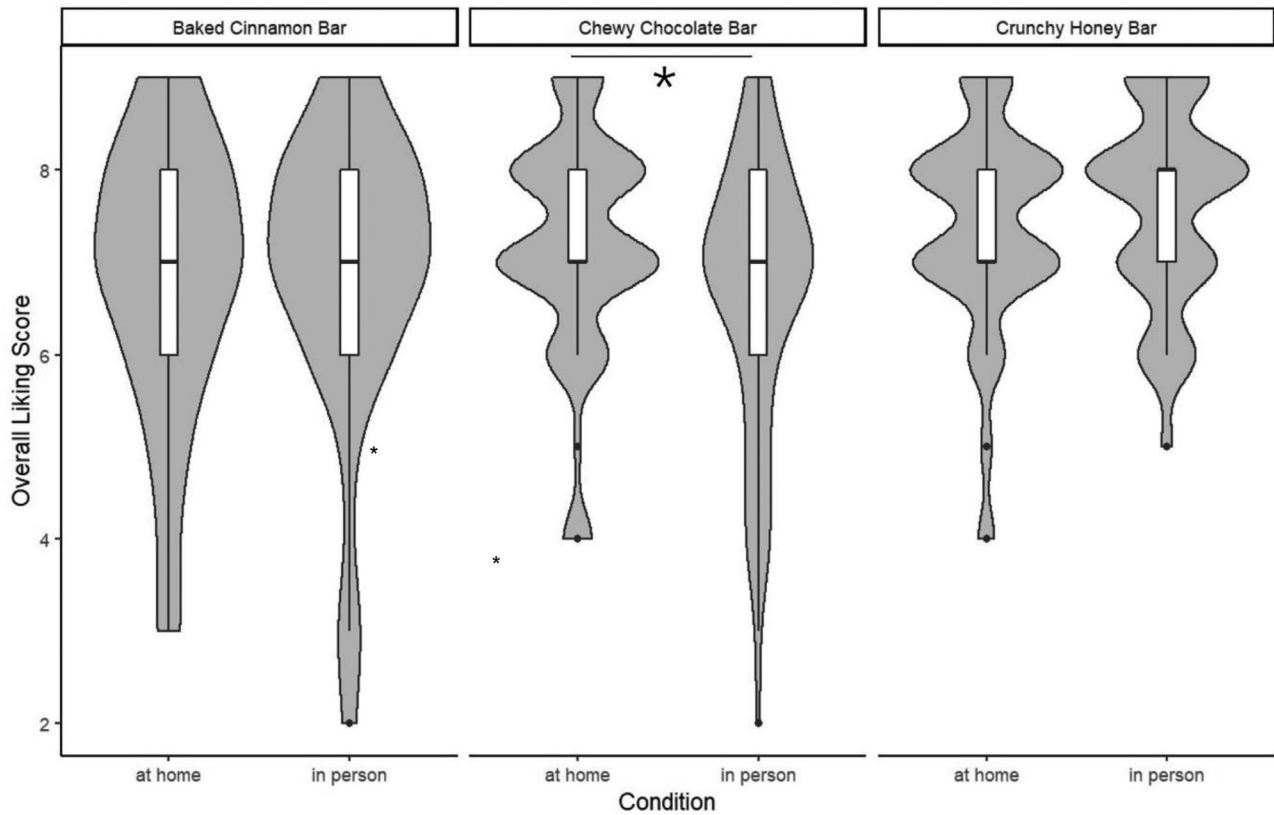
FIGURE 2 Violin plot of overall liking scores at different testing conditions for the three snack bars. *y*-axis denotes liking, 9-point scale. Stars denote statistical significance, *p < 0.05

Table S1 in the Supporting Information summarizes the individual mean scores with pairwise comparisons, along with 95% confidence intervals. The mean overall liking score of the chocolate snack bar was significantly higher in HUT than CLT testing (mean difference = 0.456, p = 0.024). Data were visualized using R ggplot in a violin graph in Figure 2. When included the product*panelist effect as a random effect, the original model improved significantly in its ability to explain the overall liking score with product*condition interaction. The major conclusions remained the same and other hedonic ratings were not affected significantly.

