Food Quality and Preference 51 (2016) 65–71

Contents lists available at ScienceDirect

Food Quality and Preference

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

Short Communication

Rate-all-that-apply (RATA) with semi-trained assessors: An investigation of the method reproducibility at assessor-, attribute- and panel-level

Davide Giacalone [a,*], Pia Ingholt Hedelund [b]

[a] Department of Technology and Innovation, University of Southern Denmark, Odense, Denmark
[b] Danish Technological Institute, Århus, Denmark

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form 18 February 2016
Accepted 24 February 2016
Available online 24 February 2016

Keywords:
Rate-all-that-apply
RATA
Assessor reproducibility
Panel agreement
Attribute stability
Sensory profiling
Chocolate

ABSTRACT

Rate-all-that-apply (RATA) is a variant of check-all-that-apply (CATA) questions that allows assessor to rate the intensity of selected attributes. Compared to CATA, RATA has the potential to improve sample description and discrimination, and might be more useful when only a small number of assessors are available. Before advocating its use with confidence, investigations on the method validity and reproducibility are necessary.

Within this context, this short paper examined the reproducibility of results obtained by RATA within a test–retest paradigm, drawing on data from a case study involving sensory assessment of common defects in chocolate production. Criteria considered were within-assessors reproducibility, attribute stability, and configurational agreement between samples spaces obtained across replicated evaluations. The results showed that although within-assessors reproducibility was moderate, RATA exhibited a very good reproducibility at panel level, as indicated by the high configurational agreement between product maps obtained from individual replicates. The method showed a good reproducibility also at the level of individual attributes. Indications were obtained that the reproducibility of RATA with semi-trained subjects might be similar to that of a simple checklist, in spite of the addition of the intensity rating step.

Overall, the work presented in this short paper supports the validity of RATA as a sensory profiling tool, and suggests that its application with semi-trained assessors may be particularly advantageous for industrial applications.

1. Introduction

1.1. Rate-all-that-apply (RATA)

Check-all-that-apply (CATA) questions are a fast product profiling technique consisting in presenting assessors with a product and checklist of predefined attributes, from which the assessor is asked to select the ones he or she finds appropriate for describing the sample (Adams, Williams, Lancaster, & Foley, 2007). This method has shown several advantages in terms of reproducibility, ease of use, and rapidity which have contributed to its increasing popularity (Meyners & Castura, 2014). The lack of bias on hedonic response (with regards to concurrent elicitation of sensory and liking on the same ballot) has also contributed to its widespread adoption, particularly with regards to its application with consumer panels (Jaeger et al., 2013b).

On the other hand, CATA questions present some limitations that may limit their use in other contexts than large-scale consumer studies. The most relevant ones are that (1) CATA produces dichotomous data (1/0) which may lack sufficient power to discriminate between samples with relatively subtle sensory differences (Reinbach, Giacalone, Ribeiro, Bredie, & Frøst, 2014), thus requiring a substantially large sample size (Ares, Tárrega, Izquierdo, & Jaeger, 2014c), and (2) that the method does not encourage a deep processing of the attributes on the ballot and may therefore prompt "satisficing" strategies in the assessors (Meyners & Castura, 2014).

To obviate these shortcomings, some authors have proposed a rating-based variant of CATA (Reinbach et al., 2014), in which assessors are required to evaluate the intensity of every applicable attribute, an approach referred to as "rate-all-that-apply" or RATA (Ares et al., 2014b). At an overall level, comparisons of RATA and CATA have shown that the two methods provide similar information (Ares et al., 2014c; Reinbach et al., 2014), but also some evidence for a greater discriminative capacity of the RATA format (Ares et al., 2014c).

* Corresponding author at: Campusvej 55, DK-5230 Odense M, Denmark.
E-mail address: dg@iti.sdu.dk (D. Giacalone).

http://dx.doi.org/10.1016/j.foodqual.2016.02.017
0950-3293/

To date, RATA has received some attention mostly with regards to comparisons with its CATA counterpart in applications with consumers (Ares et al., 2014b; Jaeger & Ares, 2015; Reinbach et al., 2014). However, little is known about other methodological aspects of RATA such as its method reproducibility. To address this paucity of information, this paper seeks to investigate the reproducibility of the results obtained across replicated evaluations using the RATA method.

1.2. Aims of the present research

The larger context of this research was a collaboration with a confectionary company, which was interested in using a fast sensory method in their quality control (QC) of their chocolate production. Several comprehensive approaches for the use of sensory methods in QC have been proposed (Civille, Carr, & Munõz, 1992), but they generally require substantial resources, and the necessary number of qualified employees is not always available. Quicker methods that can be employed with a few assessors, possibly using the company own employees, would therefore be advantageous (Costell, 2002).

