| Project acronym: | EDULIA |
| :--- | :--- |
| Project title: | BRINGING DOWN BARRIERS TO CHILDREN'S HEALTHY |
|  | EATING |
| Grant number: | H2020 - MSCA - ITN - 2017: Project no. 764985 |
| Coordinator: | NOFIMA, Norway |
| Website: | www.edulia.eu |

## Deliverable D3.1:

Submitted paper: Methods for studying children's sensory and hedonic perception of food products

| Authors: | Martina Galler, Tormod Næs, Valérie L. Almli, Paula Varela |
| :--- | :--- |
| WP/WP-leader: | WP3 / Erminio Monteleone |
| Task/Task leader: | Task 3.1 / Tormod Næs, Nofima |
| Dissemination level: | PU |
| Deliverable type: | PDE |
| Approval Task/WP: | 15 May 2020 |

Approval Supervisory board: 25 May 2020

Submission date:
29 June 2020

## Table of contents

Manuscript Food Quality and Preference - Food Quality and Preference - How children
approach a CATA test influences the outcome. Insights on ticking styles from two case
studies with 6-9-year old children.................................................................................. 3

## Food Quality and Preference

How children approach a CATA test influences the outcome. Insights on ticking styles from two case studies with 6-9-year old children. --Manuscript Draft--

| Manuscript Number: |  |
| :--- | :--- |
| Article Type: | VSI: Sensometrics 2020:Research Paper |
| Section/Category: | Check-all-that-apply; Children; Ticking style; Rapid methods; Sensory description; <br> Protocol recommendations |
| Keywords: | Martina Galler, MSc <br> Nofima AS <br> Ås, NORWAY |
| Corresponding Author: | Martina Galler, MSc |
| Order of Authors: | Martina Galler, MSc |
| Tormod Næs, phd |  |
| Valérie L. Almli, phd, prof |  |
| Paula Varela, phd, prof |  |
| Suggested Reviewers: | Due to its simplicity, Check-all-that-apply (CATA) is a promising method for consumer <br> studies with children to generate sensory and other descriptions of samples, and to <br> find their drivers of liking. This paper explores how children's approach to the CATA <br> test influences the outcome , based on two case studies that illustrate suitable setups <br> for CATA tests with children of the age group 6-9. The first, conducted with <br> experimenter assistance, and the second designed to ensure the autonomy of the <br> children during the test . The children's approach to the CATA task was described with <br> ticking style indicators which revealed three ticking style groups. One group ticked only <br> a few attributes probably due to cognitive limitations, e.g. lack of reading skills, limited <br> vocabulary or ability to focus on the task. The second group gradually increased their <br> number of ticked attributes per sample over the test, while the third subgroup ticked a <br> steady number of attributes throughout the test. The two latter groups are likely to <br> represent different test strategies: one using the CATA list relatively to the sample <br> space, and one using the CATA list as in a more absolute way. Analysis regarding <br> data validity assessed by the detection of pre-defined Design of Experiment (DoE) <br> sample differences and the alignment to a trained panel using Quantitative Descriptive <br> Analysis (QDA) revealed that ticking style played a crucial role. This study shows the <br> importance of analysing "ticking style" as a validation strategy for CATA tests run with <br> children and as a tool to gain insights into underlying test strategies. |
| Monica Laureati |  |
| Monica.Laureati@unimi.it |  |
| experienced with sensory testing with children |  |$|$| Rosires Deliza, phd |
| :--- | :--- |
| rosires.deliza@embrapa.br |
| she has experience with CATA testing with children |
| Begoña Alfaro, phd |
| Senior researcher. New Foods |
| balfaro@azti.es |

## Dear Editor

Please find attached the manuscript entitled "How children approach a CATA test influences the outcome. Insights on ticking styles from two case studies with 6-9-year old children." for your consideration. Authors are Martina Galler, Tormod Næs, Valérie Legard and Paula Varela.

This research aims at understanding how children's approach to the Check-all-that-apply (CATA) test influences their results. We defined ticking style indicators to describe children's CATA usage and validated their data based on ticking style. Based on practical experiences from the case studies and based on our data analysis we provide advices how CATA test protocols can be adapted in a child-friendly way.

I hope you find it worth considering for publication.
With kind regards,

## Martina Galler

Phd candidate
Nofima, Ås Norway
Tel: +47 47717616
https://edulia.eu/

## Highlights

- Ticking style was analysed in CATA tests with 6- to 9-y.o. children
- Data validation based on sample discrimination and alignment to trained panel was carried out
- Ticking style indicators unveiled groups with different test strategies
- Differences among groups may reflect different cognitive development levels
- Recommendations for suitable child-friendly adaptations of CATA protocols are provided


## How children approach a CATA test influences the outcome. Insights on ticking styles from two case studies with 6-9year old children.

Martina Galler ${ }^{12}$, Tormod $\mathrm{Næs}^{1}$, Valérie L. Almli ${ }^{12}$, Paula Varela ${ }^{12}$
${ }^{1}$ Sensory \& Consumer Sciences, Nofima As, Norway
${ }^{2}$ The Norwegian University of Life Science, Department of Chemistry, Biotechnology and Food Science (KBM), Ås, Norway

Corresponding author: Martina Galler,
Tel: +4747717616

Email: martina.galler@nofima.no


#### Abstract

Due to its simplicity, Check-all-that-apply (CATA) is a promising method for consumer studies with children to generate sensory and other descriptions of samples, and to find their drivers of liking. This paper explores how children's approach to the CATA test influences the outcome, based on two case studies that illustrate suitable setups for CATA tests with children of the age group 6-9. The first, conducted with experimenter assistance, and the second designed to ensure the autonomy of the children during the test. The children's approach to


the CATA task was described with ticking style indicators which revealed three ticking style groups. One group ticked only a few attributes probably due to cognitive limitations, e.g. lack of reading skills, limited vocabulary or ability to focus on the task. The second group gradually increased their number of ticked attributes per sample over the test, while the third subgroup ticked a steady number of attributes throughout the test. The two latter groups are likely to represent different test strategies: one using the CATA list relatively to the sample space, and one using the CATA list as in a more absolute way. Analysis regarding data validity assessed by the detection of pre-defined Design of Experiment (DoE) sample differences and the alignment to a trained panel using Quantitative Descriptive Analysis (QDA) revealed that ticking style played a crucial role. This study shows the importance of analysing "ticking style" as a validation strategy for CATA tests run with children and as a tool to gain insights into underlying test strategies.

Keywords: Check-all-that-apply, Children, Ticking style, Rapid methods, Sensory description, Protocol recommendations

1. Introduction

Rapid sensory methods such as Check-all-that-apply (CATA) and Projective Mapping are now used in a broad range of applications, both in research and industry (Delarue, Lawlor, \& Rogeaux, 2015; Varela \& Ares, 2012). These methods can produce similar results as traditional descriptive methods with the advantage that they are more flexible and less time consuming. In their review, Varela \& Ares (2012) describe how the emergence of rapid methods has blurred the line between sensory and consumer studies. Rapid methods have been validated both in studies with trained panellists (Dehlholm, Brockhoff, Meinert, Aaslyng, \& Bredie, 2012) and with consumers (Ares, Barreiro, Deliza, Giménez, \& Gámbaro, 2010;

Bruzzone, Ares, \& Giménez, 2012; Dooley, Lee, \& Meullenet, 2010; Jaeger et al., 2013). As validation, they mainly used the comparison to results generated with traditional descriptive methods. Jaeger et al. (2013) evaluated the within-assessor reproducibility of several CATA datasets with repetitions generated by consumers.

Many rapid methods are simple to perform and therefore promising to use in consumer studies with special populations such as children. In recent years, various applications of rapid methods with children have been published. Daltoe et al. (2017) used projective mapping with food stickers to understand the perception of fish of different age groups. Varela and Salvador (2014) concluded that children from the age of five years old could perform a structured sorting task with images. The most common rapid method used with children has, however, been the CATA method. Researchers used the CATA method with sensory attributes (Cardinal, Zamora, Chambers, Carbonell Barrachina, \& Hough, 2015; Laureati et al., 2017; Lima, Ares, \& Deliza, 2018; Schouteten, De Steur, Lagast, De Pelsmaeker, \& Gellynck, 2017), emotional attributes (De Pelsmaeker, Schouteten, \& Gellynck, 2013; Schouteten et al., 2017; Schouteten, Verwaeren, Gellynck, \& Almli, 2019; Schouteten, Verwaeren, Lagast, Gellynck, \& De Steur, 2018) and hedonic attributes (Yoo et al., 2017) to investigate children's perception and their drivers of liking. Table 1 provides an overview of previous CATA studies with children generating sensory descriptions.

