



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

Remote testing: Sensory test during Covid-19 pandemic and beyond

Caterina Dinnella^{a,*}, Lapo Pierguidi^a, Sara Spinelli^a, Monica Borgogno^b, Tullia Gallina Toschi^c, Stefano Predieri^d, Giliana Lavezzi^e, Francesca Trapani^f, Matilde Tura^c, Massimiliano Magli^d, Alessandra Bendini^c, Erminio Monteleone^a

^a Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Italy

^b Mérieux NutriSciences, Italy

^c Department of Agricultural and Food Science, Alma Mater Studiorum-University of Bologna, Italy

^d Institute of BioEconomy, National Research Council, Italy

^e Kerry Ingredients & Flavours, EMEA Region, Italy

^f Enrico Giotti S.p.A. a Subsidiary of McCormick & Company, Inc, Italy

ARTICLE INFO

Keywords:

Sensory analysis
Discrimination tests
Descriptive analysis
TDS
CATA
Liking
Home test

ABSTRACT

Restrictions adopted by many countries in 2020 due to Covid-19 pandemic had severe consequences on the management of sensory and consumer testing that strengthened the tendency to move data collection out of the laboratory. Remote sensory testing, organized at the assessor's home or workplace and carried out under the live online supervision of the panel leader, represents a trade-off between adequate control and the convenience of conducting testing out of the lab. The Italian Sensory Science Society developed the "Remote sensory testing" research project aimed at testing the effectiveness and validity of the sensory tests conducted remotely through a comparison with evaluations in a classical laboratory setting. Guidelines were developed to assist panel leaders in setting up and controlling the evaluation sessions in remote testing conditions. Different methods were considered: triangle and tetrad tests, Descriptive Analysis and Temporal Dominance of Sensations tests, all of which involved trained panels, and Check-All-That-Apply and hedonic tests with consumers. Remote sensory testing provided similar results to the lab testing in all the cases, with the exception of the tetrad test run at work. Findings suggest that remote sensory testing, if conducted in strict compliance with specifically developed sensory protocols, is a promising alternative to laboratory tests that can be applied with both trained assessors and consumers even beyond the global pandemic.

1. Introduction

Covid-19 pandemic has impacted our lives tremendously from several points of view. This includes food behaviours and preferences (Marty et al., 2021; Li et al., 2021), but also how sensory and consumer testing are conducted due to the restrictions to limit the spread of the virus. In fact, due to government health and safety directives, having assessors gathering in a facility may not be possible, or may be perceived as unsafe by the assessors. This was very clear in the period of strict lockdowns adopted by many countries in 2020 but has consequences that also impact the management of sensory testing even when less restrictive governmental measures are adopted.

The push to move out of the laboratory and to conduct more and more home use testing in sensory studies has been observed far before the pandemic (Galmarini et al., 2016; Martin et al., 2014; Nogueira-

Terrones et al., 2006) and predicted as a growing trend in the future (Meiselman, 2013). Furthermore, data collection out of the lab is being made easier by internet technology, which permits data collection anywhere, anytime and in real time. Recently the possibility of conducting sensory tests in the assessor's own vehicle has been proposed as a valid alternative to lab sensory booth setting allowing for participants to feel safe from the risk of COVID-19 while performing sensory evaluations (Seo et al., 2021). Live video calls were used for conducting observational studies on children's food preference and intake and were proposed as a valid instrument for food behavioural studies at participant's home even beyond pandemic (Venkatesh and DeJesus, 2021)

Internet technology, together with the constraints of the global pandemic, allowed the setup of a solution that represents a trade-off between adequate control and the convenience of conducting testing at home: remote sensory testing. Remote sensory testing is sensory

* Corresponding author.

E-mail address: caterina.dinnella@unifi.it (C. Dinnella).

<https://doi.org/10.1016/j.foodqual.2021.104437>

Received 1 August 2021; Received in revised form 28 September 2021; Accepted 10 October 2021

Available online 15 October 2021

0950-3293/© 2021 Elsevier Ltd. All rights reserved.

testing organised out of the lab, for example at assessor's home or workplace. However, remote sensory testing radically differs from home use tests as it includes a constant control of the conditions in which the test is performed. The panel leader is in fact connected in videoconferencing with the assessors for all the duration of the test, having always the possibility to monitor the evaluations and to interact with the assessors, similarly to what happens in a lab environment. These specificities make remote sensory testing a very useful tool, particularly for sensory tests that require highly controlled conditions, such as with trained assessors. While remote sensory testing has been more and more applied starting from the spring/summer 2020, in particular by sensory test providers, food and personal/home care industries, and some attempts of operative guidelines were made (White Paper, [Compusense, 2020](#)), at present there is no published scientific literature that supports the validity of this methodology.

The Italian Sensory Science Society (SISS) has responded to this urgent need, that was very large not only for research centers and universities but particularly for industries, by developing the "Remote sensory testing" research project aimed at testing the effectiveness and validity of the sensory tests conducted remotely through a comparison with evaluations in a classical laboratory setting. Five different discriminant or descriptive methods were considered: triangle and tetrad tests ([Lawless & Heymann, 2010](#)); Descriptive Analysis ([Lawless & Heymann, 2010](#)); Temporal Dominance of Sensations ([Pineau et al., 2003](#)); Check-All-That-Apply ([Adams et al., 2007](#); [Meyners and Castura, 2014](#)). All the tests involved trained panels apart from Check-All-That-Apply that involved consumers with no specific preliminary sensory training. An hedonic test with consumers was also considered. The aim of this paper is to present the results of this study and to illustrate the guidelines developed for remote sensory testing.

General Methods

1.1. Overview of the experimental plan

The research project was run in 2020 and involved, on a voluntary basis, six sensory laboratories of public and private organizations belonging to SISS across Italy. A working group, open to SISS members from the laboratories who joined the project, selected the sensory methods to be included in the project to cover quality control activities, descriptive methods performed by a trained panel, and a consumer

study. Internal (assessors recruited from personnel of the organization running the test) and external panels (assessors recruited out of the organization running the test) took part in the study. Internal panels were selected as an example of procedures generally adopted by food and food ingredient companies for quality control purposes. To be consistent with the usual conditions of internal panel activities, at the working place and during working hours, the assessor's workstation was selected as remote testing location (RT-W). Assessors' home was selected as the remote testing location (RT-H) for external panels.

The working group revised the procedure applied for data collection in lab conditions and defined the procedure for remote testing conditions. Each laboratory team performed data analysis relevant to the sensory test conducted under its own responsibility. The experimental plan and the testing dates are summarized in [Table 1](#). Sensory tests in lab settings during 2020 were performed according to Italian government regulations to control for virus spread, that include: controlled access to the lab after testing for the absence of COVID-19 symptoms; compliance with the minimum interpersonal distance of 1.8 m; wearing masks apart while testing the sample; environment and individual workstation sanitization after every use.

1.2. Procedure for remote testing

Evaluations in remote condition (RT) were performed by video call from the assessor's home (H) or at work (W, i.e. at the company of the participating research partners of SISS, where sensory testing is usually performed by the trained panellists, but in their own office instead that in the sensory lab) under the guidance of the panel leader. An "evaluation box" with all the equipment needed was delivered to the assessor's home for RT-H evaluations, while a tray with the samples under evaluation, water and crackers for rinsing procedure was brought to the assessor desk in his/her own office by sensory lab personnel for RT-W. Video calls were operated using multimedia platforms (Google Meet and Microsoft Teams). Data were collected with the web versions of softwares on the market for sensory data acquisition with the only exception of hemp seed oils where a paper evaluation sheet was used. Data acquisition software used for each test are described in further detail in the upcoming sections.

Table 1

Summary of sensory evaluations performed in lab (LAB) and remote testing (RT) conditions: type of test, samples (product and sample number), location and dates (period) for lab and remote testing conditions (home -H, at work -W), number of assessors and experimental design.

Test	Samples	LAB			RT			Experimental Design
		Location	Assessor (n)	Period	Location	Assessor (n)	Period	
Study 1 Discrimination tests								
Tetrad test	Lime Flavour (2 samples)	Kerry (Mozzo, Bergamo Italy)	36	Jul 2019	W	36	Oct 2020	within-subjects
Triangle test	Orange Flavour (2 samples)	Giotti McCormick (Scandicci, Firenze, Italy)	36	Feb 2020	W	36	Jun 2020	between-subjects
Study 2 Descriptive Analysis								
Re-training	Coffee descriptor standards	Mérieux-Nutriscience (Prato, Italy)	23	Jan 2020	H	23	Apr 2020	within-subjects and between-subjects
Evaluations	Mocha Coffee (3 samples)	Mérieux-Nutriscience (Prato, Italy)	15	Jan 2020	H	24	May 2020	within-subjects and between-subjects
	Hemp seed oils (4 samples)	Dept. of Agricultural and Food Sciences, University of Bologna (Bologna, Italy)	9	May-Jun 2020	H	9	Jun-Jul 2020	within-subjects
Study 3 Temporal Dominance of Sensations								
	Chewing gums (3 samples)	CNR-Istitute for Bioeconomy (Bologna, Italy)	9	3rd week of June	H	9	4th week of June	within-subjects
Study 4 Check All That Apply								
	Gluten free breads (4 samples)	SensoryLab, Dept. of Agriculture, Food, Environment and Forestry, University of Florence (Firenze, Italy)	60	Nov 2018	H	60	Apr 2020	between-subject

1.3. General guidelines for panel leaders

Panel leaders were provided with general guidelines for setting up and controlling the evaluation sessions. A list of preliminary documents to lead the remote testing was defined, which included the list of the participants in the session with their contact (email and mobile number, to contact the assessors in private during evaluations in case of need) and the MasterCard to assist assessors in the correct sample evaluation order. Guidelines specified that panel leaders were responsible for supervising the preparation of the “evaluation box” (RT-H) or the “tray” (RT-W) to be delivered to assessors. They were recommended to include in the box all the equipment needed for carrying out the evaluation in order to standardize as much as possible the evaluation conditions. Thus, it was suggested to include in the box three-digit coded vessels for sample evaluation, small white paper towels to cover the working surface, napkins, cutlery if necessary, and unsalted crackers for the mouth rinsing procedure. Moreover, panel leaders were asked to select the most appropriate packaging to assure sample stability and avoid leakage during transportations. It was recommended to minimize the time between sample preparation and evaluation and to eventually perform preliminary tests with sensory lab personnel to identify the maximum time allowed between preparation and evaluation to avoid sample perceptual changes. Panel leaders were also requested to make available to assessors the instructions for the correct sample storage and handling before evaluation. Panel leaders were recommended to schedule remote testing sessions and “evaluation box” delivery controlling for time between sample shipping and evaluation. They were invited to run the sessions with a maximum of six participants per time to allow the easy monitoring of the assessor’s behaviour during evaluation and facilitate eventual corrective action toward a single assessor. They were also recommended to define the evaluation’s duration time considering the time needed for checking for the workstation setting up and managing possible delays due to assessor connection difficulties. Guidelines also included detailed instructions on how to plan and carry out the video call on popular multimedia platforms. Finally, panel leaders were recommended to identify (or act themselves as) a contact person for assessor’s assistance requests before the evaluation session.