Results are in line with several other reports comparing CLT and HUTs. We sought to quantify the level of agreement between all available data making such a comparison, with a meta-analysis of available data. When examining results of previous reports using 9-point hedonic scaling and 10-point hedonic scaling of liking in HUT and CLT schema, we saw a general shift in liking behavior whereby users in their own home gave ratings with a mean difference of 0.41 points higher on the scale than when in a CLT environment (Figure 3), indicated by a shift to the right in the Forest plot, where ratings in the HUT are higher. SMDs show the same trend, in Figure S1 in the Supporting Information. This is in line with results from

our own study, and with that of multiple previous results with an analogous setup (Boutrolle et al., 2007; Karin et al., 2015; Zhang et al., 2020), although it should be noted, this may not be the case when other factors are introduced (i.e. results from the blind condition in Schouteten et al., 2017 illustrated in Figure 3 differed when panelists were offered an informed condition), and speculatively may have been due to a higher level of comfort and/or well-being when in one's own home versus a somewhat sterile sensory booth. Whether this is the case or not, this is a bias worth noting while home use testing remains prevalent, especially when aiming to make longitudinal comparisons with previous CLT results. Several articles have provided qualitative summaries to compare sensory differences among CLTs and HUTs, also including other contextual environment (Boutrolle & Delarue, 2009; Delarue & Boutrolle, 2010; Jaeger & Porcherot, 2017), but no paper currently has provided a quantitative synthesis to summarize the effect size between HUT and CLT in this field. While both the mean difference (MD, Figure 3) and SMD (Figure S1 in the Supporting Information) were used to compute effect sizes, a similar result was found regardless of reporting method, indicating in general, an increasingly higher overall liking score in the HUT than those of the CLT.

Study	Weight%	MD [95% CI]
Schouteten et al, 2016_Private Label Everyday Yogurt	3.25%	-0.60 [-1.22, 0.02]
Pound et al, 2000_Chocolates	2.98%	-0.48 [-1.16, 0.20]
Schouteten et al, 2016_Private Label 365 Yogurt	2.73%	-0.40 [-1.14, 0.34]
Shi et al, 2021_Honey bar	4.65%	-0.24 [-0.58, 0.10]
Shi et al, 2021_Cinnamon bar	3.89%	-0.06 [-0.55, 0.43]
Schouteten et al, 2016_Private label Carrefour Yogurt	3.08%	0.00 [-0.66, 0.66]
Boutrolle et al, 2007_Cracker Private Label	4.62%	0.07 [-0.28, 0.42]
Karin et al, 2015_ND Cinnamon Chocolate bar	4.33%	0.10 [-0.30, 0.50]
Boutrolle et al, 2007_Milk Beverage standard	4.69%	0.28 [-0.05, 0.61]
Zhang et al, 2020_Beverage-153	2.42%	0.30 [-0.53, 1.13]
Zhang et al, 2020_Beverage-886	2.42%	0.30 [-0.53, 1.13]
Zhang et al, 2020_Beverage-424	2.42%	0.40 [-0.43, 1.23]
Boutrolle et al, 2007_Milk Beverage Diet *	4.59%	0.40 [0.05, 0.75]
Shi et al, 2021_Chocolate bar *	4.21%	0.46 [0.04, 0.88]
Zhang et al, 2020_Beverage-718	2.42%	0.50 [-0.33, 1.33]
Schouteten et al, 2016PremiumDanone Yogurt	3.20%	0.50 [-0.13, 1.13]
Karin et al, 2015_Nutr Caramel Chocolate bar *	4.33%	0.60 [0.20, 1.00]
Schouteten et al, 2016_Premium Pur Natur Yogurt	2.95%	0.60 [-0.09, 1.29]
Zhang et al, 2020_Beverage-016	2.42%	0.70 [-0.13, 1.53]
Zhang et al, 2020_Beverage-661	2.42%	0.70 [-0.13, 1.53]
Boutrolle et al, 2007_Cracker National Brand *	4.85%	0.77 [0.47, 1.07]
Boutrolle et al, 2007_Sparkling water A *	4.83%	0.80 [0.50, 1.10]
Karin et al, 2015_ND Caramel Chocolate bar *	4.33%	0.80 [0.40, 1.20]
Karin et al, 2015_ND Chocolate Chocolate bar *	4.39%	0.80 [0.41, 1.19]
Zhang et al, 2020_Beverage-369	2.42%	0.80 [-0.03, 1.63]
Zhang et al, 2020_Beverage-523 *	2.42%	0.90 [0.07, 1.73]
Boutrolle et al, 2007_Sparkling water B *	4.64%	0.94 [0.60, 1.28]
Karin et al, 2015_Nutr Chocolate Chocolate bar *	4.13%	1.10 [0.66, 1.54]
RE model for all studies(k=28)	100.00%	0.41 [0.24, 0.58]