In their review about sensory testing with children, Laureati, Pagliarini, Toschi, and Monteleone (2015) highlighted the importance of adapting test protocols to the cognitive level of the targeted age group to ensure that the results reflect the actual perception, not the cognitive limitations of understanding the task. One such limitation could be difficulties to understand the words of the CATA list. To avoid this potential issue, Laureati et al. (2017) and Schouteten et al. (2017) generated a CATA list with a panel of children while Lima et al. (2018) did a pilot study to test if the children understood the CATA list.

How to evaluate the suitability of a test protocol for the respective age group regarding the validity of results is still a rather unexplored area. Schouteten et al. (2017) showed that children were able to discriminate samples with the CATA method. Laureati et al. (2017) and Lima et al. (2018) could further show that pre-defined sample differences were detected. Cardinal et al. (2015) and Lima et al. (2018) compared children's discrimination capability to adults. To the authors' knowledge, none has compared sensory profiling by children to a trained panel which is still the "golden standard" regarding the objectivity of sensory descriptive results. In their recent book, Næs, Varela, and Berget (2018) suggested the analysis of ticking style to understand how consumers use the CATA list which could potentially be used to study how children approach the test.

The objective of this paper is to explore the analysis of ticking style as a way of validating CATA testing with 6-9-year-old children. We investigate children's ticking style in two case studies, one on bread and the other on fruit smoothies. Further, based on the practical experiences and data analysis findings in each of the studies, we draw practical recommendations for conducting CATA tests with children.

## 2. Materials \& Methods

The two case studies, Bread and Smoothie, illustrate how a CATA test with children of the age group 6-9 can be set up, the first (Bread) conducted with experimenter assistance and the second (Smoothie) designed to ensure the autonomy of the children during the test. We defined three ticking style indicators to describe and group the children based on their usage of the CATA list: number of ticks, standard deviation of the number of ticks per sample, and number of different attributes used in the test. Then we analysed data validity regarding detection of sample differences based on the Design of Experiment (DoE) and regarding similarity to the sensory description of a trained profile panel by Quantitative Descriptive Analysis (QDA).

### 2.1. Samples

Bread and smoothie samples were constructed to vary systematically in their sensory profiles based on a $2^{3}$ factorial design, resulting in 8 different samples. Each factor covered a different sensory modality (Darkness, Coarseness and Saltiness for Bread; Colour intensity, Thickness and Acidity for Smoothie; Table 2). The bread samples were baked at the cereals pilot plant at Nofima, based on a non-commercial recipe. The smoothie samples were prepared in lab scale by a commercial partner, using one of their commercial smoothies as a base. The base smoothie contained $100 \%$ fruit juice of raspberry, blueberry, strawberry, banana, apple and orange and naturally displayed a red colour. Figure 1 shows the visual differences between the bread samples.

### 2.2. Consumer test with 6 to 9 -year-old children

Three school grades from local schools in the Akershus county (Norway) participated in the consumer tests. Both studies were run in the respective schools and each school participated in one study only. The majority of the children were between 7 and 9 -years-old. However, as the school grade is based on the year of birth in Norway, some 6-year-old children participated in the test as well. Parental informed consent forms, including allergy information, were collected before the tests. Children gave their informed assent to participate and were informed they could leave the test at any point. The data collection followed the ethical recommendations from the Norwegian Centre for Research Data.

### 2.2.1. Bread test

The check-all-that-apply (CATA) list was established by researchers based on the main sample properties as described by a trained panel. They defined ten attributes (Light colour, Dark colour, Not grainy, Grainy, Easy to chew, Hard to chew, Not coarse, Coarse, No salty taste, Salty taste). In each case, two attributes stretched the same dimension as antonyms,
e.g. "Salty taste" and "No salty taste". Prior to the test, the understanding of the CATA list was tested through a pilot study with children of the age group.

In total, 103 children participated in the test. The test questions were presented on a paper questionnaire, as shown in Figure 2. The children executed the test in subgroups of five, with three experimenters available for assistance in, for example, tasting the right sample, reading challenging words or remembering to rinse between samples. In the first page of the questionnaire, the children were asked to indicate age and gender. The eight samples were presented in a sequential monadic balanced presentation order, coded with single capital letters A-H (Figure 2). Each sample was first evaluated for overall liking on a 1 to 7 -point scale with three emojis (unhappy, neutral, happy) as anchors, followed by the Check-all-that-apply (CATA) evaluation on the same page. Attributes were randomized across children to prevent position biases but kept constant across sample evaluation as per the recommendation by Meyners and Castura (2016). Between the samples, the children were instructed to rinse their mouth with water. At the end of the test, an ideal (imaginary) sample was evaluated for liking and CATA.

Even though experimenter assistance allowed a successful data collection, there were some challenges observed during the test. The reading was challenging for some 6 - and 7 -year-olds, which reduced their attention to their taste perception itself. Also, some children monopolised experimenter support. Some children did not understand all CATA attributes, especially the attribute "coarse". Nearly a fourth of the children skipped pages, a few confused samples and some forgot to rinse their mouth with water in-between sample tasting. Finally, many explanations were necessary for the abstract question about their ideal product. In general, the experimenters noted a loss of interest after four to six samples.

### 2.2.2. Smoothie test

The Smoothie test tried to overcome some of the challenges encountered in the Bread test. The main focus was to improve the autonomy of the children during the test, particularly with regards to attribute reading and understanding. To ensure a good understanding of the CATA attribute list, children of the age group developed attributes with the repertory grid method. Twelve children established 59 attributes. The experimenters reduced their attributes based on the frequency of elicitation and synonym reduction to the following 15: Light colour, Dark colour, Bubbles, Thin, Thick, Slimy, Very sour, Banana, Lemon, Strawberry, Raspberry, Blueberry, Strong smell, Yummy, Yuck. The list included two hedonic attributes "Yummy" and "Yuck" as well as an odour attribute "Strong smell".

To address reading challenges previously observed with the 6-and 7 -year-olds ( $2^{\text {nd }}$ graders), the children read the attributes with the teachers in class and with parents when they signed the consent form before the test. The questionnaire was electronic with little text to minimize the reading effort. A monkey story was introduced in the test in order to increase engagement: the participants were asked to help the experimenters find out what type of smoothies a monkey that had broken into a smoothie factory had produced.

In total, 93 children participated in the test. The test was performed on tablets. The test pages are displayed in Figure 3. At the start of the test session, the experimenters explained and demonstrated the test. Then the children conducted the test independently. The children executed the test in subgroups of ten, with three experimenters available for assistance. The first page of the questionnaire asked school grade and gender, followed by the sample-related questions. The original smoothie with the low factor levels (no colour added, no thickener, no lemon juice added) was first evaluated as "warm-up" sample (sample 1_1, Table 2). The eight samples were then presented in sequential monadic balanced presentation order, coded with distinct symbols (e.g. a lightning, see Figure 3c and d). Each sample was first evaluated for overall liking on a 1 to 7 -point scale with seven emojis (from unhappy to happy) followed by the CATA evaluation on the next page (Figure 3). As in the Bread case study, the attributes
were randomized across children, but kept constant across samples. Between the samples, the children were instructed to rinse their mouth with water. No ideal sample was evaluated in this case.

### 2.3. Quantitative descriptive analysis with trained panel (QDA)

A Generic descriptive analysis (based on QDA as described by Lawless \& Heymann, 2010) was performed for each set of samples by the trained profile panel of Nofima. Nofima's panel is highly trained and very stable. The assessors are solely hired as tasters, and some of them have more than 30 years' experience working with descriptive analysis. Panel performance is checked for every project, based on three qualities: discrimination, repeatability and agreement. The descriptive terminology of the products was created in a pre-trial session using extreme samples, selected for showing extremes examples stretching the sensory space. After a 1-h pre-trial session, the descriptors and definitions were agreed upon by the assessors; all assessors were able to discriminate among samples, exhibited repeatability, and reached an agreement with other members of the group. For the bread samples the following 18 attributes were defined: Acidic odour, Grain odour, Cloying odour, Colour hue, Colour strength, Whiteness, Hardness, Juiciness, Coarseness, Chewing resistance, Sticky, Doughy, Acidic taste, Sweet taste, Salty taste, Bitter taste, Corn taste, Cloying taste. For the smoothie samples, the following 18 attributes were defined: Intensity smell, Acidity smell, Fruity Berry smell, Artificial smell, Colour intensity, Whiteness, Taste intensity, Acidity, Sweetness, Sourness, Bitterness, Metallic, Fruit Berry, Artificial, Fullness, Viscosity, Astringency, Pungency. After a pre-testing, nine panellists rated each sample in duplicate on a $10-\mathrm{cm}$ scale.