Health status of personnel in charge for sample preparation both for RT-W and RT-H was daily controlled for the absence of COVID-19 symptoms according to the measures adopted from the Italian government to limit the spread of the virus (*Gazzetta Ufficiale della Repubblica Italiana*, n. 61, anno 161°). Procedures reported in “Interim provisions on food hygiene during the SARS-CoV-2 epidemic” from ISS COVID-19 Working group on Veterinary public health and food safety (ii, 17p. *Rapporto ISS COVID-19n. 17/2020*) were followed for sample handling.

1.4. General instructions to assessors

Assessors received by e-mail the general instructions to participate in the sensory evaluations. They were requested to have a stable internet connection and an appropriate device allowing for both audio and video connection (tablet or pc) to participate in the experiment. They were recommended to have their mobile phone available during the evaluation. Assessors were informed that they would have received instructions for video-call connection and for setting up their workstation at home in the days immediately preceding the test. They were also informed that an “evaluation box” would have been delivered at home with all the equipment needed for setting up the workstation and with instructions for sample storage (RT-H) or, alternatively, that a tray would have been delivered to them by sensory lab personnel (RT-W). A text message advised the assessors that the evaluation box had been shipped, and they were asked to let the sensory laboratory contact person know when they had received it. They were instructed to set-up the workstation in a quiet room, possibly with a window in order to assure ventilation, where they could be alone during the whole test; they were also instructed to choose a working surface wide enough (i.e.

90x60 cm) to comfortably position the connection device and all the equipment needed for the evaluations. Furthermore, they were asked to avoid cooking and using household cleaners one hour before the test and to follow the general behaviour rules preceding sensory evaluations (e.g., do not smoke, eat or drink, apart from water, prior to the test). Assessors were asked to sign the informed consent attached to the message and send it back to the sensory lab. A contact was provided for further clarifications, if needed.

1.5. General procedure for remote testing

The panel leader opened the video call, checked that the name of the participants in the video call was included in the assessor list and individually tested audio and video connection ensuring that the assessor’s faces were visible in the frame. Then, assessors declared that they signed the informed consent and confirmed they had no allergies or intolerances to the sample ingredients. The panel leader invited assessors to position all the equipment delivered in the “evaluation box” on the working surface (e.g., a table or a desk), to access the software for data acquisition and to position the samples according to the evaluation order (the first sample on the left). The panel leader individually checked with assessors the sample positioning according to the MasterCard. Then, evaluation aim and modality were recalled, assessors were told that they were not allowed to talk while the evaluation was ongoing and were instructed to use their mobile to contact the panel leader for assistance (with microphones of the video call muted to avoid interferences with the other panelists’ evaluations). Assessors were invited to open the data acquisition software page in full screen mode (so that they could not see each other anymore while their faces were all visible to the panel leader), mute their microphone and start the evaluation. Once assessors completed the evaluation, they were requested to share their screen with the panel leader to show the final page of the session and were allowed to quit the video call. The panel leader monitored the assessor’s behaviour during the evaluation and in case of incorrect actions privately contacted the assessor and drew his/her attention on the strict compliance to the evaluation procedure.

Equipment, environmental conditions and panel leader activities for sensory testing at assessor’s home are summarized in [Table 2](#).

Table 2

Summary of equipment, environmental conditions and panel leader activities to perform sensory evaluations at assessor’s home (RT-H).

Equipment and environment	
“Evaluation box” content	<ul style="list-style-type: none"> • three-digit coded vessels • white paper towel • napkins and cutlery • unsalted crackers
Evaluation station	<ul style="list-style-type: none"> • quiet room with possibility of air exchange • wide working surface • avoid cooking/household cleaner use 1 h before evaluation start
Panel leader “To do list”	
Days preceding evaluation	<ul style="list-style-type: none"> • evaluation box delivery • instruction to assessors for sample handling • instruction to assessors for video call connection
Before evaluation start	<ul style="list-style-type: none"> • collection of informed consent forms from assessors • participant list and document check • audio and video connection test • workstation equipment and organization inspection • sample positioning and sample order assistance • reminder to assessors about evaluation aim and modality
During evaluation	<ul style="list-style-type: none"> • reminder to assessors that talking is not allowed • data acquisition page in full screen mode • assessor’s microphone muted • monitor assessor behaviour
End of evaluation	<ul style="list-style-type: none"> • final page on assessor screen check • closing connection

2. Study 1 - discrimination tests

Tests were performed by internal panels of two flavour manufacturer companies using the assessor's workstation (their own desk) as RT location (RT-W). Personnel from the two companies independently selected the samples to be tested in the study considering both the availability of results from discrimination tests collected before pandemic in lab conditions and their quality control activities at the time of the study. The possibility to cover different types of discriminant tests was set as a further sample selection criterion. These considerations lead to the selection of samples considered in the present study with no preliminary assumptions on their relative perceptual differences.

2.1. Materials and methods

2.1.1. Participants

Two groups of assessors from company internal trained panels with previous experience in discrimination tests took part in the study. The number of assessors was set up according to recommendation for difference test selecting a Pd of 40%, α risk equal or lower than 0.05 and β risk within the range 0.1–0.05 (O'Sullivan, 2017). Two different panels participated in tetrad tests in the sensory lab ($n = 36$; 61% women; 40 ± 10.74 y.o.) and RT-W condition ($n = 36$; 53% women; 40 ± 9.07 y.o.), with eleven assessors that participated in both panels. In triangle tests the same assessors participated in both conditions ($n = 36$; 35% women; 27 ± 9.81 y.o.). A small gift was given to the assessors to motivate their participation in the study.

2.1.2. Samples

Lime flavours with raw material from two suppliers (LF1 and LF2) prepared in 5% sucrose solution were evaluated in tetrad test. Orange flavours with two different formulations (OF1 and OF2) prepared in 4% sucrose solution were evaluated in the triangle test. Samples (15 ml) were presented in disposable cups identified by a three-digit code. Samples did not differ for appearance according to a preliminary evaluation performed by the panel leader and the sensory lab personnel.

2.1.3. Evaluations

Assessors participated in one session held at the sensory lab (lab) and in one session in remote condition (RT-W). In the tetrad test, assessors were presented with one sample set consisting in two pairs of samples LF1 and LF2. The presentation order was randomized and balanced across subjects. One sample set consisting of three samples, two from the same and one from a different formulation of OF1 and OF2, was presented for the triangle test. All the six possible serving orders counter-balanced across assessors were presented.

Assessors were instructed to put the whole sample in their mouth and swallow after a few seconds. A short break was taken between samples. Water and plain crackers were provided for mouth cleaning between samples. Assessors were asked to identify the pair of identical samples or the odd sample in tetrad and triangle tests, respectively. They were requested to make a choice even in case of uncertainty.

Evaluations were performed under white light, with the only exception of the tetrad test at the sensory lab that was performed under red light since it was the standard condition for flavour evaluation at the Kerry sensory lab. Evaluations in lab conditions were performed in individual booths, while evaluations in RT-W were performed at the assessor's desk under natural or artificial light in offices shared by three-five employees doing their regular work (general administrative work).

Tetrad test data were collected using FIZZ Software Version 2.40G (Biosystèmes, Couternon, France); the FIZZ web version was used for remote testing. Triangle test data were collected using Compusense Software 3.8 (2021 Version).

2.1.4. Data analysis

The critical number of correct answers to conclude that samples were

perceptibly different was fixed at 18 ($\alpha = 0.05$) in both triangle and tetrad tests in all the conditions.

2.2. Results

Results from the tetrad test showed that LF1 and LF2 were perceived as different when evaluated in lab conditions, while no significant difference was found when evaluations were performed in RT-W conditions (Table 3). Results from the triangle test showed that OF1 and OF2 samples were perceived significantly different both in lab and RT-W conditions (Table 3).

3. Study 2 – Descriptive analysis (DA)

3.1. Materials and methods

3.1.1. Participants

Twenty-four trained assessors (90% women, mean age 45.0 ± 9.3 y.o.) who participated in several descriptive analyses of coffee samples in the two years preceding the study took part in the mocha coffee (MC) evaluations. The whole panel (P24) evaluated coffee samples in RT-H condition, while fifteen assessors evaluated coffee samples in lab conditions (P15).

Nine trained assessors (44.4% women, mean age 37.0 ± 13.15) with previous experience in descriptive sensory testing, mainly on olive oil, participated in hemp seed oil (HSO) evaluations.

3.1.2. Samples

Three mocha coffees from the market (MCA, MCB and MCC) from different batches were evaluated. Coffees were prepared using a three-cup mocha filled with 17 g of ground coffee and bottled water. Assessors received an "evaluation box" for RT-H sessions with all the materials needed for home coffee preparation including mocha, cups, spoons, pre-weighted samples and water, and crackers for rinsing procedure. Fourteen batches of MCA, MCB and MCC evaluated at the sensory lab facilities in 2018–2019 were considered for further comparison purposes.

Four hemp seed oils (HSO1, HSO2, HSO3, HSO4) were selected for evaluation to cover the sensory variability of HSO on the Italian market. The evaluation box for HSO tasting in RT-H condition included pre-weighted samples (20 ml) in three digit coded sealed containers, disposable glasses identified by the same three-digit codes, the same standards for colour and flavour attributes used for training, the evaluation sheet, napkin and crackers for rinsing procedure and instructions for sample storage (6–8 °C, room temperature 40 min before evaluation).

3.1.3. Evaluations

3.1.3.1. Coffee samples

3.1.3.1.1. *Re-training.* In 2020 panel trained for coffee descriptive analysis participated in three re-training sessions in the lab before restrictions due to pandemic and three in RT-H conditions during the pandemic. Re-training sessions were held two weeks before the evaluation sessions (lab condition: retraining on 8–10th January 2020, evaluation 29–30th January 2020; RT-H retraining 28–30th April 2020,

Table 3

Results from tetrad and triangle test performed in lab (LAB) and remote testing conditions at work (RT-W): number of correct answers and p values.

Test	Condition	N° of correct answers	p-values
Tetrad	LAB	19	$p < 0.05$
	RT-W	14	n.s.
Triangle	LAB	24	$p < 0.0001$
	RT-W	22	$p < 0.0001$

evaluation 12-13th May 2020). Re-training sessions consisted in the evaluation of standard solutions of sensory descriptors. The same method was adopted for re-training in RT-H and lab conditions, that consisted in two tasks: 1. recognition of eight coffee aroma standard solutions (honey, floral, tobacco, roasted, burnt, spicy, roasted cereal, citrus); 2. ranking of four sour taste standard solutions (citric acid: 0.012–0.095 mg/ml) and of three astringent standard solutions (alum sulphate: 0.05–0.20 mg/ml) presented at different concentrations. Standard aroma solutions were prepared to induce moderate intensity, sourness and astringency solutions were prepared to induce intensities from weak to strong on a 11-point scale.