CLT Higher   HUT Higher

Mean Difference

FIGURE 3 Forest plot displaying mean differences (MDs) for comparisons of HUT versus CLT overall liking scores from various food products from each study ($k = 28$). As most studies provided more than one unique comparison of products, each product subgroup was separately entered into the meta-analysis from each study. All studies were collected on 9-point hedonic scaling, except for Boutrolle et al. (2007) on 10-point hedonic scaling. Bars in black represent liking scores significantly different between HUT and CLT setups, with grey not significantly different, boxes to the right indicating increasingly higher scores in the HUT, with those to the left higher in the CLT. Diamond represents the estimated pooled mean difference, indicating HUT score was generally 0.41 MD higher than that of CLT; box sizes reflect the relative weight apportioned to studies in the meta-analysis

3.2 | JAR scale measures

In line with common consumer testing practices, after liking scores of samples were measured, a series of questions on flavor, texture, and appearance were also asked of the panelists, using JAR scales. These measures were used in calculating penalties given to the samples, with penalty analysis in Table S2 in the Supporting Information and Figure 4 (for more information, see Rothman & Parker, 2009). Interestingly, it became clear that there were far more penalties in general applied to the same samples when the panelists were in the CLT than when testing in their own homes (Figure 4).

This would seem to suggest that panelists in HUTs are in less of an analytical mindframe, possibly implying a more relaxed attitude to testing, despite all wording in the tests being consistent from one test to the next. Sosa et al. (2008) previously suggested that such an effect may be due to panelists being in a less critical mindset when at home. This also aligns with other previous work, such as that suggesting that the attributes of oven-baked cod loin

samples were more ably appreciated when in a CLT situation than HUTs (Sveinsdottir et al., 2010), samples of soup were better discriminated in paired preference tests (Miller et al., 1955), and that consumers testing chocolate bars at home gave much fewer comments on the samples than when in sensory booths (Karin et al., 2015). An alternative hypothesis for why this may occur lies in the nature of data input in the booth versus at home. Most sensory booths use large computer monitors and both a keyboard and mouse, which when compared to a cellphone screen, using touch screen input would require less effort for a long answer, and would give more accuracy in inputting precise scaling data. However, De Bruijne and Wijnant (2013) suggested that data input fidelity remained high across input device, with Visalli et al. (2016) actually suggesting superior results for a task with higher cognitive load, such as a temporally varying rating. We would conclude from this result that central location testing would appear to be more sensitive in diagnosing specifics of a product's attribute liking than in home testing, and thus again care must be taken regarding such ratings taken in a panelist's home, as
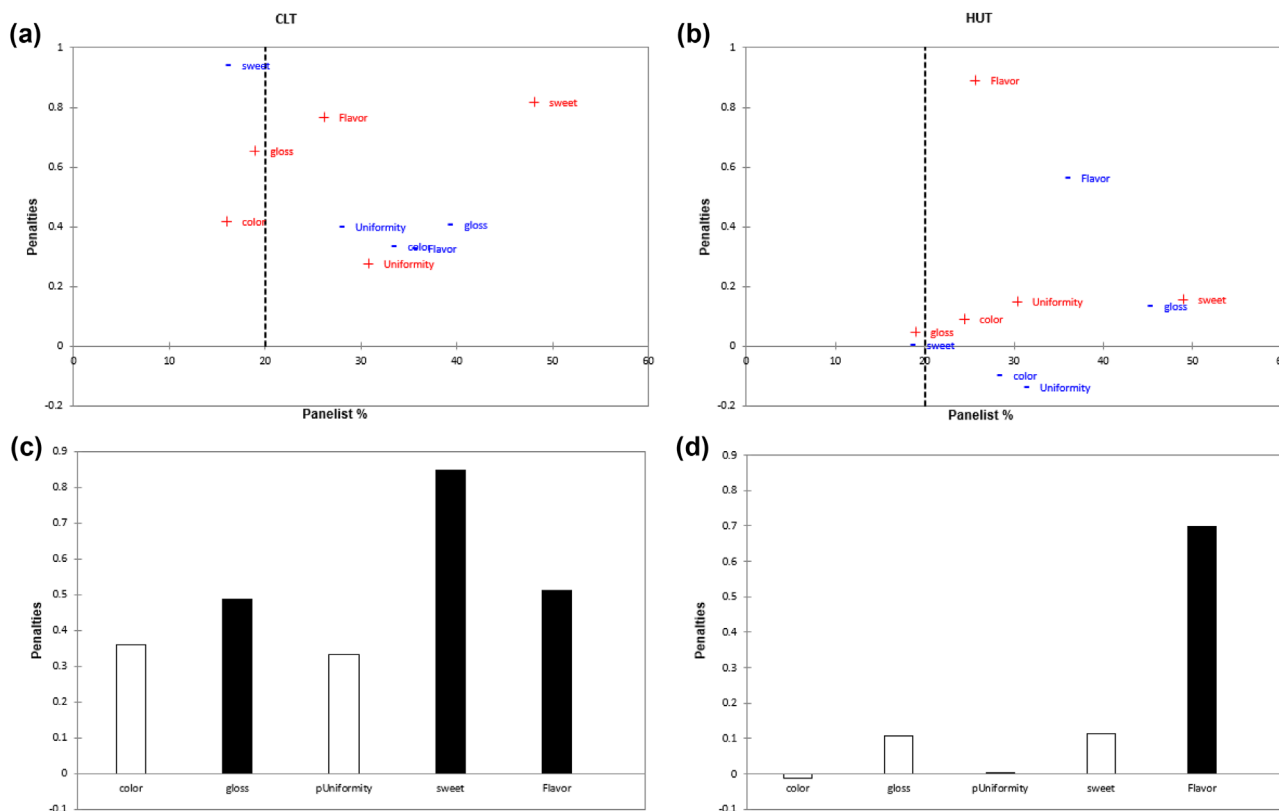
FIGURE 4 Penalty analysis of samples in CLT (left) versus HUT (right). Penalties (mean drops, *y*-axis) assigned to samples from (a) CLT and (b) HUT, versus percent of panel (*x*-axis) assigning the penalty. Dashed line indicates where 20% of panel began to cite issues. Red indicates attribute too high, blue too low. Total penalties across samples assigned in (c) CLT and (d) HUT. Black bars indicate the attributes' penalty scores (*y*-axis) were significantly different from 0 ($p < 0.05$). Only attributes that were cited by more than 20% of panelists were tested