### 2.4. Statistical analysis

### 2.4.1. Usage of CATA list, ticking style indicators

To find out how the children used the CATA attributes, we described their ticking behaviour with three ticking style indicators. The total number of ticks for the eight randomized samples
(called "number") and the standard deviation in the number of ticks per sample (called "std") were calculated as described by Næs, Varela, and Berget (2018). As a third indicator, we considered the number of different CATA attributes (called "attributes") used per child as well. "Attributes" was regarded as relevant to compare the usage of a researcher-developed CATA list in the Bread test and a child-generated CATA-list in the Smoothie test.

Three equally sized ticking style groups were built based on the first two components of the PCA of the standardized ticking style indicators. The ticking style groups were compared regarding age in the Bread dataset and school grade in the Smoothie dataset with a $\mathrm{X}^{2}$-test.

### 2.4.2. Analysis of CATA data

The Cochran's $Q$ test was used to test for differences between samples regarding the number of ticks of a CATA attribute. The ticking or no ticking of an attribute was defined as the binary response variable, sample as a fixed factor and child as a random factor.

A correspondence analysis (CA) of the contingency table of the CATA attributes was performed. The not significant attributes were included for better comparability of the ticking style groups where the significance was not conclusive due to their smaller sample size. For better interpretation and comparability of the score plots, the levels of the three design of experiment (DoE) factors were projected as supplementary qualitative variables into the plot. The "Ideal" sample in the Bread study and the "Warm-up" sample "1_1" in the Smoothie study were projected as supplementary rows into the score plot. The projection of the supplementary variables was done with the FactoMineR R package according to Lê, Josse, and Husson (2008). The supplementary variables did not influence the configuration.

To compare the perceptual space of the three ticking style groups, a multiple factor analysis (MFA) was performed using the contingency tables of each ticking style group defined as a frequency table. For better readability, the plot only displayed the DoE factor levels of the
overall configuration as well as the partial coordinates of the ticking style groups. Again, the DoE factor levels were projected into the plot as supplementary qualitative variables.

### 2.4.3. Analysis of QDA data

The significance of the QDA attributes regarding sample discrimination was determined with a Mixed effect ANOVA. The rating on a scale (1 to 10 ) of the attributes was defined as the continuous response variable, samples as a fixed factor and trained assessors as well as the assessor x sample interaction were considered as random factors.

A principal component analysis (PCA) was performed with the significant unstandardized QDA attributes. The levels of the three design of experiment (DoE) factors were projected as supplementary qualitative variables into the score plots.

### 2.4.4. Liking

The influence of the DoE sample differences on the liking rating were analysed with a Mixed ANOVA, with the DoE factors and second order interactions as fixed and child as well as second order interaction of child $\times$ DoE factors as random.

The correlation of the average liking of the samples with the first three components of the perceptual space of the children (CA and MFA) and trained profile panel (PCA) was calculated and displayed in correlation circles.

### 2.4.5. Similarity Index

The similarity between the perceptual space of the children and the trained profile panel was measured with the similarity index (SMI) introduced by Indahl, Næs, and Liland (2018); the first component, the first two components, as well as the first three components of the score plots, were compared. The SMI was chosen over the more frequently used RV coefficient because it weighs the three components more equally while the RV coefficient weighs the first
component most. It must be noted that the SMI, as well as the RV coefficient, overestimate the similarity of the present matrices because the row versus column ratio was relatively small in the score plot matrices.
2.4.6. Investigation in underlying reasons for ticking style

In order to further analyse ticking behaviour, we analysed the influence of three different variables on the ticking number per sample. It was of interest if the number was linked to certain samples, the hedonic response to them or tasting order. The liking ratings were transformed to ranks within child to avoid scale effects, the sample with the lowest rating was assigned the lowest rank, 1 and the sample with the highest rating was assigned the highest rank, 8. A mixed regression model then analysed sample, ranked liking and tasting position as fixed effect and child as random effect.

### 2.4.7. Software

$R$, version 3.5.1 was used for the data analysis. The package FactoMineR for CA, MFA and PCA, the ImerTest and mixIm as packages for Mixed effect models, the RVAideMemoire package for Cochran's $Q$ test and the MatrixCorrelation package for SMI calculation.

## 3. Results

### 3.1. Usage of the CATA list

The bread questionnaire was paper-based and assisted by researchers. Several children's evaluations contained missing answers. These incomplete datasets, 26 in total, were excluded from this data analysis. The remaining 83 children used the CATA list in different ways. Figure 4 presents a summary of ticking style indicators for the Bread and Smoothie studies. The distribution of the ticking style indicators is displayed as a histogram, the lower plots show the correlation between the variables, and the upper plots display their Pearson correlation values. One child only ticked twice during the whole test while the most active child ticked 33 times
(see ticking style indicator: "number" in Figure 4). Some children used one of the ten available attributes across all samples, while others used up to eight different attributes across all samples (ticking style indicator: "attributes" in Figure 4). None of the children used all ten available attributes. Some children displayed a high standard deviation in the number of ticks per sample (ticking style indicator: "std" in Figure 4) indicating an unsteady ticking behaviour. In contrast, others ticked a similar number of attributes for all samples.

The electronic questionnaire of the Smoothie test required the evaluation of all samples, ticking at least one CATA attribute per sample. Therefore, no answers were missing, and all 93 answers could be considered for the analysis. The minimal number of ticks was eight, corresponding to one tick per sample. In this test, some children used all 15 available CATA attributes across all samples ("attribute") which indicates that the child-developed attributes were well applicable. The analysis of the ticking style revealed one outlier displaying an extremely high standard deviation. The inspection of this child's ticking data showed that he/she had ticked almost all attributes for half of the samples while for the other half, he/she had only made one tick per sample which was required by the electronic questionnaire in order to continue. It can be assumed that this child did not use the CATA list to describe his/her perception of the samples and his/her data were excluded from further data analysis.

Next, it was of interest how the different ticking styles influenced the perceptual space generated by the children. A PCA of the three ticking style indicators, "number", "attributes" and "std", indicated a tendency for a split in three groups of children in both datasets (Figure 5). One group consisted of children that only used a few attributes of the CATA list, the "few tickers". Another group consisted of children that ticked frequently displaying a high standard deviation, the "unsteady tickers". The third group represented the children that ticked frequently displaying a low standard deviation, the "steady tickers". Three subgroups of the data set with almost the same size were established based on this interpretation.

The "few tickers" ratio decreased with age, as displayed in Figure 6, indicating that this ticking style might be related to cognitive limitations, e.g. difficulties to read and understand the CATA attributes. However, no significant difference between the age groups in either of the datasets was found with the $\mathrm{x}^{2}$-test, ( $p$-value: 0.428 in Bread, 0.476 in Smoothie).

### 3.2. Check-all-that-apply and liking of children

Tables 3 (Bread) and 5 (Smoothie) show the number of ticks in total and the significance of each CATA attribute for the total panel as well as for the ticking style groups. It was of interest if the children discriminated the samples with CATA attributes representing the three DoE differences between the samples. Table 4 (Bread) and 6 (Smoothie) show the influence of the DoE differences on liking. It was of special interest if the children could describe their drivers of liking with CATA attributes.

In the Bread case study (Table 3), the two attributes "Light colour" and "Dark colour" representing the DoE factor Darkness were significant for all ticking style groups. Coarseness was represented by the three antonym pairs "Grainy", "Not grainy", "Coarse", "Not coarse" as well as "Easy to chew", "Hard to chew". One or both antonyms representing grainy and coarse were significant in each ticking style group. Only the "unsteady tickers" differentiated the samples regarding the chewing aspect. The overall ticking number suggests that all samples were perceived as "Easy to chew" which was ticked 405 times while "Hard to chew" was only ticked 95 times. So, the "unsteady tickers" were the only group that described the relative difference between the samples. For the DoE factor Salt one of the two antonyms, "Salty taste", was significant. Conclusive analyses of the ticking style groups regarding discrimination are not possible due to the small group sizes of the ticking style groups. However, p-values suggest that the "unsteady tickers" did discriminate the samples with the attribute "Salty taste" (pvalue $=0.06$ ) while the "few tickers" clearly did not ( $p$-value $=0.56$ ) and the "steady tickers" also did not sufficiently ( $p$-value=0.18). It can be summarized that the "steady tickers" and the "few
tickers" did mainly discriminate the samples regarding two of the three DoE factors. The "unsteady tickers" discriminated most attributes representing all three DoE factors.

The liking evaluation based on the pre-defined DoE factors (Table 4) revealed different preference patterns for the ticking style groups. For the overall panel as well as for the "unsteady tickers" and "steady tickers", Salt was the main driver of liking. However, for the "few tickers" the texture aspect, Coarseness was their main driver of liking.