Among the several fast sensory methods proposed in the literature (see e.g., Varela & Ares, 2012), the RATA method has the potential to deliver on this need. There are several reasons why this is the case. First, since the method is based on a pre-defined list of attributes, the ballot can be tailored to describing key production errors, which is important in QC applications (Costell, 2002). Moreover, in addition to the advantages it shares with the CATA format (speed, cognitive ease), RATA is known to increase processing of the ballot and the number of attributes used by the assessors (Ares et al., 2014b). Finally, the rating step increases its discriminating power which is important when working with small panels, and/or when sensory differences between the test samples are subtle. The latter is often the case in QC where a defect sample might share several attributes with the reference product, but differ with regards to their intensity.

However, before advocating its adoption with confidence, it is important to ascertain whether this method produces valid results.

In this research, we focus on reliability, i.e. the degree to which a method produces results that are stable and reproducible across repeated evaluations. Reliability is a relevant facet of validity to focus on in the present context, because in daily QC applications replication may not be an option due to practical constraints.

Method reliability can be investigated within a test–retest paradigm, that is, by obtaining responses from the same group of assessors at different time points, such as when replicate evaluations are conducted on the same sample set. This has been done in similar research on CATA questions, both with trained assessors (Campo, Do, Ferreira, & Valentin, 2008) and consumers (Jaeger et al., 2013; Ares et al., 2014a).

This work extends such methodological investigation to the RATA approach. Specifically, the purpose of this paper is to investigate the reproducibility of the results from a RATA task collected using a semi-trained panel performing replicated evaluations. Three main evaluative criteria will be addressed:

(1) assessor reproducibility: the degree to which individual (semi-trained) assessors use a RATA ballot similarly across replicate evaluations;
(2) attribute stability: at the panel level, the degree to which individual attributes are used identically across replicated evaluations;
(3) reproducibility of global sensory characterizations: at the panel level, the degree to which sample spaces obtained in replicated evaluations are congruent, and whether they lead to (dis)similar sensory conclusions about the samples.

2. Materials and methods

2.1. Samples

A total of eight chocolate samples were produced in a small production site and represented standard recipes (i.e., chocolate without production errors), and a range of critical production errors, well known in the chocolate industry, that are not easily detectable by instrumental or microbiological methods. Table 1 provides a list and brief description of the samples.

Sample B provided the basis for all the chocolate defects (samples D–H) which represented production errors relative to this standard. Samples A and C are standard recipes and were introduced to increase the complexity of the task for the assessors and to cover a wider and more multidimensional product space. As an additional diagnostics measure, sample B was also used as a blind duplicate in all sensory evaluations, meaning that the assessors evaluated nine chocolate in total. Samples are referred to by the product codes indicated in Table 1 in the remainder of the paper.

2.2. Panelists

An in-house panel composed of 16 company employees was initially recruited on the basis of interest and availability. All assessors had considerable technical expertise about chocolate production, but little or no prior experience with formal sensory evaluation.

The panel underwent four training sessions (35–45 min each), during which the assessors were introduced to sensory science, received instructions and reference materials for the RATA ballot (see Section 2.3), and were exposed to the focal samples. Additionally, they were screened for sensory acuity in relation to the basic tastes. The screening procedure was conducted in accordance with ISO 8586-1 (1993) and included both a recognition test and a threshold test. Five assessors were excluded from the panel during the training phase: three of them because they could not complete the training, and two of them because they performed below expectations during the screening procedure.[1]

The final panel was therefore composed of 11 assessors (7 women, 21–51 years of age).

2.3. Vocabulary development and RATA ballot

A trained external panel ($N = 6$) from the University of Copenhagen, Denmark was used to develop an initial list of attributes. To this end, the trained assessors evaluated all the samples in four consecutive sessions and wrote down all attributes they could think of for describing the samples considering all relevant sensory modalities. The sensory attributes most frequently mentioned, and for which a reference or a common definition could be identified, were chosen for inclusion in the RATA ballot. The final list comprised 65 attributes. For brevity, these are not discussed but interested readers will find them reported in Table 3.

The RATA ballot listed the attributes broken down by sensory modalities, because this format of presentation has been reported to improve attribute processing and reduce cognitive burden in similar tasks (Ares & Jaeger, 2013). Within sensory modalities, attributes appeared in a fixed order. To further ease attribute processing, the order in which the modalities appeared was in line with the expected 'dynamics of sensory perception' (Ares &

[1] At the request of the company, even those two assessors who did not meet the minimum performance according to ISO 8586-1 were allowed to take part in the study. However, their results were omitted from the analysis.