TABLE 3 Overall liking mean difference during COVID-19 and prior to COVID-19

Sample name	Mean difference	df	Sig	95% confidence interval for difference	
				Lower bound	Upper bound
Honey bar	0.206	516.089	0.342	−0.219	0.630
Cinnamon bar	0.102	516.089	0.636	−0.322	0.527
Chocolate bar	−0.409	516.089	0.059	−0.833	0.016

sensory testing is predominantly occurring during COVID. In fact, the notion of which mode better models a true measure of liking is an interesting one, with panelists at home, and in this less analytical mindframe, possibly more reflective of the eating experience than when in the sensory lab.

3.3 | Year-to-year analysis

A further hypothesis we tested was that panelists may be experiencing a general malaise due to worry or stress over COVID-19, affecting hedonic measures in consumer tests. The samples tested in the testing were selected to match those from a test in 2018, before the COVID-19 pandemic. No statistical differences (Table 3) were observed between

liking scores for the same products from the CLT in 2018 versus testing above, suggesting that in fact that liking scores remained consistent from year-to-year.

3.4 | COVID-19-related stress

Finally, a series of questions were asked of panelists on how concerned they were about a number of factors related to COVID-19, while both in the CLT and HUT. Two-tailed paired *t*-tests evaluated differences in a panelist's attitudes to testing in either environment (doubt, nervous, apprehensive, anxious, worried, comfort) during CLT and HUT. No statistical differences were observed between the two testing conditions in any measure (all $p > 0.05$; Table S3 in

the Supporting Information), indicating that panelists are equally comfortable in the booths versus at home.

3.5 | Limitations

While the project garnered interesting results, there are some limitations that should be discussed. First, while we did attempt to compare results from identical samples taken in the same lab setting in a previous year, this would not entirely decouple the effects from home testing alone versus the effects of testing while the COVID-19 pandemic was still occurring. Additionally, not all samples saw changes from one testing location to the next. This is in line with previously reported results, and may or may not relate to the specific qualities of the samples. We acknowledge that best practices already state that the comparison of liking scores directly with those from another test is generally not valid, due to results only being fully valid against scores obtained within the same test. Finally, we tested a single food category, and such effects may not be universal. This was the original motivation for performing the meta-analysis, where effects seemed at a reasonable level of agreement across categories.

