Food Quality and Preference 25 (2012) 140–147

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

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

Comparison of rating, best–worst scaling, and adolescents' real choices of snacks

Line Holler Mielby, Merete Edelenbos, Anette Kistrup Thybo *

Department of Food Science, Faculty of Science and Technology, Aarhus University, Kirstinebjergvej 10, DK-5792 Aarslev, Denmark

ARTICLE INFO

Article history:
Received 30 June 2011
Received in revised form 25 October 2011
Accepted 15 February 2012
Available online 23 February 2012

Keywords:
Adolescents
Best–worst scaling
Rating
Real choices
Snacks

ABSTRACT

Despite great power on the consumer market, adolescents are an overlooked segment in sensory and consumer science. This segments' ability to master different sensory and consumer tests have not been studied. This study aimed to investigate the efficacy of adolescents' use of best–worst scaling and rating of 21 different snacks by investigating these methods' predictability of adolescents' real choice of snacks. 387 Danish adolescents (11–16 years old) participated. Rating and best–worst scaling were both able to predict real choice of snack on an individual level but rating performed best. However, best–worst scaling showed greater sample discrimination. With regards to easiness, the adolescents found rating the easiest to perform. The adolescents' real choice of snacks and background data showed that boys reported a high level of hunger and chose baked savory and sweet snacks, whereas girls chose fruit snacks. Best–worst scaling was found more time consuming to work with, both during the experiment as well as during data handling before data analysis.

1. Introduction

Acceptance, preference and sensory properties of foods are some of the most important criteria for determining food choice. The evaluation techniques used to measure these are of major importance for the results retrieved. However, it is still a matter of opinion which of the existing testing methods is the best and this seems to be dependent on factors such as sample size and product (Hein, Jaeger, Tom Carr, & Delahunty, 2008; Jaeger & Cardello, 2009; Jaeger, Jørgensen, Aaslyng, & Bredie, 2008). As for the evaluation techniques applied the output of sensory and consumer science tests are also largely determined by the subjects used. Adolescents have a huge impact on food choices and are a large market in the western world, both through their parents and through their own spending power (O'Dougherty, Story, & Stang, 2005; Popper & Kroll, 2005; Tufte, 2007). They have more power over their own diet than ever before and are confronted with more choices (Popper & Kroll, 2005). Despite this, hardly any studies have been published on this age group. Most studies are performed using adults and many studies have also been published using children up to the age of 12 years old. These studies show that as children's cognitive skills develop they are able to perform and execute more advanced evaluation techniques. At the age of 12 years, adolescents have been reported to be capable of using sensory techniques both discrimination tests, e.g. paired comparisons and duo-trio tests and preferences tests, e.g. preference rankings and hedonic scales (Guinard, 2000; James, Laing, & Oram, 1997; Kimmel, Sigman-Grant, & Guinard, 1994; Kroll, 1990; Léon, Couronne, Marcuz, & Köster, 1999; Zandstra & de Graaf, 2001). Within health economics adolescents have been found to differ in their ability to use different techniques and the Case 2 best–worst scaling techniques was found easier for them to use compared to standard gamble and time trade-off methods (Ratcliffe et al., 2011). As young consumers move into adolescence (11–19 years old) they have more freedom to select foods. Many young consumers seek and develop their own individual food preference behavior independence trait, particularly apparent through the number and styles of meals eaten outside home, within the school and social environments (Brown, Mcllveen, & Strugnell, 2000). Adolescents have been known to eat on the go and to frequently snack (Nu, MacLeod, & Barthelemy, 2007; Vergetaki, Linardakis, Papadaki, & Kafatos, 2011). Due to their great power on the food market it is highly relevant to include adolescents in sensory and consumer studies and to study their food choices. To do this it is important to elucidate the efficacy of different hedonic methods using adolescents as subjects.

In sensory and consumer research, it has been standard practice to use acceptance scaling and monadic ratings on category or line scales (Jaeger et al., 2008; Lawless & Heymann, 2010b). Acceptance methods measure the degree to which a product is liked/disliked and gives interval or ratio data directly as opposed to preference methods where choices are given and intervals are indirectly measured (Thurstone, 1928). There are nevertheless other ways of assessing foods hedonically. A preference testing methodology new to sensory and consumer science is best–worst scaling (Finn & Louviere, 1992). Best–worst scaling was developed by Louviere and Woodworth (1990) and the first application of best–worst scaling was published in 1992 (Finn & Louviere, 1992). It has

* Corresponding author. Tel.: +45 87158394; fax: +45 89993495.
E-mail address: anette.thybo@agrsci.dk (A.K. Thybo).

0950-3293/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2012.02.007

among others been applied in marketing (Finn & Louviere, 1992) and health economics (Flynn, Louviere, Peters, & Coast, 2007). Jaeger et al. (2008) introduced it to sensory science as an alternative method to monadic rating with the underlying assumption that data as a minimum possess interval properties. Fundamental benefits of choice experiments including best–worst scaling is that it forces subjects to make choices between competing options and is based on the well tested theory of human decision-making (McFadden, 1974; Thurstone, 1927). While three types of best–worst scaling methodologies exist Case 1 "object case" best–worst scaling is the method which has been used within sensory science (Flynn, 2010). Best–worst scaling is a discrete choice task that forces subjects to select the best and the worst option available within a subset of samples generated from a block design. These two choices can then be converted into individual scores for each sample. These are afterwards subtracted from each other resulting in a best-minus-worst score for each sample. Best–worst scaling data can be analyzed using the multinominal logit (MNL) or by simply using the best-minus-worst scores for each sample. Due to simplicity of the analysis it is standard practice to use the best minus worst scores for each sample (Hein et al., 2008; Jaeger & Cardello, 2009). This method has been shown to give the same results as the more theoretically appropriate MNL method (Finn & Louviere, 1992; Jaeger & Cardello, 2009; Jaeger et al., 2008).