3.1.3.1.2. Procedure. The trained panel participated in four sessions for MC evaluations; two sessions were held in the lab by the panel consisting of fifteen assessors (P15) and two sessions were held in RT-H conditions by the panel consisting in twenty-four assessors (P24, this panel consisted in the same fifteen subject participating in lab evaluation plus further nine assessors). Three samples (in replicates, i.e. assessed twice) were evaluated in each session. Samples were evaluated monadically. The order of the sample presentation was randomized among assessors using a balanced Latin square design. Coffee samples (30 ml) were presented in ceramic cups identified by a three-digit code in lab evaluation conditions. In RT-H conditions assessors were instructed to prepare one coffee sample per time and to use the provided ceramic cup for evaluation. Samples were served at $65^{\circ}\text{C} \pm 5$ in the lab, while serving temperature was not controlled in RT-H condition. The order of attribute presentation was randomized across samples by sensory modality to avoid the proximity error, that is the tendency of assessors to rate as more similar the descriptors “that follow one another in close proximity on the ballot sheet than those that are either further apart or rated alone” (Poste et al., 1991). Assessors were asked to smell samples for aroma evaluation and then to take a spoon of coffee sample for flavour and mouthfeel attribute evaluation. A 11-point category scale (0 = none/extremely weak, 10 = extremely strong) was used for intensity ratings. After the evaluation of each sample, subjects had a break of 90 s and were instructed to rinse their mouths with water and crackers. In RT-H conditions assessors were instructed to prepare the following sample after completing the mouth rinsing procedure.

MC data was collected using the RedJade software (RedeJade Sensory Solution LLC) both in lab and RT-H conditions.

3.1.3.2. Hemp seed oils

3.1.3.2.1. Training. Trained assessors for HSO evaluation participated in 20 training sessions held at the sensory lab (in 2019) consisting of the generation of a list of attributes describing HSO sensory profile (four sessions) and panel calibration (sixteen sessions). For term generation, assessors were asked to taste sixteen samples representative of the sensory variability of HSO on the market. Samples used for training were purchased both in large scale distribution and retail stores considering production process as the main selection criterion. Only oils obtained by mechanical or physical processes (i.e. labelled as “cold-pressed” or “obtained only by mechanical/physical processes”) were selected. The panel consensus was reached on a list of nine attributes: yellow, green, rancid, paint, roasted, fishy, sunflower/pumpkin seeds, toasted hazelnuts, and hay. An evaluation sheet was then set up following the official olive oil evaluation sheet layout (COI/T.20/Doc. No 15/Rev. 2 September 2007). To facilitate the calibration of the panel on descriptors, participants were familiarized with standard solutions prepared to induce a moderate intensity (corresponding to the central point of a 100 mm unstructured scale) of the seven flavour descriptors. Reference HSO samples were provided as a standard for yellow (two references corresponding to 20 and 60 on the scale) and green (two references corresponding to 50 and 80 on the scale).

3.1.3.2.2. Procedure. The trained panel participated in eight sessions for HSO evaluations; four sessions were held in the lab and four sessions were held in RT-H condition. Two samples (in replicates, i.e.

assessed twice) were evaluated in each session. Colour and flavour standard solutions were made available to assessors before evaluation. They were instructed to observe colour and smell flavour standards in order to help identification and ratings of the relevant sensations in HSO samples. In lab evaluation conditions, samples (20 ml) were presented in disposable glass coded with random three-digit codes. In RT-H conditions, assessors were instructed to fill the provided disposable glass with the corresponding sample (20 ml). The presentation order of the samples was randomized among assessors using a balanced Latin square design. Assessors were asked to take a sip and rate descriptors’ intensity on the paper evaluation sheet. An unstructured 100 mm scale was used for intensity ratings (0 = extremely weak; 100 = extremely strong). After each sample, subjects were asked to take a short break and rinse their mouths with water and crackers.

HSO data was collected on a paper evaluation sheet. In RT-H conditions assessors were asked to take a picture of the sheet using their mobile and send it to the Sensory Lab contact person.

Evaluations of MC and HSO in lab conditions were performed in individual booths under white light.

3.1.4. Data analysis

The effect of re-training conditions on the performance panel for MC evaluation was assessed by McNemar test for the aroma recognition task and by Kruskal-Wallis test for ranking tasks.

Intensity data collected in lab and RT-H conditions from MC and HSO samples were independently submitted to a 3-way ANOVA Mixed model (fixed factors: samples and replicates; random factor: assessors) with interactions. The effect of condition (lab and RT-H) on MC significant attributes was tested computing two independent two-way ANOVA models (factors: samples and conditions): the first one comparing the data of P15 in lab and in RT-H conditions and the second one comparing P15 in lab with P24 in RT-H. A further two-way ANOVA model (samples and conditions) was computed to test the effect of condition on HSO significant attributes. Post-hoc Fisher (LSD) multiple comparison tests were carried out to determine significant differences between samples in each condition.

In all studies the significance level was fixed at 95% ($p \leq 0.05$).

A Principal Component Analysis was performed on the mean intensity data of fourteen batches of the same three coffee samples evaluated at the sensory lab facilities in 2018–2019. Mean intensity data of MC evaluations performed in lab and in RT-H conditions (P24 and P15 assessors) were plotted as supplementary variables on the perceptual map.

ANOVA models and Principal Component Analysis were performed using XLSTAT, (version 2021.3.1, Addinsoft, NY, USA).

3.2. Results

3.2.1. Re-training

Results of the re-training on MC standards performed in lab and RT-H conditions were compared. No significant differences in the number of correct answers were found for the aroma recognition task ($p = 0.383$). Results of the ranking post hoc tests on sour and astringent solutions are reported in Table 4. The ranking of sour taste solutions did not differ in the two conditions: significant differences were found between 0.012, 0.026 and 0.04 g/l citric acid solutions while 0.04 and 0.095 g/l solutions did not significantly differ between them. The ranking of astringency solutions differed slightly between lab and RT-W. In lab condition, significant differences were found between all the three astringent solutions. In RT-W, 0.05 g/l aluminium sulphate solution significantly differed from 0.10 and 0.20 g/l solutions while 0.10 and 0.20 g/l solutions did not significantly between them.

3.2.2. Descriptive analysis

The comparison of results of the descriptive data on MC samples in lab and RT-H conditions indicates that the highest discrimination among

Table 4

Ranking task performed in lab (LAB) and remote condition at home (RT-H) for sour and astringency standard solutions. Different letters indicate significantly ranking differences ($p \leq 0.05$)

	Location test condition			
	LAB		RT-H	
	Frequency	Rank mean	Frequency	Rank mean
Astringency (aluminum sulphate g/100 ml)				
0.05	23	3.000a	23	2.783a
0.10	23	1.957b	23	1.783b
0.20	23	1.043c	23	1.435b
p-values		<0.0001		<0.0001
Sourness (citric acid g/100 ml)				
0.012	23	4.000a	23	4.000a
0.026	23	2.957b	23	2.739b
0.040	23	1.913c	23	1.739c
0.095	23	1.130c	23	1.522c
p-values		<0.0001		<0.0001

samples was observed in lab condition with 15 assessors for flavour and aroma and in RT-H condition with P24 for tastes (Figure 1). A significant sample effect on roasted flavour and aroma, astringency, bitter and sour taste attributes was found in lab (P15) and RT-H condition with the P24, while four attributes (roasted flavour, astringency, bitter and sour tastes) significantly discriminate among samples in RT-H condition with the P15. No significant effects of replicates ($p \geq 0.05$) and interaction assessor*sample were found ($p \geq 0.1$) with the only exception of an interaction effect for sour in RT-H condition in both panels (F value = 2.48; $p = 0.002$ for P24; F value = 2.68; $p = 0.006$ for P15). However, considering the low F value for these interactions compared to the F value for the product effect (F = 103.96 and F = 68.76 for P24 and P15, respectively), the interaction effect could be assumed as negligible (Næs et al., 2010).

The effect of condition on intensity of attributes significantly discriminating among samples was tested. The comparison between data collected with P15 both in lab and in RT-H conditions showed a significant effect of samples for all attributes ($p \leq 0.001$). The conditions significantly affected roasted aroma and flavour ($p \leq 0.002$) that were both rated significantly lower in the lab than in the remote conditions.

Interactions sample*condition were not found with the only exception of bitterness ($p = 0.002$); this attribute was rated as lower in the lab than in the remote conditions only in sample MCB (Figure 2a).

The comparison between data collected in the lab with P15 and in the remote conditions with P24 showed a significant samples effect for all attributes ($p \leq 0.001$), with the only exception of astringency ($p = 0.123$). The conditions significantly affected only roasted aroma ($p = 0.007$) that was rated significantly lower in the lab by P15 than in the remote conditions by P24. All interactions sample*condition were significant ($p \leq 0.017$) thus indicating changes in the relative differences among samples evaluated in the two conditions (Figure 2b). In lab conditions MCC showed the highest intensity of roasted aroma and flavour, astringency and bitterness, MCB and MCA did not differ in astringency while MCB was rated higher than MCA for roasted aroma, flavour and bitterness. In remote conditions, differences for roasted aroma and flavour among samples were less evident with no significant differences between MCC and MCB; MCB was the sample described by the highest intensity of astringency, MCA was still the sample with the lowest bitterness intensity while no significant differences were found between MCC and MCB. Sourness showed the highest intensity in sample MCA while only slight differences were observed between MCB and MCC both in lab and in remote conditions.

In order to explore the possible influence of the intrinsic sample variability on coffee sensory profile obtained in the two evaluation conditions, results from DA performed on fourteen batches of MCs (A, B and C) samples in lab conditions during 2018–2019 were compared to those considered in the present study. A Principal Component Analysis was computed on mean scores of the attributes describing the sensory profile of all MCs. The obtained perceptual map shows that the different batches of the three coffee samples are spread on the space thus indicating a within-sample variation for sensory profile (Figure 3). However, different batches of the same sample tend to be located in the same area. The first component (explained variance of 59.92%) discriminates coffee MCA, associated with sour taste, from sample MCC, mainly associated with roasted (flavour and aroma) and bitter. On the other hand, sample MCB tended to be positioned across the two sides of the map and its perceptual space superimposed to those of the other two samples, sample MCA on the left and MCC on the right. The second component (explained variance of 21.63%) discriminates batches of sample MCC according to astringency. In general, the same samples evaluated in lab and RT-H conditions in this study are positioned close to each other. They also fall in the perceptual space occupied by samples of batches

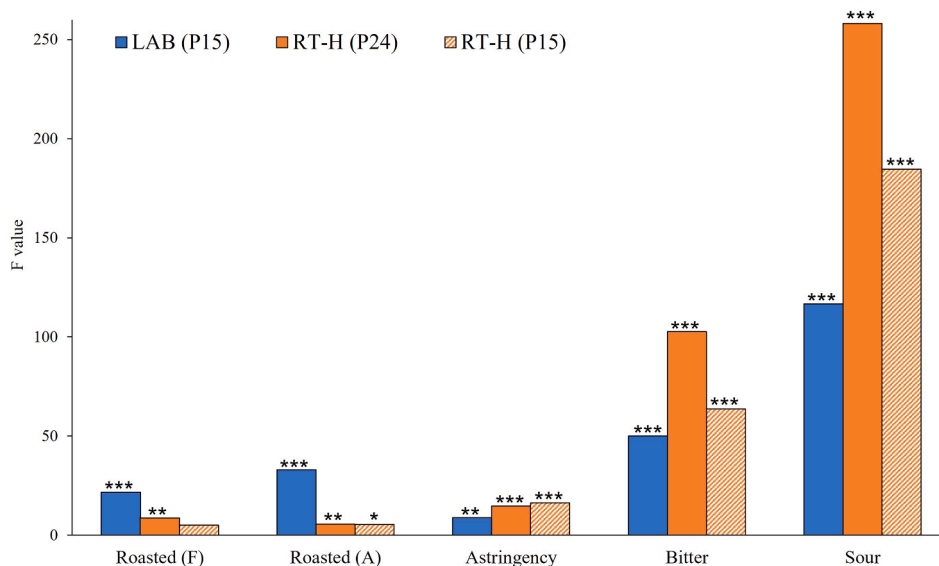


Fig. 1. Three-way ANOVA mixed model on mocha coffees (MC) intensity data: F-values of sample effect from lab (LAB) and remote sessions at home (RT-H) considering the whole panel of 24 assessors (P24) and a panel of 15 assessors (P15). Significance: ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