The Smoothie case study (Table 5) included some attributes that did not represent the DoE differences directly. Some of them were significant in the discrimination between samples, e.g. the two hedonic attributes ("Yummy" and "Yuck") and fruit flavour attributes not directly referring to the difference in Acidity ("Banana" "Strawberry", "Blueberry"). The "unsteady tickers" and "steady tickers" discriminated the samples with a high number of the CATA attributes covering all three DoE factors. The "few tickers" discriminated less and did not display any significant texture attributes that could represent the DoE difference in Thickness.

For all ticking style groups the DoE factor Acidity was their main driver of liking (Table 6). For the "few tickers", the texture aspect Thickness might have also had an influence on their liking $(p$-value $=0.054)$.

The average values of the liking rating by children are displayed in the Appendix Table A1 (Bread) and A2 (Smoothie).

### 3.3. Perceptual space and correlation of liking, comparison to trained panel

The analysis of the perceptual space allowed to check if the children discriminated the samples according to the underlying DoE factors and to evaluate the correlation of the components with the average liking. Further, it allowed the comparison with the trained profile panel. Figure 7 (Bread) and 8 (Smoothie) show a CA of the CATA contingency table, an MFA comparing the contingency tables of the ticking style groups as well as a PCA of the QDA rating by the trained
panel. The first three components of the score plots with the DoE factor levels projected for better interpretability are displayed as well as the correlation with the average liking. The definition of the sample numbers can be found in Table 2. Loading plots as well as separate CAs for each ticking style group can be found in the supplementary material (Supplementary Figures 1-10). Table 7 displays the similarity index (SMI) between the CA score plots of the children and the PCA score plot of the trained panel. The average values of the QDA of the trained panel, as well as the p-values are displayed in Tables A3 (Bread) and A4 (Smoothie).

In the Bread case study, the three DoE factors were each represented by one component of the perceptual space of the children as well as of the trained panel (Figure 7). The colour difference Darkness was represented by the first component, Coarseness by the second component and Salt by the third. The perceptual difference in Salt was relatively small compared to the other two DoE factors, although it most strongly correlated with liking. The MFA plot where the ticking style groups are compared shows that the "unsteady tickers" described the most liking-relevant difference in Salt level most. The "few tickers" differed in their preference from the other groups. For this group, the DoE factor Coarseness was more correlated with their liking. The imaginary ideal sample (Ideal) was well aligned with the liking correlation and placed outside of the sample space in the third component, which was most important for liking. The placement of the Ideal sample indicates that the children understood the concept of the imaginary ideal sample and did not rate it identically to their preferred real sample.

In the Smoothie case study (Figure 7), Acidity was most strongly correlated with liking and also represented by the first component. All ticking style groups could discriminate the samples regarding Acidity. In the second component, the thinner and lighter samples and thicker and darker samples were more often described by the same attributes, so that the DoE factors Thickness and Colour overlapped. Considering the third component, the factors Thickness and Colour were separated, however. The trained panel showed a similar perceptual space also in
this case. The association of DoE factors Thickness and Colour in component 2 was not apparent, however. The warm-up sample 1_1, which was composed of the low factor levels and identical to sample 1, was well placed in the first two components, Acidity_0 and Thickness_0, but not in the third component, Colour_int_0. The colour attributes "Light colour" and "Dark colour" only became applicable over the test once darker samples had been presented. In contrast, the attributes describing Acidity and Thickness were applicable in a more absolute way, less relative to the sample space.

In both case studies, the similarity index (SMI) between the trained profile panel and the complete child panel was high, 0.94 in the Bread dataset and 0.93 in the Smoothie dataset (Table 7). The "few tickers" were the least aligned with the trained panel over the three components in both studies while the "unsteady tickers" as well as the "steady tickers" were well aligned with the trained panel.

### 3.4. Investigation in unsteady ticking behaviour

In the presented datasets, the "unsteady tickers" produced a good sample discrimination and detection of pre-defined sample differences. We first hypothesized that the unsteady ticking behaviour was sample induced, e.g. by the intensity of the DoE factor level or by the children's hedonic responses to them.

However, the present data suggest that the tasting position of the sample played a more important role than the sample properties or the hedonic response (Table 8). The "unsteady tickers" increased their ticking number along the test. In the beginning, they ticked fewer, and in the end, they ticked more attributes in both datasets, as shown in Figure 9. The good results of the "unsteady tickers" could indicate that learning took place over the test. The attributes became relevant and more applicable once the sample space was apparent. In the Bread study, all ticking groups showed a slight increase in the number of ticks which might be linked
to antonym-based attribute structure, while in the Smoothie dataset, this trend is only observable in the unsteady ticking group.

## 4. Discussion

### 4.1. Assessment of CATA for sensory description with children and determination of

 their drivers of likingAs shown by Laureati et al. (2017) and Lima et al. (2018), children were able to discriminate samples regarding pre-defined sample design differences. The two case studies analysed in the present paper also showed for the first time that the alignment with a trained profile panel was generally very high, for the consensus perceptual space. The high alignment to the trained panel indicates that the majority of the children's usage of the CATA list was guided by their sensory perception, which they could accurately point out with the CATA list. However, our results indicate that ticking style plays an important role regarding data validity which is discussed further in the next section 4.2.

In both studies, the design factor representing the sensory modality taste was the main driver of liking. In the Smoothie study Acidity was also the predominant factor of the perceptual space. In the Bread study, Salt was the least important factor in terms of product description, only apparent in the third component. As this factor was also only visible in the third dimension of the perceptual space of the trained panel, it can be assumed that it was the least salient DoE factor difference regarding perception.

### 4.2. Implications of ticking style

The analysis of the ticking style indicators revealed some participants that could not use the CATA list accurately to describe their perception. Ticking style indicators can, therefore, be valuable to find outliers, e.g. eliminating consumers from the data analysis with a low ticking
"number" or low number of "attributes". The elimination of the "few tickers" from the data analysis might be especially relevant when the setup of an electronic questionnaire requires a minimal number of ticks, and when young children participate in the test. In the two case studies, the proportion of children in the few ticking group decreased by age in trend. Therefore, the few ticking behaviour is likely linked to cognitive limitations. In her review Anderson (1998) described how executive functions such as ability to resist distraction and verbal fluency, of which a certain degree is a pre-requirement for the successful performance of a CATA test, are only mature by the age of 12 and older and large individual differences occur.

Against the observation that the children tended to get bored over the test which could lead them to tick a smaller number of CATA attributes, a hypothesis, e.g. also mentioned by Jaeger et al. (2015) for adults, our analysis of ticking style indicators showed the opposite. While the "steady tickers" kept their ticking number constant over the test, the "unsteady tickers" increased their number of ticks over the test. This increase makes perfect sense for the relative nature of sensory evaluations, especially in the case of the CATA method where the response to a continuous stimulus has to be transformed into a binary answer. To describe a sample as "Salty" becomes more relevant once a less salty sample has been tasted. In the smoothie study, the absence of an "anchoring" effect was foreseen: A warm-up sample 1_1 where all three factor levels were low was presented prior to the eight randomized test samples. The occurrence or non-occurrence of this increased-ticking behaviour points to different underlying test strategies: The "steady tickers" might use the CATA attributes in a more absolute sense. In contrast, the "unsteady tickers" might use them in a more relative sense considering the sample space that gradually unfolds to them during the test. The "unsteady tickers" are likely to produce similar results as with a similarity-based method, such as the projective mapping, while the "steady tickers" might generate different results with this approach. More generally, these findings point to a phenomenon likely to underly many sensory consumer tests where samples are presented in a sequential monadic design. Consumers are generally instructed
to rate a sample independently of previously tasted samples. However, many consumers are likely to switch to a strategy where they contrast previously tasted samples, adjusting the scale to the sample space of the test. Lawless and Heymann (1998) described this effect as contrast effect, attributing it to an axiom of perceptual psychology: "Humans are very poor absolute measuring instruments but are very good at comparing things."

### 4.3. Implications of the test protocol

Table 9 highlights the learnings from the two case studies for future CATA test setups with children.

In the first case study with Bread where researchers developed sample-relevant CATA attributes, some CATA attributes were not understood by all children. Our data analysis showed that the sample- and age-relevant CATA list developed by children in the Smoothie case study was more fully used than the list developed by researchers, both regarding the ticking style indicator "number" which might also be related to the higher number of available attributes, but also regarding the number of different attributes used throughout the test, "attributes". Moreover, no attribute explanations were necessary during the smoothie data collection, while "coarse" generated several questions in the bread study. Regarding data validity, both the sample-relevant CATA list based on antonyms and the sample- and agerelevant CATA list were suitable. The sample-relevant CATA list produced a perceptual space that divided the samples based on one DoE factor in each component. In comparison, the less systematic sample- and age-relevant CATA list revealed an interaction between two sample design factors, Colour and Thickness which was not found in the perceptual space of the trained panel. Whether this can be attributed to the type of CATA list is not conclusive as the two case studies vary in too many aspects.