4 | CONCLUSION

Our results imply that panelists report their liking for products to be slightly greater when testing from home, suggesting care should be taken in comparing such HUT results with those from previous CLTs. Further, panelists appeared more critical when in the sensory booth than at home. No difference was seen between CLT scores during COVID-19 versus before. Finally, while panelists reported being slightly less comfortable testing in the booth versus at home, overall they were still comfortable with their role as panelists, assuming care was taken to assure their safety.

AUTHOR CONTRIBUTIONS
Menghan Shi: formal analysis; investigation; visualization. **Alina Stelick:** formal analysis. **Susan Licker:** formal analysis; investigation. **Robin Dando:** conceptualization; project administration; writing original draft.

ORCID
Alina Stelick https://orcid.org/0000-0001-5490-9349
Robin Dando https://orcid.org/0000-0001-8539-3971

REFERENCES
Amsterdam, J. D., Settle, R. G., Doty, R. L., Abelman, E., & Winokur, A. (1987). Taste and smell perception in depression. *Biological Psychiatry*, *22*(12), 1481–1485.

ASTM Special Technical Publication 913 (1986). J. Eggert & K. Zook (Eds.) ASTM Publication Code Number (PCN) 04-913001-36.

Bakbergenuly, I., Hoaglin, D. C., & Kulinskaya, E. (2020). Estimation in meta-analyses of mean difference and standardized mean difference. *Statistics in Medicine*, *39*(2), 171–191. https://doi.org/10.1002/sim.8422

Boutrolle, I., & Delarue, J. (2009). 8—Studying meals in the home and in the laboratory. In H. L. Meiselman (Ed.), *Meals in science and practice* (pp. 128–165). Woodhead Publishing. https://doi.org/10.1533/9781845695712.3.128

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*(3), 490–499.

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*(8), 704–713.

Daillant-Spinnler, B., & Issanchou, S. (1995). Influence of label and location of testing on acceptability of cream cheese varying in fat content. *Appetite*, *24*(2), 101–105.

De Bruijne, M., & Wijnant, A. (2013). Comparing survey results obtained via mobile devices and computers: An experiment with a mobile web survey on a heterogeneous group of mobile devices versus a computer-assisted web survey. *Social Science Computer Review*, *31*(4), 482–504.

Delarue, J., & Boutrolle, I. (2010). 8—The effects of context on liking: Implications for hedonic measurements in new product development. In S. R. Jaeger & H. MacFie (Eds.), *Consumer-driven innovation in food and personal care products* (pp. 175–218). Woodhead Publishing. https://doi.org/10.1533/9781845699970.2.175

Delarue, J., & Lageat, T. (2019). 22.1 Testing contexts and product experience. *Context: The Effects of Environment on Product Design and Evaluation. Woodhead Publishing, Sawston, UK*.

Esteves, E., & Aníbal, J. (2021). Sensory evaluation of seafood freshness using the quality index method: A meta-analysis. *International Journal of Food Microbiology*, *337*, 108934. https://doi.org/10.1016/j.ijfoodmicro.2020.108934

Gupta, R., Grover, S., Basu, A., Krishnan, V., Tripathi, A., Subramanyam, A., Nischal, A., Hussain, A., Mehra, A., Ambekar, A., Saha, G., Mishra, K. K., Bathla, M., Jagiwala, M., Manjunatha, N., Nebhinani, N., Gaur, N., Kumar, N., Dalal, P. K., … Avasthi, A. (2020). Changes in sleep pattern and sleep quality during COVID-19 lockdown. *Indian Journal of Psychiatry*, *62*(4), 370.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2021). *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. www.training.cochrane.org/handbook

Huang, Y., & Zhao, N. (2020). Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: A web-based cross-sectional survey. *Psychiatry Research*, *288*, 112954.

Ileri-Gurel, E., Pehlivanoglu, B., & Dogan, M. (2013). Effect of acute stress on taste perception: in relation with baseline anxiety level and body weight. *Chemical Senses*, *38*(1), 27–34.