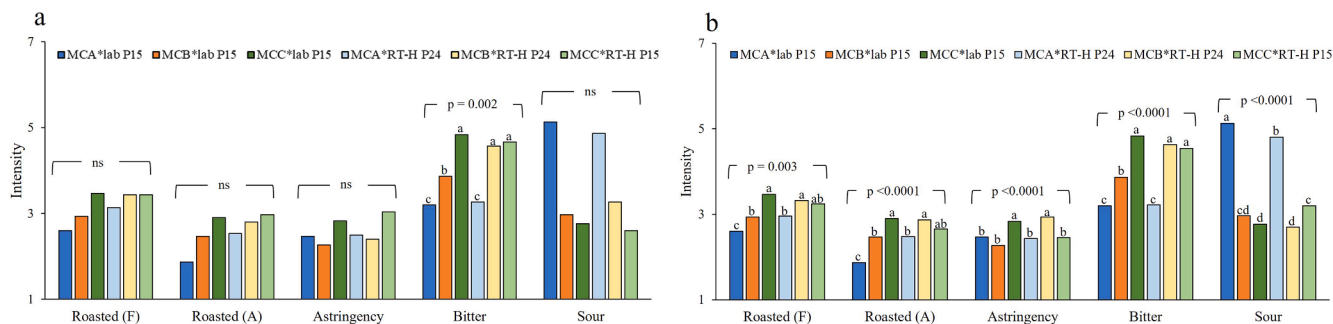


Fig. 2. Two-way ANOVA model on mocha coffees (MCA, MCB and MCC) intensity data: a) Mean scores of P15 panel in lab (lab P15) and in remote sessions at home (RT-H P15); b) Mean scores of P15 panel in lab (lab P15) and P24 panel in remote sessions at home (RT-H P24). P-values for product*condition interaction effects are reported. Different letters indicate a significant difference at $\alpha = 0.05$ as determined by Fisher's least significant difference (LSD).

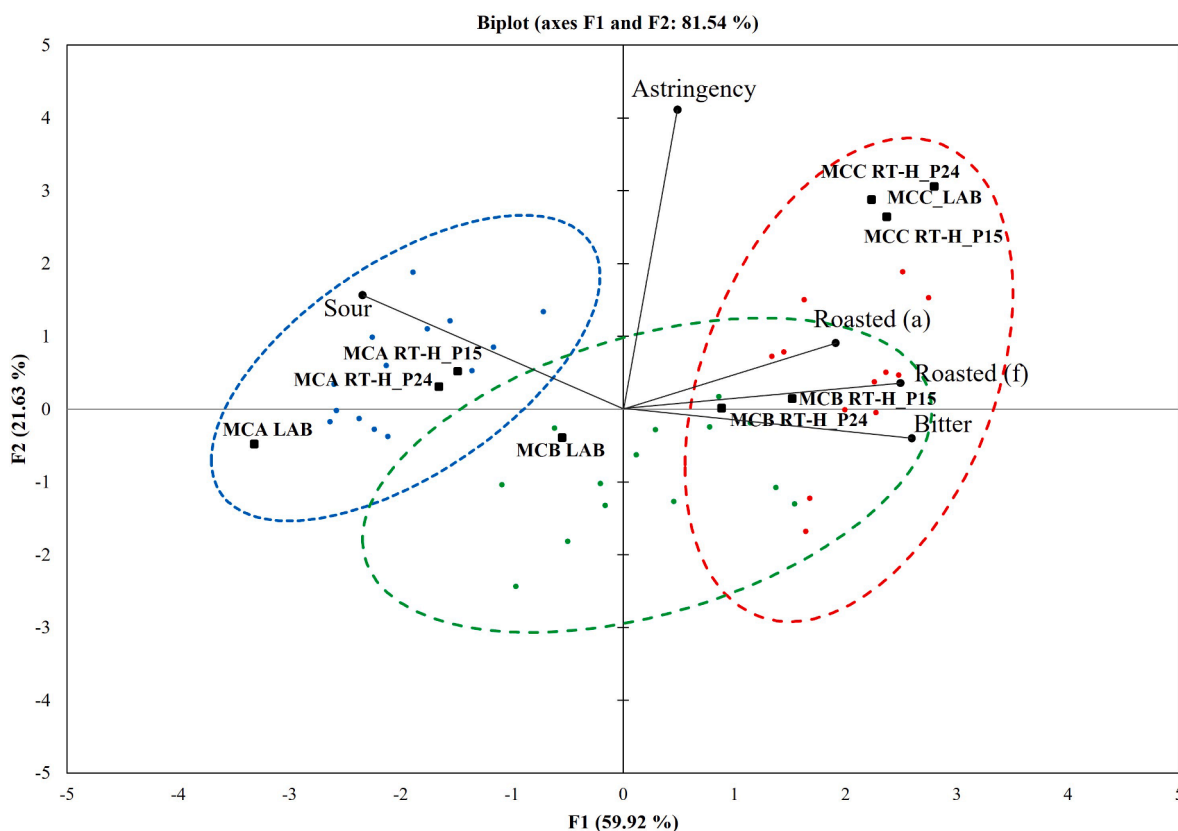


Fig. 3. Bi-plot of Principal Component Analysis on Descriptive analysis data of 14 batches of coffee: MCA (blue dots), MCB (green dots) and MCC (red dots), tested in lab conditions during 2018–2019. Confidence intervals are reported. Coffee samples evaluated in the present study in the two conditions (lab conditions: LAB; remote with the whole panel: RT-H P24; remote with the panel of 15 assessors: RT-H P15) were plotted as supplementary variables (in black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

belonging to the same product previously evaluated in the lab. The only exception is represented by sample MCB evaluated in lab condition that is located on the left side of the first component in opposition to the corresponding samples evaluated in remote condition (MCB RT-H P15, and P24). This difference is mainly driven by the sensory opposition sour taste vs bitter taste, roasted aroma and flavour.

Analysis on data from HSO evaluations separated by condition showed a significant sample effect on the same six (yellow, green, fishy, pumpkin seeds, toasted hazelnut and hay) out of eight attributes ($p \leq 0.002$). A significant assessor*sample interaction was found for fishy in lab conditions ($F = 2.20, p = 0.029$). However, considering the low F values for this interaction compared to the F values for the product effect ($F = 41.27$), it could be assumed that the interaction effect is negligible (Næs et al., 2010). F values for fishy, toasted hazelnuts and hay were

very similar in both conditions; F values for yellow was higher in lab than in RT-H condition while the opposite was observed for green and sunflower seeds (Figure 4).

The effect of condition on HSO sample evaluation for discriminating attributes was further investigated. The comparison between data collected in the lab and in RT-H conditions showed a significant effect of samples for all attributes ($p \leq 0.001$). Furthermore, a significant effect of evaluation conditions ($F = 4.4; p = 0.038$) and interaction sample*condition ($F = 5.7; p = 0.001$) for yellow was reported: no significant difference in yellow intensity evaluated in lab and remote conditions were found for HSO2 and HSO4 while HSO1 and HSO3 were rated higher in remote than in lab conditions (Figure 5).

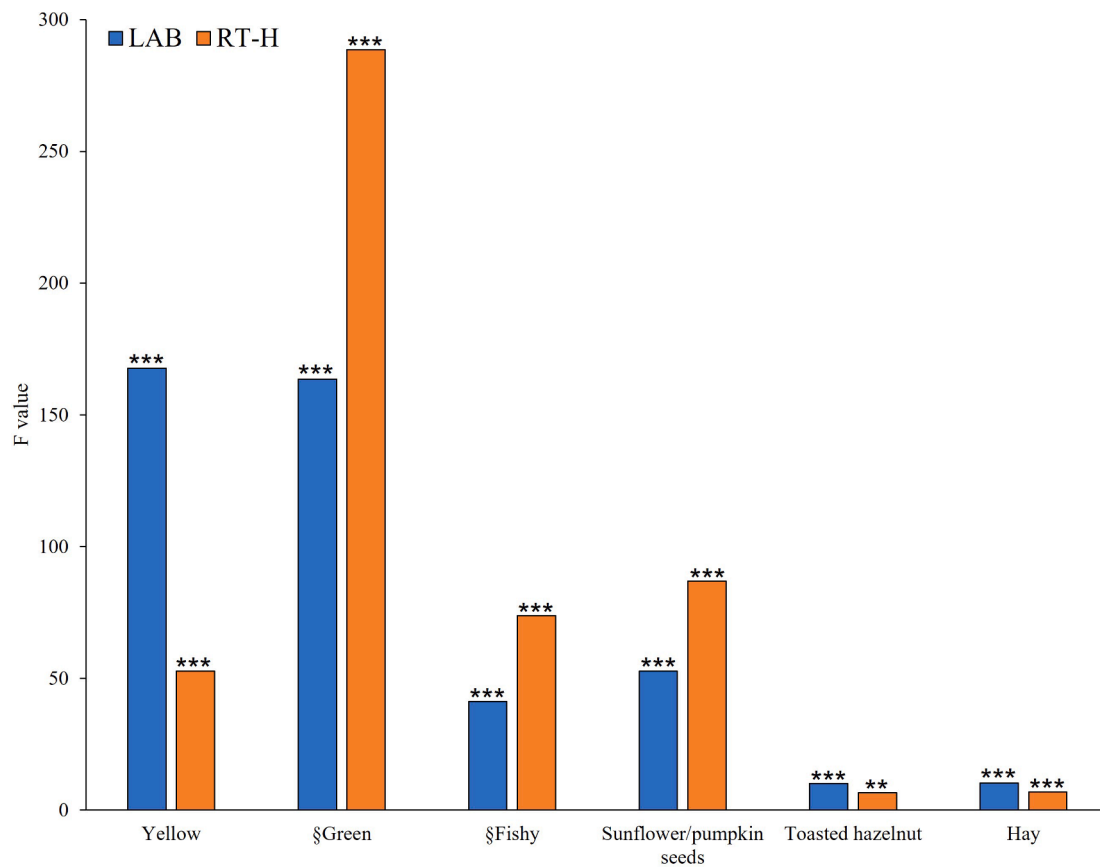


Fig. 4. Three-way ANOVA on hemp seed oil (HSO) intensity data: F-values of sample effect in lab (LAB) and remote sessions at home (RT-H). Significance: ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$. § indicates that F value was $\times 10^{-1}$.