Special care should be taken setting up the questionnaire. The text throughout the test should be reduced to the minimum because reading takes more time for children. Instead of written
instructions, a live demo of the test is useful and recommended. To increase children's motivation, the Smoothie study included a story explaining the purpose of their task. This ensured that the children engaged and fulfilled the test despite its high level of repetitiveness.

Overall, the electronic questionnaire offered advantages over the paper questionnaire where children skipped pages, forgot to rinse their mouth with water between samples and needed a higher degree of assistance. An electronic questionnaire can include a page between samples as a reminder to rinse the mouth. Also, missing answers can be avoided. Another advantage is that with tablets the test looks and feels much more like a game. It has to be kept in mind, that the mandatory answers might trigger some wrong data as seen in the outlier discussed in section 3.1.

Labelling samples with symbols instead of three-digit codes or letters makes the selfadministered tasting easier. Care should be taken in the choice of suitable symbols to avoid cross-modal influences of the symbols on taste perception. Deroy and Valentin (2011) for example, showed an association of certain shapes with certain tastes. Symbols differing in emotional valence might bias hedonic ratings of samples as well.

The ideal sample in the Bread case study was well aligned with the liking and outside the real sample space for the most liking-relevant DoE difference Salt. However, at data collection stage an explanation for the evaluation of an "ideal" sample is necessary as children are likely to think in a less abstract way than adolescents and adults, corresponding to the operational development stage described by Piaget (1964).

Our data analysis revealed that the CATA attributes became more relevant for one group, the "unsteady tickers", once the sample space was apparent. This sample space-relative ticking would speak for a training session or at least an anchoring "Warm-up" sample as done in the Smoothie case study in order to improve data quality.

### 4.4. Limitations and future research

This study sheds light into the topic of individual differences in approaching a consumer profiling method with children, i.e. CATA. Results highlighted groups of children performing the test in different ways. The segments found is this study were small ( N appr. 30), so more research with larger groups of children would be desired to confirm these findings. To the authors' knowledge, no study has been done studying the ticking style with CATA data of adults. It would be desirable to do so, preferably with a DoE underlying the test design, for a more controlled interpretation. Added to this, further studies are needed on different food categories, with smaller and larger differences among samples, to see to what extent these potential ticking groups may affect the outcome of the studies.

## 5. Conclusion

This paper unveils that individual differences underly how children 6-9 y.o. approach CATA tests, influencing the outcome, with potential implications for test design, validity check and interpretation of results. We propose three ticking style indicators to study this: number of ticks, standard deviation of number of ticks per sample, and number of different attributes used in the test. Our analysis revealed one group, the "few tickers", that used the CATA list scarcely and produced less informative data, potentially due to cognitive limitations. The other two groups produced valid data, closer to QDA by a trained panel, indicating that the test protocols were suitable for the majority of children.

Further analysis revealed that the latter two groups likely adopted different test strategies: The "unsteady tickers" increased their number of ticks over the test, implying a sample spacerelative test strategy. In contrast, the "steady tickers" might have used the list in a more absolute way. Future research may investigate if children displaying a sample-space relative strategy in CATA are more capable of conducting other sample-space relative methods, such as projective mapping, than those relying on absolute strategies.

In our discussion we provided an overview of suitable child-friendly adaptations of the CATA test protocol for future studies. Future research should also aim at better understanding the effects of ticking style in other product categories and potential ticking groups in adult population.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764985, the Research Council of Norway (project no. 233831/E50) and the Norwegian Foundation for Research Levy on Agricultural Products FFL, through the research program "FoodSMaCK, Spectroscopy, Modelling and Consumer Knowledge" (2017-2020). The authors would like to thank the research assistants and the children for their participation in the study. Also, a warm thanks to professor Kristan Hovde Liland and Ingunn Berget for their valuable advices regarding $R$ programming.

## References

Anderson, V. (1998). Assessing Executive Functions in Children: Biological, Psychological, and Developmental Considerations. Neuropsychological Rehabilitation, 8(3), 319-349. doi:10.1080/713755568
Ares, G., Barreiro, C., Deliza, R., Giménez, A. N. A., \& Gámbaro, A. (2010). Application of a Check-All-That-Apply Question to the Development of Chocolate Milk Desserts. Journal of Sensory Studies, no-no. doi:10.1111/j.1745-459X.2010.00290.x
Bruzzone, F., Ares, G., \& Giménez, A. N. A. (2012). Consumers' Texture Perception of Milk Desserts. li Comparison with Trained Assessors' Data. Journal of Texture Studies, 43(3), 214-226. doi:10.1111/j.1745-4603.2011.00332.x
Cardinal, P., Zamora, M. C., Chambers, E., Carbonell Barrachina, Á., \& Hough, G. (2015). Convenience Sampling for Acceptability and CATA Measurements May Provide Inaccurate Results: A Case Study with Fruit-Flavored Powdered beverages Tested in Argentina, Spain and U.S.A. Journal of Sensory Studies, 30(4), 295-304. doi:10.1111/joss. 12158
Daltoe, M. L. M., Breda, L. S., Belusso, A. C., Nogueira, B. A., Rodrigues, D. P., Fiszman, S., \& Varela, P. (2017). Projective mapping with food stickers: A good tool for better understanding perception of fish in children of different ages. Food Quality and Preference, 57, 87-96. doi:10.1016/j.foodqual.2016.12.003
De Pelsmaeker, S., Schouteten, J., \& Gellynck, X. (2013). The consumption of flavored milk among a children population. The influence of beliefs and the association of brands with emotions. Appetite, 71, 279-286. doi:10.1016/j.appet.2013.08.016
Dehlholm, C., Brockhoff, P. B., Meinert, L., Aaslyng, M. D., \& Bredie, W. L. P. (2012). Rapid descriptive sensory methods - Comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. Food Quality and Preference, 26(2), 267-277. doi:10.1016/j.foodqual.2012.02.012
Delarue, J., Lawlor, J. B., \& Rogeaux, M. (2015). Rapid Sensory Profiling Techniques, Applications in New Product Development and Consumer Research (1. Ed.): Woodhead Publishing.
Deroy, O., \& Valentin, D. (2011). Tasting Liquid Shapes: Investigating the Sensory Basis of Cross-modal Correspondences. Chemosensory Perception, 4(3), 80-90. doi:10.1007/s12078-011-9097-1
Dooley, L., Lee, Y.-s., \& Meullenet, J.-F. (2010). The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. Food Quality and Preference, 21(4), 394-401. doi:10.1016/j.foodqual.2009.10.002
Jaeger, S. R., Beresford, M. K., Paisley, A. G., Antúnez, L., Vidal, L., Cadena, R. S., . . . Ares, G. (2015). Check-all-that-apply (CATA) questions for sensory product characterization by consumers: Investigations into the number of terms used in CATA questions. Food Quality and Preference, 42, 154-164. doi:10.1016/j.foodqual.2015.02.003
Jaeger, S. R., Chheang, S. L., Yin, J., Bava, C. M., Gimenez, A., Vidal, L., \& Ares, G. (2013). Check-all-thatapply (CATA) responses elicited by consumers: Within-assessor reproducibility and stability of sensory product characterizations. Food Quality and Preference, 30(1), 56-67. doi:10.1016/j.foodqual.2013.04.009
Laureati, M., Cattaneo, C., Lavelli, V., Bergamaschi, V., Riso, P., \& Pagliarini, E. (2017). Application of the check-all-that-apply method (CATA) to get insights on children's drivers of liking of fiberenriched apple purees. Journal of Sensory Studies, 32(2). doi:ARTN e12253
Laureati, M., Pagliarini, E., Toschi, T. G., \& Monteleone, E. (2015). Research challenges and methods to study food preferences in school-aged children: A review of the last 15 years. Food Quality and Preference, 46, 92-102. doi:10.1016/j.foodqual.2015.07.010
Lawless, H. T., \& Heymann, H. (1998). Context Effects and Biases in Sensory Judgment. In Sensory Evaluation of Food: Principles and Practices (second ed.): Springer.