Jaeger, S. R., & Porcherot, C. (2017). Consumption context in consumer research: Methodological perspectives. *Current Opinion in Food Science*, *15*, 30–37. https://doi.org/10.1016/j.cofs.2017.05.001

Karin, W., Annika, Å., & Anna, S. (2015). Exploring differences between central located test and home use test in a living lab context. *International Journal of Consumer Studies*, *39*(3), 230–238.

Krop, E. M., Hetherington, M. M., Nekitsing, C., Miquel, S., Postel-nicu, L., & Sarkar, A. (2018). Influence of oral processing on appetite and food intake – A systematic review and meta-analysis. *Appetite*, *125*, 253–269. https://doi.org/10.1016/j.appet.2018.01.018

Lahne, J., Trubek, A. B., & Pelchat, M. L. (2014). Consumer sensory perception of cheese depends on context: A study using comment analysis and linear mixed models. *Food Quality and Preference*, *32*, 184–197.

Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference*, *22*(8), 733–744. https://doi.org/10.1016/j.foodqual.2011.05.008

Lv, W., Finlayson, G., & Dando, R. (2018). Sleep, food cravings and taste. *Appetite*, *125*, 210–216.

Miller, P. G., Nair, J. H., & Harriman, A. J. (1955). A household and a laboratory type of panel for testing consumer preference. *Food Technology*, *9*(9), 445–449.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petti-crew, M., Shekelle, P., & Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. https://doi.org/10.1186/2046-4053-4-1

Murphy, E. F., Clark, B. S., & Berglund, R. M. (1958). A consumer survey versus panel testing for acceptance evaluation of Maine sardines. *Food Technology*, *12*(5), 222–226.

Nakagawa, M., Mizuma, K., & Inui, T. (1996). Changes in taste perception following mental or physical stress. *Chemical Senses*, *21*(2), 195–200.

Noel, C., & Dando, R. (2015). The effect of emotional state on taste perception. *Appetite*, *95*, 89–95.

Oliver, G., & Wardle, J. (1999). Perceived effects of stress on food choice. *Physiology & Behavior*, *66*(3), 511–515.

Oliver, G., Wardle, J., & Gibson, E. L. (2000). Stress and food choice: A laboratory study. *Psychosomatic Medicine*, *62*(6), 853–865.

Platte, P., Herbert, C., Pauli, P., & Breslin, P. A. (2013). Oral perceptions of fat and taste stimuli are modulated by affect and mood induction. *Plos One*, *8*(6), e65006.

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.

Qiu, J., Shen, B., Zhao, M., Wang, Z., Xie, B., & Xu, Y. (2020). A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: Implications and policy recommendations. *General Psychiatry*, *33*(2), e100213.

Rothman, L., & Parker, M. (2009). *Just-About-Right JAR scales: Design, usage, benefits and risks*. Manual MNL-63-EB. ASTM International.

Schouteten, J. J., De Steur, H., Sas, B., De Bourdeaudhuij, I., & Gellynck, X. (2017). The effect of the research setting on the emotional and sensory profiling under blind, expected, and informed conditions: A study on premium and private label yogurt products. *Journal of Dairy Science*, *100*(1), 169–186.

Singh, M. (2014). Mood, food, and obesity. *Frontiers in Psychology*, *5*, 925.

Sosa, M., Martinez, C., Marquez, F., & Hough, G. (2008). Location and scale influence on sensory acceptability measurements among low-income consumers. *Journal of Sensory Studies*, *23*(5), 707–719.

Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, *25*(3), 267–284.

Sveinsdottir, K., Martinsdottir, E., Thorsdottir, F., Schelvis, R., Kole, A., & Thorsdottir, 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.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Visalli, M., Lange, C., Mallet, L., Cordelle, S., & Schlich, P. (2016). Should I use touchscreen tablets rather than computers and mice in TDS trials? *Food Quality and Preference*, *52*, 11–16.

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *International Journal of Environmental Research and Public Health*, *17*(5), 1729.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*(5), 820.

Willner, P., & Healy, S. (1994). Decreased hedonic responsiveness during a brief depressive mood swing. *Journal of Affective Disorders*, *32*(1), 13–20.

Wright, K. P. Jr, Linton, S. K., Withrow, D., Casiraghi, L., Lanza, S. M., de la Iglesia, H., Vetter C., & Depner, C. M. (2020). Sleep in university students prior to and during COVID-19 stay-at-home orders. *Current Biology*, *30*(14), R797–R798.

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.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.