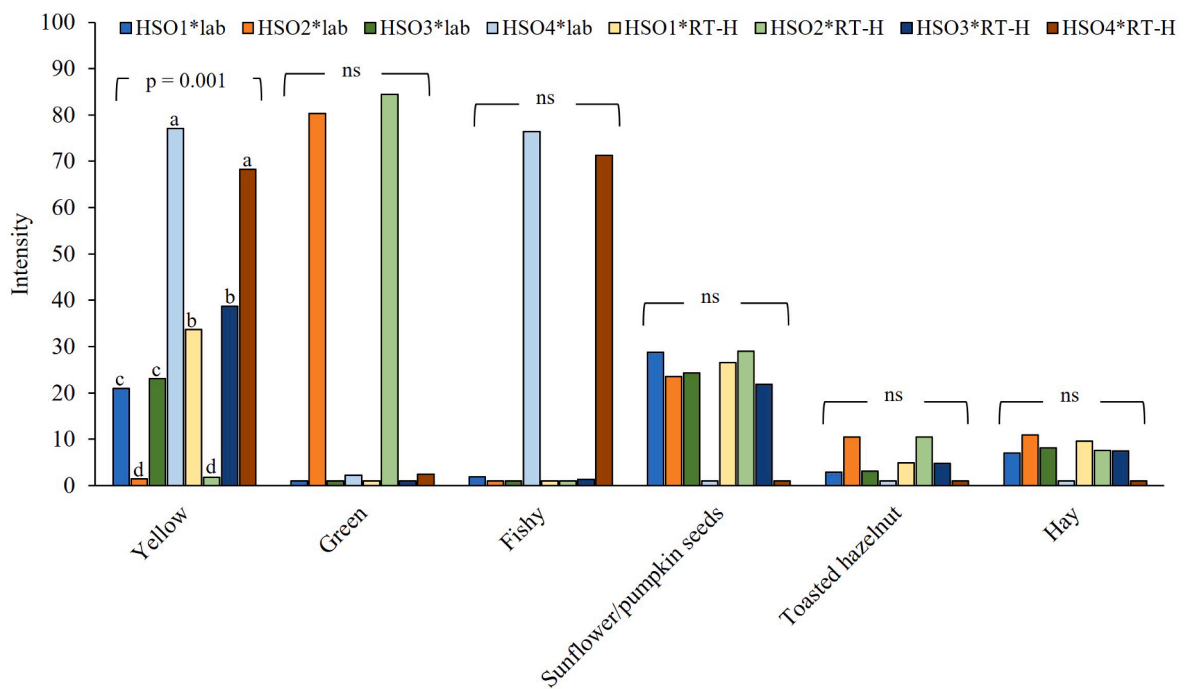


Fig. 5. Two-way ANOVA model on hemp seed oils (HSO1, HSO2, HSO3 and HSO4) intensity data: comparisons of mean sample scores between lab (LAB) and remote sessions at home (RT-H). P values for product*condition interaction effects are reported. Different letters indicate a significant difference at $\alpha = 0.05$ as determined by Fisher's least significant difference (LSD).

4. Study 3 - Temporal dominance of sensations (TDS)

4.1. Materials and methods

4.1.1. Participants

Nine trained assessors (5 women, mean age $36,4 \pm 11.1$ SD) took part in the evaluations in both lab and remote conditions.

4.1.2. Samples

Three 100% xylitol chewing gums were evaluated: Green Mint (GM), Spearmint (SP) and Peppermint (PE) (Vivident®Xylit, Perfetti Van Melle, Lainate, Italy). The “evaluation box” for RT-H included samples each one wrapped in an aluminium foil film, coded with three-digit codes.

4.1.3. Evaluations

4.1.3.1. Training. Assessors participated in three training sessions held at the sensory lab consisting, respectively, in the generation of a list of attributes describing the dominant sensations in chewing-gum samples, panel calibration, and test simulation. In the first session, the concept of dominance was explained to the assessors as “the attribute associated with the sensation catching the attention of the assessors at a given time, not necessarily being the one with the highest intensity” (Bruzzone et al., 2013). For term generation, assessors were asked to taste five samples (including the three considered in the study), indicate the dominant sensations, and describe their temporal evolution. After a common discussion, the panel consensus was reached on a list of ten attributes: crunchiness, mouth coating, elasticity, firmness, juiciness, sweetness, mint flavour, freshness, pungency, bitterness. To facilitate the calibration of the panel on descriptors of dominant sensations, participants were familiarized with bitter taste, pungency, and mint flavour standard solutions. All the solutions were prepared to induce a moderate intensity of target sensations, corresponding to the central point on a nine-point scale, based on preliminary tests. The last training session was performed to familiarize participants with the software procedure for data recording. Three samples were evaluated. Assessors were trained to click on the “Start” button as soon as they start chewing and to immediately start the selection of dominant attributes. They were also told that not all the attributes have necessarily to be selected as dominant and that a given attribute could be selected as dominant several times during the evaluation.

4.1.3.2. TDS procedure. The temporal evolution of sensations induced by each sample was described using the ten-attribute list. The presentation order of attributes was randomized across participants but was always the same for a given assessor for all samples. In total, assessors participated in four evaluation sessions; the first two sessions were performed in the lab and the third and fourth sessions were held in RT-H condition. Three samples were evaluated in each session. Each sample was assessed two times. Samples were presented to participants wrapped in an aluminium foil, coded with random three-digit codes. The presentation order of the samples was randomized among assessors using a balanced Latin square design. Assessors were asked to put the sample in their mouths, begin chewing and start the evaluation; after 300 s assessors were prompted by a screen signal to spit out the gum and to continue the evaluation for further 120 s. The total evaluation time was 420 s. Data were recorded every 1 s. After each sample, subjects rinsed their mouths with water for 3 min, then waited at least 12 more minutes before continuing the test. Evaluations in lab conditions took place in individual booths under white light. Data were collected by using the FIZZ Software version 2.51 c02 (Biosystèmes, Couternon, France) and automatically plotted as TDS curves (Dominance Rate – DR vs evaluation time).

A virtual environment based on VMWare® vSphere Web Client was

used to make available the FIZZ terminal to assessors in RT-H condition. Assessors had available the connection to the FIZZ terminal through the web using HTML5 with their own computer, without installing additional software and could use their mouse and keyboard to control their virtual terminal. This software configuration allowed the panel leader to control and monitor sessions and individual terminals; the panel leader access to each terminal was allowed if necessary.

4.1.4. Data analysis

Total evaluation time (420 s) was split in seven time periods of 60 s each, five referred to the chewing phase (1. 0–60 s; 2. 61–120 s; 3. 121–180 s; 4. 181–240 s; 5. 241–300 s) and two referred to aftertaste (6. 301–360 s and 7. 361–420 s). TDS curves were visually inspected and for each sample, in each time interval, only attributes with a dominance rate higher than the chance level were selected for further analysis. TDS difference curves were computed for each sample based on differences in dominance rates recorded in lab and RT-H conditions. The limit of significance for each difference curve over time was obtained using the test to compare two binomial proportions as reported by Pineau et al. (2009). To assess differences among samples in the frequency of selection of dominant attributes in the two evaluation conditions data were pre-processed according to Dinnella et al. (2013). Raw data of dominant attributes (1 = selected; 0 = not selected) were downloaded, and the frequency values were computed for each attribute in each time interval, each sample, each assessor and each replicate. Frequency values for the same attribute dominant in different samples were summarized for each assessor, in both lab and RT-H conditions, in each time interval. To the purpose of identifying assessors with a low ability to detect differences between samples, for each attribute evaluated in the two conditions, in each time interval, frequency values of each assessor were independently submitted to a one-way ANOVA considering sample as a factor. A p -value ≤ 0.2 was fixed as an assessor inclusion criterion. According to this criterion all assessors were included. A two-way ANOVA model with interaction (fixed factors: sample and conditions) was applied on frequency of dominant attributes on each time interval, independently (Tab. 1 in Supplementary Materials).

Data analysis was performed using R software (Version 4.0.3) and tempR package (version 0.9.9.16) was used for TDS difference curve computation.

4.2. Results

Visual comparison of TDS curves for each sample in the two evaluation conditions (lab and RT-H) showed similar evolution patterns (Fig. 1S-6S in Supplementary materials). The most part of texture-related descriptors (crunchiness, firmness, juiciness) dominated the beginning of the chewing phase (1–60 s) while elasticity dominated the last part of the evaluation (60–300 s) in all samples. Sweet taste was above the chance level for almost the whole evaluation time in the GM and SP samples, while it was dominant from 120 s to the end in the PE sample. Mint flavour and freshness dominated the whole evaluation time in all the samples. Pungency showed a sample-specific evolution pattern. This sensation dominated only the beginning of evaluation in the GM (0–60 s) and PE (0–120 s) samples, while it dominated the SP sample's profile to a lesser extent: it was close to the chance level from the beginning to around 300 s in lab condition and for short periods at the beginning (around 30 s) and at the end (around 280 s) of the evaluation in RT-H condition. Bitter taste and mouth coating never reached the chance level in either of the two conditions.

The TDS curves of the difference between each sample evaluated in lab and RT-H conditions indicate very limited differences in the dominance rate (Figure 6a-c), where difference values significantly higher than zero indicate dominance rate values higher in the lab than in the remote conditions, and vice-versa. Dominance of bitter taste at around 100 s and juiciness at around 400 s of the evaluation time of GM sample was rated significantly higher in remote than in lab conditions. In