Lê, S., Josse, J., \& Husson, F. (2008). FactoMineR: A Package for Multivariate Analysis. Journal of Statistical Software, 25(1), 1-18. doi:10.18637/jss.v025.i01
Lima, M., Ares, G., \& Deliza, R. (2018). Children and adults' sensory and hedonic perception of added sugar reduction in grape nectar. Journal of Sensory Studies, 33(2). doi:10.1111/joss. 12317
Meyners, M., \& Castura, J. C. (2016). Randomization of CATA attributes: Should attribute lists be allocated to assessors or to samples? Food Quality and Preference, 48, 210-215. doi:10.1016/j.foodqual.2015.09.014
Næs, T., Varela, P., \& Berget, I. (2018). Individual Differences in Sensory and Consumer Science, Experimentation, Analysis and Interpretation: Woodhead Publishing.
Piaget, J. (1964). Development of Learning. Journal of Research in Science teaching, 2, 176-186.
Schouteten, J. J., De Steur, H., Lagast, S., De Pelsmaeker, S., \& Gellynck, X. (2017). Emotional and sensory profiling by children and teenagers: A case study of the check-all-that-apply method on biscuits. Journal of Sensory Studies, 32(1). doi:10.1111/joss. 12249
Schouteten, J. J., Verwaeren, J., Gellynck, X., \& Almli, V. L. (2019). Comparing a standardized to a product-specific emoji list for evaluating food products by children. Food Quality and Preference, 72, 86-97. doi:10.1016/j.foodqual.2018.09.007
Schouteten, J. J., Verwaeren, J., Lagast, S., Gellynck, X., \& De Steur, H. (2018). Emoji as a tool for measuring children's emotions when tasting food. Food Quality and Preference, 68, 322-331. doi:10.1016/j.foodqual.2018.03.005
Varela, P., \& Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. Food Research International, 48(2), 893-908. doi:10.1016/j.foodres.2012.06.037
Varela, P., \& Salvador, A. (2014). Structured sorting using pictures as a way to study nutritional and hedonic perception in children. Food Quality and Preference, 37, 27-34. doi:10.1016/j.foodqual.2014.04.009
Yoo, H. J., Machin, L., Arrua, A., Antunez, L., Vidal, L., Gimenez, A., . . . Ares, G. (2017). Children and adolescents' attitudes towards sugar reduction in dairy products. Food Res Int, 94, 108-114. doi:10.1016/j.foodres.2017.02.005
le
An overview of the test setup of previous sensory CATA studies with children which involved the tasting of samples

| Paper | Age group | CATA list generated by: | Numbe $r$ of CATA attribut es. | Questio nnaire form | Assistan ce during test | Validation of CATA |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Cardinal et al. (2015) | $11-12$ <br> Women <br> Food-sciencerelated consumers | Experimenter | 11** | n/a | n/a | - Comparison to women and Food-science-related consumers |
| Laureati et al. (2017) | 8-11 | Children | 11 | Paper | Yes | - Detection of pre-defined differences |
| Lima et al. (2018) | 6-12 <br> Adults | Experimenter | 6 | Paper | n/a | - Detection of pre-defined differences <br> - Comparison to adult panel |
| Schouteten et al. (2017) | $10-12$ <br> Teenagers | Children | 14 ** | n/a | No | - Discrimination of samples |

$\mathrm{n} / \mathrm{a}$ : Information not available

* pilot-tested with age group
** combined with non-sensory attributes

Table 2
Sample design with DoE factors. Low factor level=0, high factor level=1

|  | DoE Bread | DoE Smoothie |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Darkness | Thickness | Colour intensity | Acidity |
|  | Salt | Coarseness | 0 = - | $0=-$ | 0 = - | 0 = - |
| Sample | 0 = 0.4\% | 0 = fine flour | 1 = Caramel | 1 = Xanthan gum | 1 = Beetroot | 1 = Lemon juice |
| name | 1 = 1.2\% | 1 = coarse flour | colouring |  | powder |  |
| 1,1_1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 | 0 | 1 | 0 |
| 4 | 1 | 1 | 0 | 1 | 1 | 0 |
| 5 | 0 | 0 | 1 | 0 | 0 | 1 |
| 6 | 1 | 0 | 1 | 1 | 0 | 1 |
| 7 | 0 | 1 | 1 | 0 | 1 | 1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 |

Table 3
Bread: Significance of CATA attributes for total child panel and ticking style groups

| Dataset | Related to DoE factor | CATA attributes | Number of ticks total | Cochran's Q Test (p-values) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $\begin{aligned} & \text { Total }(\mathrm{N} \\ & =83) \end{aligned}$ | Few tickers $(\mathrm{N}=28)$ | Unsteady tickers ( $\mathrm{N}=27$ ) | Steady tickers ( $\mathrm{N}=28$ ) |
| Bread | Darkness | Light colour | 290 | 0.000 *** | 0.000*** | 0.000 *** | $0.000^{* * *}$ |
|  |  | Dark colour | 245 | 0.000*** | 0.000*** | 0.000*** | 0.000*** |
|  | Coarseness | Not grainy | 172 | 0.000*** | 0.026* | 0.002** | 0.000*** |
|  |  | Grainy | 273 | 0.000 *** | 0.000*** | $0.000^{* * *}$ | 0.000 *** |
|  |  | Easy to chew | 405 | 0.057 | 0.822 | 0.047* | 0.489 |
|  |  | Hard to chew | 95 | 0.179 | 0.280 | $0.069^{\circ}$ | 0.688 |
|  |  | Not coarse | 137 | 0.000 *** | 0.664 | 0.012* | $0.000^{* * *}$ |
|  |  | Coarse | 216 | 0.000 *** | 0.022* | 0.001** | $0.000 * * *$ |
|  | Salt | No salty taste | 255 | 0.094 | 0.525 | 0.875 | 0.368 |
|  |  | Salty taste | 174 | 0.012* | 0.555 | $0.063{ }^{\circ}$ | 0.184 |
| Note: signi | ts are shown with | *p $<0.05$, **p | 01, ***p < | 001, p < 0 | 5 for ticki | style grou |  |

Table 4
Bread: Influence of DoE factors on 7-point-liking rating, p -values DoE factor p-values

|  | Total <br> $(\mathrm{N}=83)$ | Few <br> tickers <br> $(\mathrm{N}=28)$ | Unsteady <br> tickers <br> $(\mathrm{N}=27)$ | Steady <br> tickers <br> $(\mathrm{N}=28)$ |
| :--- | :--- | :--- | :--- | :--- |
| Darkness | 0.283 | 0.274 | 0.159 | 0.878 |
| Coarseness | $0.012^{*}$ | $0.042^{*}$ | $0.106^{*}$ | 0.586 |
| Salt | $0.000^{* * *}$ | 0.251 | $0.000^{* *}$ | $0.021^{*}$ |
| Darkness $\times$ Coarseness | $0.000^{* * *}$ | 0.322 | $0.004^{* *}$ | $0.011^{*}$ |
| Darkness $\times$ Salt | 0.483 | 0.138 | 0.218 | 0.298 |
| Coarseness $\times$ Salt | 0.666 | 0.766 | 0.749 | 0.496 |

Table 5
Smoothie: Significance of CATA attributes for total child panel and ticking style groups

| Dataset | Related to DoE factor | CATA attributes | Number of ticks total | Cochran's Q Test (p-values) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $\begin{aligned} & \text { Total }(\mathrm{N} \\ & =92) \end{aligned}$ | Few tickers ( $\mathrm{N}=31$ ) | Unsteady tickers ( $\mathrm{N}=30$ ) | Steady tickers ( $\mathrm{N}=31$ ) |
| Smoothie | Colour intensity | Light colour | 123 | 0.000*** | 0.001** | 0.030* | 0.024* |
|  |  | Dark colour | 323 | 0.000 *** | 0.020* | 0.278 | $0.007 * *$ |
|  | Thickness | Bubbles | 250 | 0.064 | 0.165 | 0.594 | 0.037* |
|  |  | Thin | 201 | 0.005** | 0.479 | 0.257 | $0.008^{* *}$ |
|  |  | Thick | 245 | 0.000*** | 0.259 | 0.008** | 0.000*** |
|  |  | Slimy | 167 | 0.002** | 0.865 | 0.053 | 0.030* |
|  | Acidity | Very sour | 226 | 0.000 *** | $0.001^{* *}$ | 0.009* | 0.000*** |
|  |  | Lemon | 267 | 0.000*** | 0.002** | 0.048* | 0.000*** |
|  | Acidity (indirect) | Banana | 237 | 0.005** | 0.165 | 0.011* | 0.151 |
|  |  | Strawberry | 315 | 0.000*** | 0.002** | $0.136{ }^{\circ}$ | $0.001 * *$ |
|  |  | Raspberry | 308 | 0.155 | 0.559 | 0.152 | 0.772 |
|  |  | Blueberry | 267 | 0.005** | $0.069{ }^{\circ}$ | 0.200 | 0.234 |
|  | Other (Odour) | Strong smell | 149 | 0.203 | 0.780 | 0.728 | 0.192 |
|  | Other <br> (Hedonics) | Yummy | 262 | 0.000*** | 0.435 | 0.000*** | 0.000*** |
|  |  | Yuck | 91 | 0.000*** | 0.018* | 0.034* | 0.003** |

