Food Quality and Preference 36 (2014) 111–121

Contents lists available at ScienceDirect

Food Quality and Preference

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

Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers

Gastón Ares [a,*], Lucía Antúnez [a], Ana Giménez [a], Christina M. Roigard [b], Benedicte Pineau [b], Denise C. Hunter [b], Sara R. Jaeger [b]

[a] Departamento de Ciencia y Tecnología de Alimentos, Facultad de Química, Universidad de la República, Gral. Flores 2124, CP. 11800 Montevideo, Uruguay
[b] The New Zealand Institute for Plant & Food Research Ltd., 120 Mt Albert Road, Private Bag 92169 Auckland, New Zealand

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 6 March 2014
Accepted 29 March 2014
Available online 8 April 2014

Keywords:
Consumer research
Within-assessor reproducibility
Attribute stability
CATA
Sensory properties

ABSTRACT

The purpose of the present research was further investigate the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization. Evaluations obtained when such questions are used by consumers are rarely replicated and therefore reproducibility of the data may be at risk. Results from the present work, which included five studies, each with 100–200 consumers across a range of product categories, revealed that sensory product characterizations obtained using CATA questions with consumers are highly reproducible. Hence, the research confirms previous research by Jaeger, Chheang, et al. (2013) and extends it to the use of CATA terms in randomised presentation order as has been recommended to avoid satisficing response behavior. In the future, if CATA studies are conducted without replication and researches seek to examine the reliability of CATA data, the use of *a posterior* bootstrapping re-sampling approach is suggested.

Introduction

Interest in consumer-based methodologies for sensory product characterization has steadily grown, motivated by the need to reduce the time and resources required for the implementation of descriptive analysis with trained assessors, and to directly include consumer input in the new product development process (Valentin, Chollet, Lelièvre, & Abdi, 2012; Varela & Ares, 2012). Several new methodologies have been developed, including check-all-that-apply (CATA) questions. A CATA question consists of a list of words or phrases from which the respondent has to select all the options they consider applicable/appropriate (Meyners & Castura, 2014). Despite their recent introduction to sensory and consumer science, CATA questions have been used for sensory characterization of a wide range of products, including beer, bread, cheese, chocolate, crackers, cosmetic products, dips, flavoured water, fruits, fruit drinks, fruit-flavoured drinks and sodas, ice-cream, milk desserts, potato chips, snacks, and whole grain breads (Adams, Williams, Lancaster, & Foley, 2007; Ares & Jaeger, 2013; Ares, Varela, Rado, & Gimenez, 2011; Ares, Jaeger, et al., 2013; Dooley, Lee, & Meullenet, 2010; Jaeger, Chheang, et al., 2013; Jaeger, Giacalone, et al., 2013; Meyners, Castura, & Carr, 2013; Parente, Manzoni, & Ares, 2011; Plaehn, 2012). This question format has been reported to be easy for participants (e.g., Ares, Jaeger, et al., 2013; Jaeger & Ares, 2014; Jaeger, Chheang, et al., 2013), while providing similar information to that obtained through descriptive analysis with trained assessors (Ares, Barreiro, Deliza, Giménez, & Gámbaro, 2010; Bruzzone, Ares, & Giménez, 2012; Dooley et al., 2010).

When using CATA questions for sensory characterization with consumers, replication is not usually performed due to cost, time, and resource constrains. Therefore, it is necessary to ensure that CATA questions when used by consumers provide reproducible results. Confidence that sensory characterizations using CATA questions, if implemented multiple times (for a focal set of samples and assessors) yield product descriptions that are highly similar, is a pertinent topic for research, particularly considering that applications of CATA questions (and variants hereof) will likely increase. Reproducibility of CATA questions can be evaluated using a test–retest paradigm, meaning that responses from the same group of respondents to the same set of stimuli at two discrete points in time are compared (Yu, 2005). By keeping all aspects of the empirical execution constant, differences between the two time points can be inferred to be due to the passing of time. Ideally, differences between the two time points are minimal/non-existent.

* Corresponding author. Tel.: +598 29248003; fax: +598 29241906.
E-mail address: gares@fq.edu.uy (G. Ares).

http://dx.doi.org/10.1016/j.foodqual.2014.03.010
0950-3293/

Jaeger, Chheang, et al. (2013) provided preliminary evidence of the reproducibility of CATA questions for sensory characterization with consumers. Across four studies these authors showed high test–retest reliability of consumer-based sensory product characterizations elicited with CATA questions. Product configurations and conclusions regarding similarities and differences among samples of different product categories were stable across test sessions. While representing a needed advance in methodological CATA research, a limitation of the investigations by Jaeger, Chheang, et al. (2013) was the use of fixed presentation order of CATA terms. It has been shown that the layout of the list of CATA terms significantly affect consumer responses (Ares & Jaeger, 2013; Ares, Jaeger, et al., 2013; Lee, Findlay & Meullenet, 2013) and a recommendation has been made to randomise term orders, preferably within products and participants (Ares et al., 2014). The present research does so and compares the use of fixed and randomized term orders on reproducibility of CATA product characterizations (Studies 1–3). Additionally, in Studies 4 and 5 also explored the influence on CATA product characterizations of the number of samples in the test and magnitude of differences between samples. Previous research with CATA questions have pointed to less stability in sample and term configurations when sample differences are smaller relative to larger (Ares et al., 2014). Hence, it is possible that reproducibility may also be compromised.