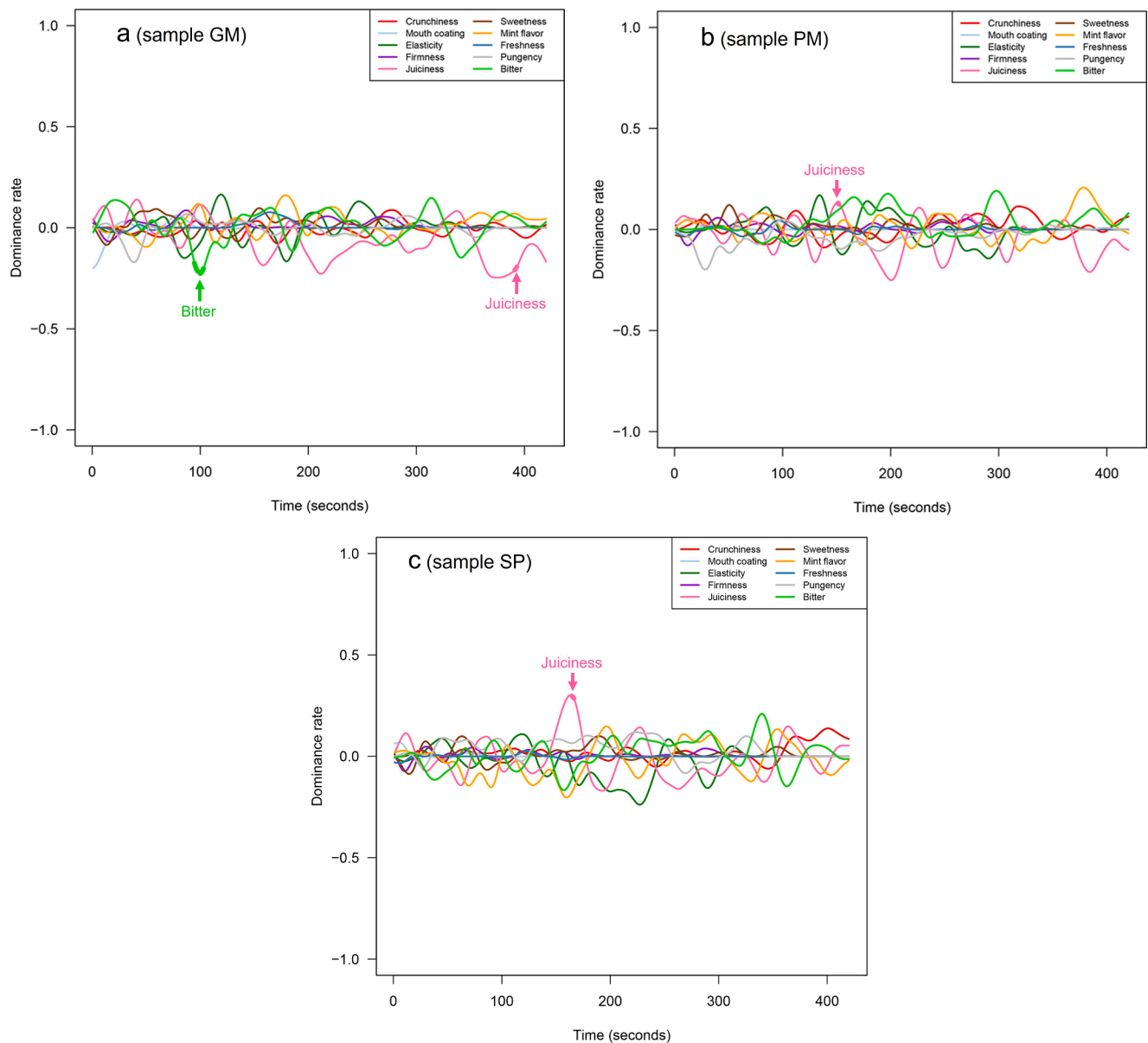


Fig. 6. Temporal Dominance of Sensations (TDS) difference curves computed for Green Mint (a, GM), Peppermint (b, PM) and Spearmint (c, SP) samples. Bolded lines (highlighted by arrows) indicate values significantly different than zero. Difference values significantly higher than zero indicate dominance rate values higher in lab than in remote conditions; difference values significantly lower than zero indicate dominance rate values higher in remote than in lab conditions. Level of significance $\alpha = 0.05$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

samples PE and SP the dominance of juiciness was higher in the lab than in remote conditions at around 150 s of the evaluation time. In all cases the dominance differences persisted only for a very short time thus indicating the almost total overlap of TDS curves in the two conditions.

Differences in the dynamic profiles among the three samples in the two evaluation conditions were further explored. Effects of samples, evaluation conditions, and their interactions on the frequency of dominant attributes during the evaluation periods are reported in Table 1 in Table 1 in Supplementary Materials (Tab. 1S). At the beginning of the chewing phase samples significantly differed for mint flavour ($p = 0.002$) and pungency ($p < 0.0001$) (period 1), elasticity ($p = 0.036$) sweetness ($p = 0.006$) and freshness ($p < 0.0001$) (period 2). Evaluation conditions did not significantly affect the evaluation during chewing nor significant sample*condition interactions were found. During aftertaste (period 6) freshness significantly differed among samples ($p < 0.0001$) and was significantly higher in lab than in RT-H conditions ($p = 0.042$). No significant sample*condition interactions were found during

aftertaste.

5. Study 4 - Check all that Apply (CATA)

5.1. Materials and methods

5.1.1. Participants

Two consumer groups participated in lab ($n = 60$, women 75%, mean age 35 ± 1.9 , celiac disease and gluten intolerance 53%) and RT-H ($n = 60$, women 77%, mean age 38 ± 1.5 , celiac disease and gluten intolerance 53%) testing conditions. Evaluations in the lab were run in 2018, while evaluations in RT-H conditions were held in 2020.

5.1.2. Samples

Five gluten free sandwich breads from the market were evaluated (SC, CR, NF, ES, OR). The “evaluation box” included two sample sets each in a paper bag labelled as “liking” and “sensory description”. Each

sample was sealed in a single plastic bag identified with a three-digit code. Samples in the two conditions came obviously from different batches.

5.1.3. Evaluation procedure

Each consumer group participated in one evaluation session consisting in two sub-sessions: the first one for liking evaluation and the second one for describing bread sensory properties using a Check-All-That-Apply (CATA) with forced choice (yes/no) methodology (Ares et al., 2010). Samples were placed in a white plastic dish immediately before evaluation in lab condition while they were presented in a single sealed plastic bag in RT-H conditions. Sensory lab expert personnel performed preliminary evaluations to exclude perceptual changes due to differences in the presentation between lab and RT-H conditions. Samples were identified by three-digit codes both in lab and RT-H conditions. Two independent sample sets, each consisting of the five bread samples, were used for liking and CATA evaluations. The presentation order of the sample was randomized among assessors using a balanced Latin square design. Assessors were asked to take a bite of the bread slice and to express their liking on a 9-point category scale (1 = dislike extremely; 9 = extremely like). A list of 23 descriptors of texture by appearance (porous, spongy, soft, moist, dry), taste (bitter, salty, sweet, sour), flavour (cereal, beans, honey, yeast, walnuts, stale, cardboard, chestnuts, sesame) and texture in mouth (gritty, sticky, chewy, puckering) was used for CATA evaluation. The order of attributes was randomized by sensory modality (appearance, taste and flavour, texture) across participants. Consumers were asked to describe sample appearance first, take a bite and evaluate taste and flavour, and then texture in mouth after a second bite. After each sample, subjects rinsed their mouths with water for 60 s.

Evaluations in lab conditions were performed in individual booths under white light.

FIZZ Software version 2.51 c02 (Biosystèmes, Couternon, France) was used to collect data in lab conditions while Compusense (version 20.0.7557.33837, Compusense Inc., Guelph, ON, Canada) was used for RT-H data collection.

5.1.4. Data analysis

Liking data in lab and RT-H conditions were independently submitted to a one-way ANOVA model to assess differences among samples in both evaluation conditions. A two-way ANOVA model (Fixed factors:

sample and condition) with interactions (sample*condition) was applied to assess the evaluation condition effect on liking. Post-hoc Fisher (LSD) multiple comparison tests ($\alpha = 0.05$) were employed to determine significant differences among samples.

Cochran Q-tests were performed on the CATA data to assess the differences between the frequency of attribute selection among samples. Post-hoc pairwise comparisons were calculated using the Sheskin method with the level of significance set at 5%. Contingency tables reporting the frequency of each descriptor for each sample in lab and remote testing were built from CATA data and chi-square statistic was applied to test the independence between rows and columns (significance level set at 5%). Multiple Factor Analysis (MFA) was computed on the two contingency tables (lab and RT-H condition) reporting the frequency of selection of attributes that significantly discriminated among samples according to the Cochran's Q test. XLSTAT, (version 2021.3.1, Addinsoft, NY, USA) was used for data analysis.

5.2. Results

Significant mean liking differences among samples were found in both conditions (lab $F = 8.34$; $p \leq 0.001$; RT-H $F = 4.35$; $p = 0.002$) (Figure 7). Discrimination among samples appears slightly higher in lab than in RT-H conditions. SC and CR were the most liked samples in both conditions, liking for NF was significantly lower than liking for SC in lab conditions while the two samples did not differ in liking in RT-H condition; OR was liked significantly less than NF in lab while the samples did not differ in RT-H. Conditions for data collection did not significantly affect results ($F = 0.170$, $p = 0.680$) nor significant sample*condition interaction effect was found ($F = 0.540$, $p = 0.706$).

Sample description using the CATA questionnaire resulted in eighteen attributes in total (17 in the RT-H and 16 in the lab condition) significantly discriminating among samples ($p < 0.05$) for texture by appearance (moist, dry, spongy and soft), tastes (salty, sweet, bitter and sour), flavour (cereal, honey, walnuts, stale cardboard and alcohol) and texture in mouth (gritty, sticky, chewy and puckering). The RV and Lg coefficients indicate a moderate correlation between the two conditions (RV = 0.58; Lg = 0.61). The first dimension is closely linked to each of the two conditions, thus representing an important direction of inertia in both of them. The two conditions are almost exactly superimposed in the relationship square figure (based on Lg coefficients) showing that, for the most part, the samples were evaluated similarly in the two

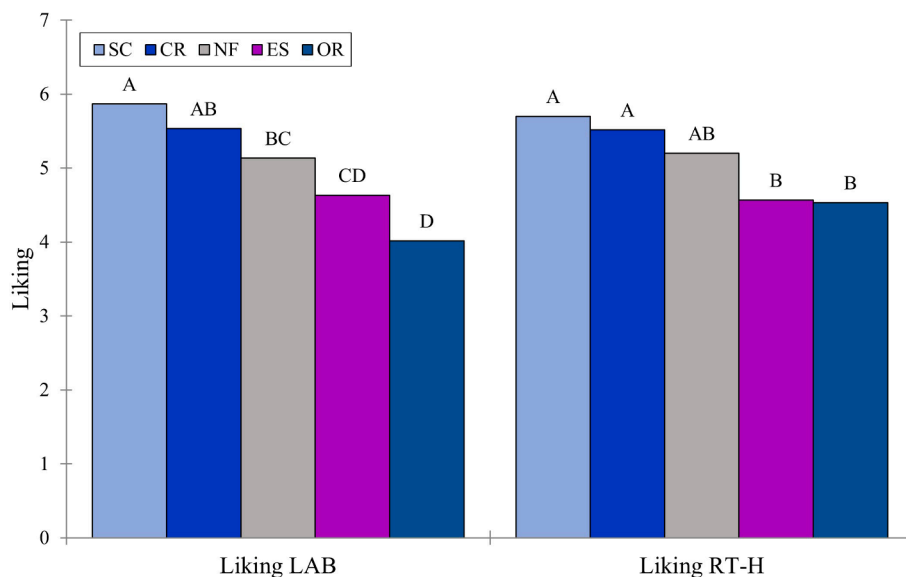


Fig. 7. Mean liking expressed by consumers for gluten free bread samples in lab (LAB) and remote testing conditions at home (RT-H). Different letters indicate significant differences after Fisher LSD post hoc test ($p \leq 0.05$).

conditions (coordinates Dim 1 RT-H: 0.85; lab:0.85 Dim 2: RT-H: 0.3; lab: 0.29).

Sample mean points and their partial individual representations according to CATA data in lab and RT-H conditions are shown in the MFA space (Figure 8a); attributes are represented on the first two dimensions of the MFA (Figure 8b). The first two factors accounted for 82.4% of variability. The first axis opposed CR and SC samples to ES and OR. ES and OR are characterised by dry and gritty texture, sour taste and cardboard and stale off flavours, while CR and SC are described by sweetness and spongy texture. Sample configuration on the first dimension looks quite similar in both conditions, with the exception of sample NF, characterised by walnut flavour, which is evaluated as more similar to OR and ES only in the lab conditions, thus less sweet, more bitter, drier and grittier than in the RT-H conditions. Some smaller differences along the second dimension were observed for samples SC, ES, OR. Both samples were described as spongy, chewy, soft and with a moist texture, with sweet taste and honey flavour in both conditions, but SC was associated with a stickier texture than CR in RT-H condition while it was assessed as more similar to CR in the lab conditions. ES and OR samples were evaluated as the most bitter, least sweet, grittiest and drier samples in both conditions, with very small differences. In the lab conditions they were also evaluated as characterised by a cereal flavour, together with the CR sample, while in the lab-condition the samples were poorly discriminated by this variable compared to the other samples.

6. Discussion

These findings showed that remote sensory testing provided similar results to the lab testing in all the cases that were considered (triangle test; Descriptive Analysis; Temporal Dominance of Sensations; Check-All-That-Apply), with the exception of the tetrad test. A discussion of the single studies is presented here including references to previous research, if possible, although it should be considered that the conditions of the quoted studies are different from the ones adopted in this study, thus making detailed comparisons difficult.