Note: significant effects are shown with ${ }^{*}:{ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001,{ }^{\circ} p<0.15$ for ticking style groups

Table 6
Smoothie: Influence of DoE factors on 7-point-liking rating, p -values

| DoE factor | p-values |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $\begin{aligned} & \text { Total } \\ & (\mathrm{N}=93) \end{aligned}$ | Few tickers ( $\mathrm{N}=31$ ) | Unsteady tickers $(\mathrm{N}=30)$ | Steady tickers ( $\mathrm{N}=31$ ) |
| Colour intensity | 0.255 | 0.198 | 0.568 | 0.174 |
| Thickness | 0.306 | $0.054^{\circ}$ | 0.897 | 0.846 |
| Acidity | 0.000*** | 0.000*** | 0.000*** | 0.000*** |
| Colour intensity x Thickness | 0.795 | 0.846 | 0.967 | 0.481 |
| Colour intensity x Acidity | 0.465 | 0.332 | 0.708 | 0.901 |
| Thickness x Acidity | 0.165 | 0.415 | $0.090^{\circ}$ | 0.901 |

Note: significant effects are shown with ${ }^{*}:{ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001,{ }^{\circ} p<0.15$ for ticking style groups

Table 7
Similarity of perceptual space: children and trained profile panel
SMI Index comparing dimension 1, dimensions 1 to 2 and dimensions 1 to 3 of the score plots
SMI: similarity between CATA and QDA

| Dataset | Component(s) | SMi: similarity between CATA and QDA |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Total (N=83/92) | Few tickers ( $\mathrm{N}=28 / 31$ ) | Unsteady tickers ( $\mathrm{N}=27 / 30$ ) | Steady tickers ( $\mathrm{N}=28 / 31$ ) |
| Bread | 1 | 0.98 | 0.97 | 0.96 | 0.98 |
|  | 1 to 2 | 0.92 | 0.92 | 0.93 | 0.88 |
|  | 1 to 3 | 0.94 | 0.73 | 0.92 | 0.90 |
| Smoothie | 1 | 0.81 | 0.66 | 0.78 | 0.87 |
|  | 1 to 2 | 0.77 | 0.39 | 0.77 | 0.74 |
|  | 1 to 3 | 0.93 | 0.44 | 0.89 | 0.94 |

$\qquad$
$\qquad$



Table 8
Potential influences on ticking number of "unsteady tickers": sample, ranked liking and tasting position

| Dataset | Variables | P-value |
| :--- | :--- | :--- |
| Bread, unsteady | Sample | 0.422 |
| ticking group, $\mathrm{N}=27$ | Liking (ranked) | 0.795 |
|  | Tasting position | $0.000^{\star * *}$ |
| Smoothie, unsteady | Sample | 0.571 |
| ticking group, $\mathrm{N}=30$ | Liking (ranked) | 0.068 |
|  | Tasting position | $0.000^{\star * *}$ |

Note: significant effects are shown with ${ }^{*}:{ }^{* * *} p<0.001$

Table 9
Challenges and recommendations for CATA tests with children

| Challenge | Recommendation | Comment |
| :---: | :---: | :---: |
| Understanding CATA attributes and relating them to samples | Vocabulary development with children of targeted age group based on samples in experiment | A repertory grid approach may be used to generate attributes |
| Reading effort dominates the task | Pre-familiarisation with the CATA list | For the youngest, reading in class and/or parents prior to the test is recommended |
| Time-consuming (reading) | Use as little text as necessary for instructions | Better to do a live instruction than explaining in text |
| Skipping pages | Usage of tablets | Pages cannot be skipped, and children handle tablets more easily than multi-page documents. |
| Forget to rinse mouth with water between samples | Reminder screen | Use an image (e.g. a glass of water) rather than a sentence |
| Losing interest after a few samples | Give a child-friendly purpose to the study | Inviting children to help adults is engaging. Use an age-appropriate cover story. It doesn't need to be credible as children under 10 y.o. enjoy fantasy. |
| Few attributes selected | Read the word, taste and tick if it applies | "click all words that apply" is too generic and they may not go through the list systematically |
| Confuse samples | Usage of distinct symbols or alphabetic letters | A,B,C,D,E,F,G,H |
| Ideal product is misunderstood | Trigger children's' imagination |  |
| CATA list is applied in an absolute manner, not restricted to the sample space | Include a "warm-up" sample | The list will be used in a sample-space relative manner. |

## Appendix

## A.1. Liking

## Table A1

Bread: Average 7-point liking rating of children

| Sample | DoE |  |  | Average Liking |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\stackrel{\rightharpoonup}{\widetilde{\sim}}$ | $$ | $\begin{aligned} & \mathscr{0} \\ & 0 \\ & \frac{1}{4} \\ & \stackrel{\rightharpoonup}{\overleftarrow{ }} \end{aligned}$ | Total $(\mathrm{N}=83)$ | Few tickers ( $\mathrm{N}=28$ ) | Unsteady tickers ( $\mathrm{N}=27$ ) | Steady tickers ( $\mathrm{N}=28$ ) |
| 1 | 0 | 0 | 0 | 5.0 | 5.4 | 4.9 | 4.8 |
| 2 | 1 | 0 | 0 | 6.0 | 6.1 | 6.0 | 5.9 |
| 3 | 0 | 1 | 0 | 4.4 | 4.7 | 4.0 | 4.5 |
| 4 | 1 | 1 | 0 | 5.0 | 5.2 | 4.9 | 5.0 |
| 5 | 0 | 0 | 1 | 4.7 | 5.4 | 3.8 | 4.7 |
| 6 | 1 | 0 | 1 | 5.2 | 5.3 | 5.3 | 5.1 |


| 7 | 0 | 1 | 1 | 4.6 | 4.8 | 4.0 | 5.0 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 8 | 1 | 1 | 1 | 5.4 | 5.0 | 5.5 | 5.5 |

Table A2
Smoothie: Average 7-point liking rating of children

| Sample | DoE |  |  | Average Liking |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{aligned} & \mathscr{0} \\ & \text { D } \\ & \text { S } \\ & \text { J } \\ & \text { N } \end{aligned}$ | $\begin{aligned} & \stackrel{\rightharpoonup}{0} \\ & \stackrel{y}{0} \\ & \stackrel{y}{5} \\ & \vdots \\ & 0 \\ & 0 \end{aligned}$ | $\frac{訁}{0}$ | Total $(\mathrm{N}=92)$ | Few tickers ( $\mathrm{N}=31$ ) | Unsteady tickers ( $\mathrm{N}=30$ ) | Steady tickers ( $\mathrm{N}=31$ ) |
| 1 | 0 | 0 | 0 | 5.6 | 5.5 | 6.0 | 5.2 |
| 2 | 1 | 0 | 0 | 5.6 | 6.0 | 5.6 | 5.1 |
| 3 | 0 | 1 | 0 | 5.5 | 5.8 | 5.2 | 5.5 |
| 4 | 1 | 1 | 0 | 6.1 | 6.5 | 6.3 | 5.4 |
| 5 | 0 | 0 | 1 | 4.2 | 4.4 | 4.2 | 3.9 |
| 6 | 1 | 0 | 1 | 4.3 | 4.7 | 4.6 | 3.7 |
| 7 | 0 | 1 | 1 | 4.4 | 4.5 | 4.8 | 4.0 |
| 8 | 1 | 1 | 1 | 4.2 | 4.7 | 3.6 | 4.3 |