Methodology

Table 1 presents an overview of the studies in this research, with details of participants, product categories, number of samples and degree of sample differences, inter-session intervals, CATA terms and presentation order design.

Participants

Five consumer studies with 105–188 participants were carried out (Table 1). To increase generalizability of the findings, the consumer studies were conducted in two locations: Auckland (New Zealand) (Studies 1–3 and 5) and Montevideo (Uruguay) (Study 4). Participants in Studies 1–3 and 5 were registered on a database maintained by a professional recruitment firm and were screened in accordance with eligibility criteria for each of the studies. Participants in Study 3 also took part in Study 5. Participants in Study 4 were recruited from the consumer database of the Food Science and Technology Department of Universidad de la República (Uruguay), based on their consumption of the focal product.

Participants were aged 18–67 years old and the percentage of female participants ranged from 50% to 70%. Participants represented households in diverse socio-economic strata. Consumer samples were not necessarily representative of the general New Zealand or Uruguayan populations.

In all the studies participants gave voluntary consent to participate and were compensated for their participation. In addition, participants in Study 5 consented in writing to supplying a DNA sample (via buccal swab) for determination of their genotypes for rs6591536, which Jaeger, McRae, et al. (2013) have identified as the causal determinant of sensitivity to the odor of β-ionone.

Samples

To increase generalizability of the findings, four product categories were investigated. The samples in Study 1 were Cadbury® chocolate (Dairy Milk, Caramello, Coconut Rough and Crunchie), while in Studies 2 and 3 flavored crackers from Arnott's® were used (Shapes Barbecue, Shapes Chicken Crimpy, Roadies Sea Salt & Vinegar, Sensations Honey Soy Chicken, and Wholegrain Vita-Wheat).

In Study 4, four sample sets with the sample product category (orange-flavoured drinks) were used, comprising different number of samples (6 or 9) and different degree of difference among samples ('different' samples and 'similar' samples). Samples corresponded to commercial products commercialized in Uruguayan market and were selected on the basis of previous research (Ares, de Saldamando, et al., 2013; Ares, Jaeger, et al., 2013).

Samples in Study 5 were aqueous solutions of β-ionone (Sigma Aldrich; analytical grade; 96% purity; CAS# 79-77-6) at two concentrations (148 and 732 ppb). These concentrations were selected on the basis of Jaeger, McRae, et al. (2013) and pilot work with Plant & Food Research staff with the expectation that they would be perceived differently by participants who were more/less sensitive to the odor of β-ionone as defined by genotypes for rs6591536. Specifically, participants with the GG or AG genotypes, who are more sensitive to β-ionone, were expected to be able to perceive positive and negative odor qualities of β-ionone and detect

Table 1
Overview of studies, with details about participants, samples, experimental treatments, CATA terms, and inter-session interval.

Study ID	Number of consumers	Inter-session interval	Product category	Number of samples and degree of difference[a]	Number of terms	Presentation order of CATA terms at S1 and S2[b]
1a	58	1 week	Chocolate	4	16	S1 [F] and S2 [RW]
1b	54					S1 [RW] and S2 [F]
2a	95	30 min	Flavoured crackers	5	16	S1 [RA-O1] and S2 [RA-O1]
2b	93					S1 [RA-O1] and S2 [RA-O2]
3a	91	30 min	Flavoured crackers	5	16	S1 [RA-O1] and S2 [RA-O2]
3b	97					S1 [RW-O1] and S2 [RW-O2]
4a	57	1 week	Orange-flavoured drinks	6 'different'	16	S1 [RW-O1] and S2 [RW-O2]
4b	55			9 'different'		
4c	53			6 'similar'		
4d	52			9 'similar'		
5a	93	5 min	Flavoured water	2 'more similar'	14	S1 [F] and S2 [F]
5b	95			2 'more different'		

[a] In Studies 4 and 5 'similar' signifies smaller sample differences, whereas 'different' signifies larger sample differences. In Studies 1–3, samples were 'different' but these studies did not systematically vary degree of sample difference.

[b] [F] = one fixed order of CATA terms; [RA] = each participant uses randomly generated order of CATA terms for all samples; [RW] = each participant uses different randomly generated order of CATA terms for each sample. In Studies 3b and 4, the terminology "S1 [RW-O1] S2 [RW-O2]" means that participants at S1 evaluates each sample using a different ordering of terms. At S1 each sample is also evaluated using a different ordering of terms. None of the orders at S1 are used at S2. In Study 2a, the terminology "S1 [RA-O1] S2 [RA-O1]" means that within participants the same order of CATA terms is used for all samples at S1 and S2. Conversely, in Studies 2b and 3a the terminology "S1 [RA-O1] S2 [RA-O2]" means that within participants all samples at S1 are all evaluated using the same order and all samples at S2 are also used the same order of CATA terms, but O1 differs from O2.