Results of lab conditions indicated significant differences among

samples both in triangle and tetrad tests. The number of correct answers in lab condition was definitely higher than the chance level for orange flavour samples in the triangle test with almost 67% of participants who correctly identified the odd sample. In the tetrad test with lime flavours in lab conditions the number of correct answers (19) was close to the chance level (18). These results might indicate clear perceptual differences among orange flavour samples and only subtle perceptual differences among lime flavour samples. Eventual noises in RT-W evaluation conditions seem more likely to bias the results of the evaluation of samples with small perceptual differences.

Furthermore, we may hypothesise that the lower discriminant ability found in the tetrad test case can be also attributed to the environmental conditions. In the case that was considered, a large source of noise could be related to the presence of odours in the work environments, very frequent in a flavour manufacturer company. The strict control for ambient odour while running sensory evaluation is highly recommended to control for olfactory adaptation and consequent changes in odour and flavour perception (Lawless and Heymann, 2010). Moreover, the presence of other employees engaged in different activities may have held assessor's attention during evaluation. These factors might lower assessor discrimination ability preventing the perception of differences among samples, especially when they are subtle as in the case considered here. This suggests that, when the workplace desk is selected for the evaluation in remote conditions of samples with subtle differences, the test should be planned when the presence of other employees is reduced and the office quieter (i.e. during lunch break, immediately after the closing time...). It may help making other employees aware that sensory evaluations are ongoing and asking them to refrain from activities that can distract assessors from their task is another possible option. Finally, food, but also non-food, manufacturers may consider conducting remote testing at home rather than in the offices according to the guidelines proposed in the present work, as in this way a larger control is ensured, particularly when the production areas/labs are close to the offices thus making difficult a control of the odour spread. In this case, stricter advice to control for environment odour biases (i.e., increasing the time between cooking/using household cleaners and evaluation, longer air exchange before running the test) should be provided. This could

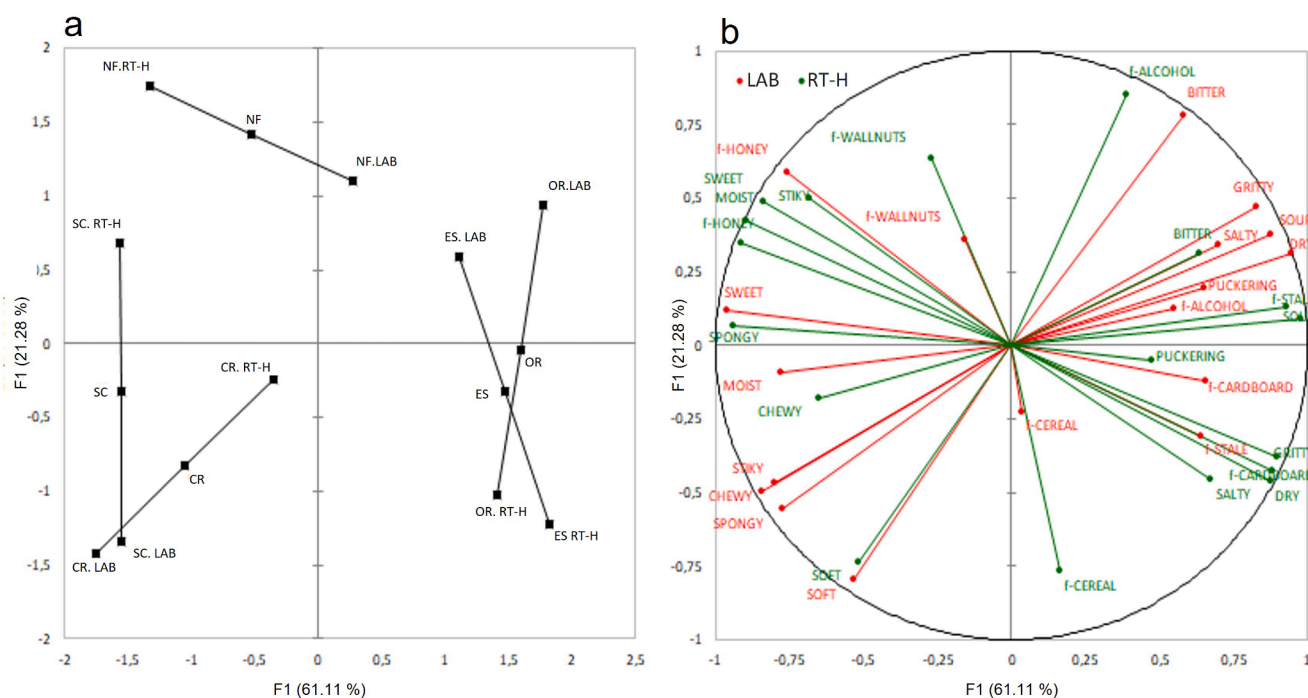


Fig. 8. a) Multiple Factor Analysis (MFA) space (first and second dimensions) showing the five bread samples as mean points (OR, ES, CR, SC and NF) and their partial individuals representing descriptor scores in the two evaluation conditions (RT-H and LAB). b) Representation of descriptors on the first plane of the MFA.

improve test reliability, particularly when evaluating at assessor's home samples with subtle differences.

The panel composition represents another point to consider in the interpretation of results of the discriminant tests. The same assessors took part in both lab and RT-W triangle tests (within-subject design) while the panel composition was different in the tetrad test, with only 30% of assessors participating in both evaluations (between-subject design). The within-subject design adopted in the triangle tests likely minimized the random noise of independent variables related to the location where the evaluation took place (i.e. odour bias, ambient noise), thus allowing for detecting significant differences both in lab and RT-W conditions. On the contrary, further random noise due to individual variability might have covered the subtle differences among samples in tetrad test run in RT-W condition. Thus, the within-subject design, when applicable, seems the more advisable experimental design when results from sensory tests performed in different conditions have to be compared.

Descriptive analysis case studies were intentionally selected to cover two different sample types. Coffee was selected as a "more difficult case" in which the sample is prepared at the assessor's home, thus requiring the experimenters the effort to develop a detailed procedure to standardise sample preparation (including the specification of all the materials needed). This type of product requires extra-time and higher engagement of the assessors. Hemp seed oil was selected as a "simpler case" in which samples are delivered as ready to be evaluated. Results from Descriptive Analysis indicated that sample description was similar between the two conditions in the case of hemp seed oils. The comparison between lab evaluation with P15 and RT-H evaluation with P15 and P24 indicate similar discrimination with slight but significant variations in coffee description. Various sources of variability can explain these results. Coffee preparation procedure step was out of panel leader control and even slight deviations from the instructions that might have occurred could have had an impact on the evaluation introducing noise in coffee sensory data and thus affecting discrimination among sample and their sensory description. Moreover, the comparison of similarity and differences among coffee samples from different batches showed a wide range of variation of their sensory profile, in particular for sample MCB. Therefore, the intrinsic variability of coffee sample sensory profile could further account for the obtained results. The perceptual map showed that the same samples evaluated in lab and remote conditions are positioned close to each other and share the perceptual space with the other batches of the same sample, thus indicating that sensory profiles from the two conditions are comparable. Possible strategies to face these issues in the case of highly variable samples could include an increase of the number of replicates. However, time and engagement required with the assessor to prepare and perform the evaluation should be taken into account to avoid fatigue and boredom particularly with "difficult case" samples. Furthermore, the individual variability related to differences in panel composition might account for the variation observed in coffee sensory description when comparing lab and RT-H evaluation with the larger panel. This further indicates that the within-subject design would be the more advisable option when comparing results from different evaluation conditions.

Results from hemp seed oil descriptive analysis indicated that the sensory profile of samples obtained in the two conditions are very similar, as well as the assessor performance. These findings clearly indicate that descriptive analysis in remote conditions represents an alternative to the lab evaluation when samples are provided ready to be used. However, a significant effect of the evaluation conditions on colour evaluation was found. Light conditions at home cannot be easily standardized. When colour represents a target attribute, further effort should be made by experimenters to provide assessors with appropriate light devices to overcome this source of variability in respect to the lab conditions.

Taken together, results from descriptive analysis indicate a substantial match between sensory descriptions obtained in classical

laboratory settings and in remote conditions at the assessor's home. This confirms previous evidence of the reliability of descriptive data from food sensory evaluations carried at the assessor home (Martin et al., 2014). Previous research using an Internet panel found some disagreement between the results of the Internet and lab reference panels and a poorer performance of the former panel (Nogueira-Terrones et al., 2006). In their study however the assessors were left free to choose the time for the evaluations and the contacts with the panel leader were only possible by phone and email (so not live). The results of our study show instead that live remote testing ensures satisfactory performance of the home panel. It should also be noted that Nogueira-Terrones et al. (2006) recruited new assessors and the whole process of training was done online, while in our case a re-training of the panel was performed. This suggests that training is a key factor when performing a remote descriptive analysis.

TDS curves obtained from chewing gum samples in both lab and remote conditions were similar, as well as the significant differences among samples. Thus, from the present findings it appears that TDS in remote conditions could be an alternative to the lab testing. This further strengthens previous evidence showing the coherence of results of other dynamic sensory tests between the evaluations at the assessor home (but without panel leader assistance) and in a classical lab setting (i.e. Progressive Profile vs Time Intensity) (Galmarini et al., 2016). Furthermore, the spreading of the application of TDS evaluation with consumers (Schlich, 2017) might be further boosted by the possibility of collecting reliable TDS profile from consumer's homes rather than in classical lab setting, particularly in the case of products that are not affected by serving temperature or that do not require intensive familiarisation with the attributes. A relatively high effort is required to make the software for data acquisition available to the assessors to allow panel leaders monitoring the evaluation according to the guidelines proposed in the present work. Thus, at present, this may represent a limiting factor for sensory teams with low skills and familiarity with the computers.

Consumer tests on gluten free bread samples gave very similar results in lab and remote conditions both for liking evaluations and sensory description using a CATA methodology. The lack of significant effect of condition (lab vs RT-H) on liking for bread supports the hypothesis that the methodology adopted for performing the test at the assessor home makes the context more similar to the standard lab setting than to home use test, for which significant changes in hedonic ratings would have been expected (Boutrolle et al., 2007). The slight differences observed in bread sensory description between lab and remote conditions, mainly limited to one sample (NF), are likely to reflect an actual change in their sensory profile due to slight formulation changes rather than differences due to the conditions for data acquisition. In fact, the lab test was performed two years before remote testing and in particular NF samples at the time of remote testing were no longer available on the Italian market (a version for a foreign market was used instead). Furthermore, panels participating in the study were different and, despite the control for their consistency in terms of demographic (gender and age) and health status (% of celiac disease and gluten intolerance), individual variability among assessors might have contributed to the small differences observed in sample description in the two conditions. This may explain for example why the RT-H panel was more discriminant than the lab panel in some attributes related to flavour, such as cereals and walnuts. The results are consistent with previous research, showing that CATA evaluation performed at the assessor's home allowed for sample discrimination according to their appearance and flavour (Mahieu et al., 2020). Furthermore, our results are also coherent with previous findings from Jaeger et al (2013) that showed that attribute use and product characterizations were stable in sensory CATA with consumers in test-retest comparison (with up to 30 days of interval).