When introducing methods it is important to validate the methods and to see if they produce similar results. Best–worst scaling has been tested against various methods including hedonic rating (Hein et al., 2008; Jaeger & Cardello, 2009; Jaeger et al., 2008). In all instances best–worst scaling was either just as good as or better than the other methodologies with regards to discrimination. Additionally, best–worst scaling was found easier to use than rating. However, the drawback of the method is that it was the most demanding to perform due to the amount of samples and designs tested (Hein et al., 2008; Jaeger & Cardello, 2009; Jaeger et al., 2008). Despite being demanding, best–worst scaling might perform well with adolescents compared to the commonly used rating. Preference methods have previously found application to overcome barriers such as language and with children and hence also with adolescents where cognitive development may prevent the use of direct measurement (Guinard, 2000; Léon et al., 1999; Popper & Kroll, 2005).

When evaluating which methods are the best, measurements such as discrimination between samples, the reliability or repeatability of the method and easiness in usage of tests are often used (Jaeger et al., 2008; Köster, Couronne, Léon, Lévy, & Marcelino, 2003). These measures are relevant, but do not necessarily explain anything about the real life situation in which we are ultimately interested in. Köster et al. (2003) states, that the first hedonic impressions are poor predictors of final liking and choice. They point out that the external validity is neglected or taken for granted. Wichchukit & O'Mahony (2011) states that to validate the predictive ability of such tests, subject's choices should be observed for months as performed in a study by Rosas-Nexticapa, Angulo, and O'Mahony (2005). Alternatively, subjects could be allowed to make actual choices and their actual choice should be observed. They state that this is not a substitute for the month long observation study, but a closer approximation compared to other hedonic assessments (Wichchukit & O'Mahony, 2011). When testing the efficacy of best–worst scaling and hedonic rating using adolescents it was thus found relevant to include a real choice test in the experimental setup.

Based on existing literature, we hypothesise that best–worst scaling will perform better or just as good as hedonic rating with adolescents. The purpose of this paper was to compare the suitability of the two methods hedonic rating and best–worst scaling when dealing with adolescents (age 11–16 years old) as subjects. This was investigated by means of: Discrimination between samples (objective 1), best prediction of real choice (objective 2), easiness of using the methods from adolescents' perspectives as well as experimenters (objective 3). As adolescents have been found to snack frequently and as many snacks exist on the Danish market, a variety of snacks were chosen as sample set. Thus it was also of interest to explore the differences in adolescents' real choices of snacks depending on their age, gender, hunger and urban/rural origin, and to see how this was related to best–worst scaling and hedonic rating (objective 4).

2. Methodologies

2.1. Samples

A selection of 21 snack products available on the Danish market was selected. Unwrapped snack products were used. Actual products were used for the real choice task whereas pictures of the snacks were used for the rating and the best–worst scaling task. The snacks were very diverse in character ranging from baked savory products to fresh vegetables. The 21 snack products are listed in Table 1.

These snacks were chosen based on interviews with adolescents about their perception of snack products ('What is a snack?') and from screening of available snack products in Danish supermarkets.

2.2. Subjects

Three hundred and eighty-seven adolescents (11–16 years old, in 5th, 7th and 9th grade) were recruited from 8 public schools, 4 urban schools and 4 rural schools. The urban schools were situated in the town Odense, Denmark (approx. 166,300 inhabitants) whereas the rural schools were situated in smaller towns on Funen, Denmark (<3000 inhabitants). The participating adolescents were equally distributed between girls and boys and between urban and rural schools (Table 2). There was a majority of adolescents from 7th grade (age $\approx$ 13 years old) which is also seen on the age distribution (Table 2).

2.3. Recruitment

The adolescents were recruited through a three-step procedure: (1) definition of recruiting criteria, (2) contact to the heads of the schools and/or teachers, and (3) contact to the adolescents and their parents. The adolescent and their parents had the opportunity to reject participation. Previous research has used similar recruiting procedures (Nørgaard, 2009).

2.4. Procedure

Classrooms and adjacent rooms at the schools were the test space. The use of schools as test settings was chosen for children and adolescents to feel comfortable and relaxed when participating in tests (Jensen, 1988). Testing was carried out just after lunch, in the months January and February 2010. Each testing began by an introduction of the procedure and how to use the Case 1 best–worst scaling and the rating method. One general introduction was made to make sure that the same information was given. Second, the adolescents filled in a self-administrated questionnaire including background questions and the best–worst scaling and the rating of the 21 snacks. The interviewer stayed in the classes to help if needed and made sure that the adolescents did not interact. Third, after filling in the questionnaires, the adolescents were taken outside the classroom individually to participate in a real