Taken together, these findings suggest that remote sensory testing, if appropriately conducted following meticulously the indicated sensory protocols, is a promising technique that can be applied even beyond the global pandemic. This is relevant in consideration of the fact that the

expected trend for the future is to have more and more employees that will work from home at least for some days per week (ISMEA, 2021). Table 5 lists the main differences between sensory laboratory, remote testing and home-use-test conditions. The main advantage of the sensory laboratory condition is its ability to control sources of variability, noise, and distraction for the assessors enabling greater sensitivity to small, but robust effects. Home Use Tests have been conceived for consumer testing and not for tests with trained assessors, in order to ensure an evaluation of products in a more natural environment (Cardello & Meiselman, 2018). Remote testing appears as a promising methodology that can be applied both with trained panel and with consumers (as an alternative to Central Location Tests), able to ensure some advantages of the sensory laboratory in terms of control of the sources of variability and some advantages of home tests in terms of lower costs and time needed for the assessors. Even if the approach was found to provide quite similar results also in the case of samples that require preparation, we observed a slight decrease in terms of discriminant ability together with a larger commitment of the assessors in those cases. Therefore, remote testing seems less appropriate, in terms of costs/benefits, when the samples require careful preparations or in cases of specific conditions of evaluations (e.g. controlled serving/storage temperature).

The results of the present work call for further studies aimed at exploring more in-depth factors affecting the sensitivity of sensory tests performed in remote conditions. The number of assessors involved in discriminant and descriptive tests appears a critical aspect especially when samples with small perceptual differences are considered. Studies specifically addressed to this topic would greatly enhance the possibility of remote testing applications. Moreover, the guidelines proposed in the present work could be adapted to cover other phases than evaluations. Among others, exploring the possibility of developing procedures for panel training in remote conditions appears a further promising perspective.

7. Conclusions

Remote sensory testing was found to be appropriate for studies with trained panellists but can also be useful with studies with consumers, when there is a need for a control of the testing conditions and a conventional home use test (in which the product is evaluated in natural conditions) is not optimal. Furthermore, consumer testing in remote conditions can help in overcoming logistic limitations when data from different regions or different countries need to be collected (i.e. cross cultural comparison). Sample characteristics limit the possibility of remote testing to products that can be handled and shipped without any hazard for participant safety and that in general show relatively high stability (physico-chemical, microbiological and sensory). Finally, it is worth noting that remote testing is time saving for participants in sensory evaluation (no time needed for travelling to sensory lab facilities) and more flexible than lab testing, and thus may facilitate participant recruitment, availability, and motivation.

Funding

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

CRedit authorship contribution statement

Caterina Dinnella: Writing – original draft, Conceptualization, Formal analysis, Methodology. **Lapo Pierguidi:** Writing – original draft, Investigation, Formal analysis, Visualization. **Sara Spinelli:** Writing – original draft, Conceptualization, Methodology. **Monica Borgogno:** Writing – review & editing, Investigation, Formal analysis, Visualization. **Tullia Gallina Toschi:** Writing – review & editing, Investigation. **Stefano Predieri:** Writing – review & editing, Investigation, Formal analysis. **Giliana Lavezzi:** Writing – review & editing, Investigation,

Table 5

Common conditions occurring in sensory laboratory, remote testing at assessor home (RT-H) or workplace (RT-W), and home use test conditions.

	Sensory Laboratory	Remote Testing (RT-H and RT-W)	Home Use Test
Control	Extremely high control of the conditions of the test (e.g., light; ambient smell, temperature)	Moderate control of the conditions of the tests (through live video-assistance)	No control of the conditions of the tests; the product is evaluated in conditions of consumption closer to reality
Time of the test	Fixed time	Fixed time	Flexibility in the choice of the time of evaluation
Assistance	Immediate assistance to the assessor in case of need	Immediate assistance to the assessor in case of need	No possible assistance to the assessor in case of need
Commitment of the assessor	Higher resources (time; energy; costs) required to the assessor to participate in the test (e.g., the assessor has to reach the facility)	Lower resources (time; energy; costs) required to the assessor to participate in the test (e.g., the assessor has not to reach the facility)	Lower resources (time; energy; costs) required to the assessor to participate in the test (e.g., the assessor has not to reach the facility)
Setting	Isolation during the test (no social setting) Interaction with other assessors in group discussion (e.g., descriptive analysis – term generation phase)	Isolation during the test (no social setting) Interaction with other assessors in group discussion (e.g., descriptive analysis – term generation phase)	No control on isolation during the test No possible interaction with other assessors, but possible interaction with family members

Formal analysis. **Francesca Trapani:** Writing – review & editing, Investigation, Formal analysis. **Matilde Tura:** Writing – review & editing, Investigation, Visualization. **Massimiliano Magli:** Writing – review & editing, Investigation, Formal analysis, Visualization. **Alessandra Bendini:** Writing – review & editing. **Erminio Monteleone:** Writing – review & editing, Conceptualization, Formal analysis, Methodology, Supervision.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodqual.2021.104437>.

References

- Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007, August). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In 7th Pangborn sensory science symposium (Vol. 16).
- Ares, G., Barreiro, C., Deliza, R., Giménez, A., & Gámbaro, A. (2010). Application of a check-all-that-apply question to the development of chocolate milk desserts. *Journal of Sensory Studies*, 25, 67–86. <https://doi.org/10.1111/j.1745-459X.2010.00290.x>
- 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. <https://doi.org/10.1016/j.foodqual.2006.06.003>
- Bruzzone, F., Ares, G., & Giménez, A. (2013). Temporal aspects of yoghurt texture perception. *International Dairy Journal*, 29(2), 124–134.
- Cardello, A. V., & Meiselman, H. L. (2018). In *Contextual Influences on Consumer Responses to Food Products* (pp. 3–54). Elsevier. <https://doi.org/10.1016/B978-0-08-101743-2.00001-7>.
- Compusense, 2021. White Papers Descriptive analysis with home panels <https://compusense.com/methodology/white-papers/>.

- Dinnella, C., Masi, C., Naes, T., & Monteleone, E. (2013). A new approach in TDS data analysis: A case study on sweetened coffee. *Food Quality and Preference*, 30(1), 33–46.
- Gazzetta Ufficiale della Repubblica Italiana (2020, march), n. 61, anno 161° <https://www.gazzettaufficiale.it/eli/gu/2020/03/09/61/sg/pdf>.
- Galmarini, M. V., Symoneaux, R., Visalli, M., Zamora, M. C., & Schlich, P. (2016). Could Time-Intensity by a trained panel be replaced with a progressive profile done by consumers? A case on chewing-gum. *Food Quality and Preference*, 48, 274–282. <https://doi.org/10.1016/j.foodqual.2015.10.006>
- ISMEA. Report sulla domanda e offerta dei prodotti alimentari 2021 Report n.4 - febbraio 2021.
- Jaeger, S. R., Chheang, S. L., Yin, J., Bava, C. M., Gimenez, A., Vidal, L., & Ares, G. (2013). Check-all-that-apply (CATA) responses elicited by consumers: Within-assessor reproducibility and stability of sensory product characterizations. *Food Quality and Preference*, 30(1), 56–67. <https://doi.org/10.1016/j.foodqual.2013.04.009>
- ISS (2020, april). Indicazioni ad interim sull'igiene degli alimenti durante l'epidemia da virus SARS-CoV-2. Versione del 19 aprile 2020. Gruppo di lavoro ISS Sanità Pubblica Veterinaria e Sicurezza Alimentare COVID-19, 2020, ii, 17 p. Rapporto ISS COVID-19 n. 17/2020.
- Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: Principles and practices*. Springer Science & Business Media.
- Li, S., Kallas, Z., Rahmani, D., & Gil, J. M. (2021). Trends in Food Preferences and Sustainable Behavior during the COVID-19 Lockdown: Evidence from Spanish Consumers. *Foods*, 10(8), 1898. <https://doi.org/10.3390/foods10081898mahieu>
- Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020). Free-comment outperformed check-all-that-apply in the sensory characterisation of wines with consumers at home. *Food Quality and Preference*, 84, Article 103937. <https://doi.org/10.1016/j.foodqual.2020.103937>
- Martin, C., Visalli, M., Lange, C., Schlich, P., & Issanchou, S. (2014). Creation of a food taste database using an in-home "taste" profile method. *Food Quality and Preference*, 36, 70–80. <https://doi.org/10.1016/j.foodqual.2014.03.005>
- Marty, L., de Lauzon-Guillain, B., Labesse, M., & Nicklaus, S. (2021). Food choice motives and the nutritional quality of diet during the COVID-19 lockdown in France. *Appetite*, 157, Article 105005. <https://doi.org/10.1016/j.appet.2020.105005>
- Meiselman, H. L. (2013). The future in sensory/consumer research:.....evolving to a better science. *Food Quality and Preference*, 27(2), 208–214. <https://doi.org/10.1016/j.foodqual.2012.03.002>
- Meyners, M., Castura, J.C. Check-all-that-apply questions. P. Varela, G. Ares (Eds.), Novel techniques in sensory characterization and consumer profiling, CRC Press, Boca Raton (2014), pp. 271-305.
- Næs, T., Brockhoff, B., & Tomic, O. (2010). Quality control of sensory profile data. *Statistics for Sensory and Consumer Science*, 11–38.
- Nogueira-Terrones, H., Tinet, C., Curt, C., Trystram, G., & Hossenlopp, J. (2006). Using the internet for descriptive sensory analysis: Formation, training and follow-up of a taste-test panel over the web. *Journal of Sensory Studies*, 21(2), 180–202. <https://doi.org/10.1111/j.1745-459X.2006.00060.x>
- O'Sullivan, M. (2017). Difference Methods. In *A Handbook for Sensory and Consumer-Driven New Product Development* (pp. 3–12). Woodhead Publishing.
- Pineau, N., Cordelle, S., & Schlich, P. (2003). Temporal dominance of sensations: A new technique to record several sensory attributes simultaneously over time. In *5th Pangborn Symposium* (p. 121).
- Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., Rogeaux, M., Etiévant, P., & Köster, E. (2009). Temporal Dominance of Sensations: Construction of the TDS curves and comparison with time–intensity. *Food Quality and Preference*, 20(6), 450–455. <https://doi.org/10.1016/j.foodqual.2009.04.005>
- Poste, L.M., Mackie, D.A., Butler, G., & Larmond, E. (1991). Laboratory methods for sensory evaluation of food. Research Branch, Agriculture Canada Publication 1864/E (p. 4). Ministry of Supply and Services.
- Seo, H.-S., Buffin, K., Singh, A., Beekman, T. L., & Jarma Arroyo, S. E. (2021). Stay safe in your vehicle: Drive-in booths can be an alternative to indoor booths for laboratory sensory testing. *Food Quality and Preference*, 94, Article 104332. <https://doi.org/10.1016/j.foodqual.2021.104332>
- Schlich, P. (2017). Temporal Dominance of Sensations (TDS): A new deal for temporal sensory analysis. *Current Opinion in Food Science*, 15, 38–42.
- Venkatesh, S., & DeJesus, J. M. (2021). Studying Children's Eating at Home: Using Synchronous Videoconference Sessions to Adapt to COVID-19 and Beyond. *Frontiers in Psychology*, 12, Article 703373. <https://doi.org/10.3389/fpsyg.2021.703373>