## A.2. QDA

## Table A3

Bread: QDA, average intensity rating (10-cm-scale) and p-values

|  | Average intensity rating per sample |  |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: |
| Attributes | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ | p-value |  |  |  |
| Acidic odor | 4.5 | 5.0 | 5.0 | 4.9 | 4.1 | 4.2 | 4.1 | 4.4 | $0.000^{* * *}$ |  |  |  |
| Grain odor | 3.9 | 4.4 | 5.2 | 4.9 | 3.8 | 3.9 | 4.0 | 4.2 | $0.000^{* * *}$ |  |  |  |
| Cloying odor | 1.6 | 1.6 | 1.1 | 1.2 | 1.8 | 1.8 | 1.5 | 1.5 | $0.003^{* *}$ |  |  |  |
| Colour hue | 4.9 | 5.0 | 4.9 | 4.9 | 5.0 | 5.0 | 5.1 | 5.1 | $0.007^{* *}$ |  |  |  |
| Colour strength | 2.0 | 1.8 | 2.0 | 2.1 | 4.2 | 4.2 | 4.3 | 4.3 | $0.000^{* * *}$ |  |  |  |
| Whiteness | 7.6 | 7.8 | 7.7 | 7.7 | 3.5 | 3.7 | 3.4 | 3.2 | $0.000^{* * *}$ |  |  |  |
| Hardness | 3.7 | 3.6 | 4.0 | 4.3 | 3.6 | 3.6 | 4.2 | 4.2 | $0.000^{* * *}$ |  |  |  |
| Juiciness | 4.6 | 4.9 | 4.7 | 4.9 | 4.6 | 4.7 | 4.5 | 4.8 | 0.098 |  |  |  |
| Coarseness | 3.9 | 3.7 | 5.1 | 5.2 | 3.8 | 3.8 | 5.2 | 5.4 | $0.000^{* * *}$ |  |  |  |
| Chewing |  |  |  |  |  |  |  |  |  |  |  |  |
| resistance | 3.7 | 3.8 | 5.0 | 5.1 | 3.6 | 3.9 | 5.1 | 5.0 | $0.000^{* * *}$ |  |  |  |
| Sticky | 3.7 | 3.7 | 3.0 | 3.1 | 3.5 | 3.6 | 3.1 | 3.1 | $0.000^{* * *}$ |  |  |  |
| Doughy | 4.0 | 3.9 | 2.8 | 3.0 | 3.7 | 4.0 | 3.0 | 3.1 | $0.000^{* * *}$ |  |  |  |
| Acidic taste | 4.1 | 4.9 | 4.5 | 4.9 | 3.3 | 3.8 | 3.5 | 4.6 | $0.000^{* * *}$ |  |  |  |
| Sweet taste | 2.9 | 2.9 | 2.8 | 2.8 | 2.8 | 2.8 | 2.6 | 3.1 | $0.007^{* *}$ |  |  |  |
| Salt taste | 2.5 | 3.6 | 2.6 | 3.7 | 2.4 | 3.5 | 2.2 | 3.6 | $0.000^{* * *}$ |  |  |  |
| Bitter taste | 3.3 | 3.6 | 3.3 | 3.5 | 3.5 | 3.8 | 3.7 | 3.7 | $0.005^{* *}$ |  |  |  |
| Corn taste | 3.8 | 4.2 | 4.8 | 5.1 | 3.4 | 3.6 | 4.2 | 4.9 | $0.000^{* * *}$ |  |  |  |
| Cloyingtaste | 3.2 | 1.9 | 2.7 | 1.2 | 3.5 | 2.2 | 3.3 | 1.5 | $0.000^{* * *}$ |  |  |  |

Note: significant effects are shown with $*:{ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

## Table A4

Smoothie: QDA, average intensity rating (10-cm-scale) and $p$-values

|  | Average intensity rating per sample |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Attributes | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ | p-value |
| Intensity smell | 5.3 | 5.2 | 5.6 | 6.0 | 5.6 | 5.7 | 5.9 | 5.8 | 0.121 |
| Acidity smell | 4.7 | 4.7 | 5.0 | 4.8 | 4.2 | 4.3 | 4.0 | 4.3 | 0.081 |
| Fruity Berry smell | 5.8 | 6.3 | 6.0 | 6.2 | 5.3 | 5.4 | 5.2 | 5.6 | $0.019^{\star}$ |
| Artifical smell | 1.6 | 1.4 | 1.9 | 2.0 | 2.5 | 2.9 | 2.7 | 2.5 | $0.018^{*}$ |
| Colour intensity | 4.5 | 4.0 | 6.4 | 6.4 | 4.5 | 4.4 | 6.3 | 5.9 | $0.000^{* * *}$ |
| Whiteness | 5.3 | 5.7 | 3.6 | 3.5 | 5.3 | 5.4 | 3.6 | 4.0 | $0.000^{\star * *}$ |
| Taste intensity | 6.1 | 5.7 | 6.5 | 6.2 | 7.0 | 7.0 | 7.0 | 6.5 | $0.000^{* * *}$ |
| Acidity | 5.2 | 4.9 | 5.1 | 5.0 | 3.9 | 3.5 | 3.8 | 3.7 | $0.000^{* * *}$ |
| Sweetness | 4.8 | 4.5 | 4.5 | 5.0 | 3.8 | 3.7 | 3.7 | 4.1 | $0.000^{\star * *}$ |
| Sourness | 4.7 | 5.1 | 5.2 | 4.5 | 6.5 | 6.7 | 6.5 | 6.3 | $0.000^{* * *}$ |
| Bitterness | 3.5 | 3.6 | 3.6 | 4.3 | 4.3 | 4.3 | 4.0 | 4.3 | $0.002^{* *}$ |
| Metallic | 2.5 | 2.8 | 2.6 | 2.7 | 2.8 | 2.7 | 2.7 | 2.5 | 0.469 |


| Fruit Berry | 6.9 | 6.3 | 6.5 | 6.3 | 5.5 | 5.3 | 5.5 | 5.4 | $0.000^{* * *}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Artificial | 1.4 | 1.6 | 2.0 | 2.7 | 3.7 | 4.3 | 3.8 | 3.8 | $0.000^{* * *}$ |
| Fullness | 4.4 | 5.5 | 4.5 | 5.8 | 4.0 | 5.6 | 4.4 | 5.6 | $0.000^{* * *}$ |
| Viscosity | 2.2 | 4.9 | 2.6 | 4.9 | 2.4 | 4.9 | 2.7 | 5.2 | $0.000^{* * *}$ |
| Astringency | 3.8 | 3.6 | 4.0 | 3.6 | 5.0 | 5.2 | 5.3 | 4.8 | $0.000^{* * *}$ |
| Pungency | 2.3 | 2.7 | 3.7 | 3.1 | 4.6 | 4.7 | 4.7 | 4.6 | $0.000^{* * *}$ |
| Note: significant effects are shown with ${ }^{*}:{ }^{*} \mathrm{p}<0.05,{ }^{* *} \mathrm{p}<0.01,{ }^{* * *} \mathrm{p}<0.001$ |  |  |  |  |  |  |  |  |  |

## Figure captions

Fig. 1 Four of the bread samples in study 1, illustrating the visual differences corresponding to sample numbers as indicated

Fig. 2 Paper questionnaire in the Bread test

Fig. 3 Electronic questionnaire in the Smoothie test: a): welcome screen, b): gender and grade input, c): liking, d): CATA, e): reminder to rinse, e): thank you end screen

Fig. 4 Ticking style indicators (number, std and attributes) for the Bread (left) and Smoothie (right) study. Histogram of distribution in the diagonal, visual correlation in the lower panel and Pearson correlation in the upper panel.

Fig. 5 PCA Biplot of ticking style indicators, individuals grouped according to ticking style. Three ticking style groups were built based on the first two PCA components of the standardized ticking style indicators. Individuals are coloured according to the ticking style group.

Fig. 6 Mosaic plot displaying the ticking style group sizes per age group in Bread / school grade (corresponding to the year of birth) in Smoothie.

Fig. 7 Bread: Score plots: left: CA, middle: MFA, right: PCA each with liking in correlation circle. For better interpretation of samples the DoE factor levels are projected as supplementary variables. The centre of text corresponds to the exact location. In the MFA the partial coordinates of the DoE factor levels of each ticking style group are connected to the overall MFA configuration. Top row: Component 1 \& 2, bottom row: Component 2 \& 3

Fig. 8 Smoothie: Score plots: left: CA, middle: MFA, right: PCA each with liking in correlation circle. For better interpretation of samples the DoE factor levels are projected as supplementary variables. The centre of text corresponds to the exact location. In the MFA the partial coordinates of the DoE factor
levels of each ticking style group are connected to the overall MFA configuration. Top row: Component 1 \& 2, bottom row: Component 2 \& 3

Fig. 9 Average number of ticks and tasting position of sample for ticking style groups

## Supplementary material

Fig. S. 1 Bread: Perceptual space of children, CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, not significant attributes are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 2 Bread: Perceptual space of "few tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with $p$-values $>0.15$ are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S.3 Bread: Perceptual space of "unsteady tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with $p$-values $>0.15$ are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 4 Bread: Perceptual space of "steady tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with p -values $>0.15$ are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component $2 \& 3$

Fig. S. 5 Bead: Perceptual space of trained profile panel, PCA with QDA data. Left: Score plots with projected DoE factor levels, Middle: loading plots, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 6 Smoothie: Perceptual space of children, CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, not significant attributes are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 7 Smoothie: Perceptual space of "few tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with p-values > 0.15 are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 8 Smoothie: Perceptual space of "unsteady tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with pvalues > 0.15 are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3

Fig. S. 9 Smoothie: Perceptual space of "steady tickers", CA of CATA. Left: Score plots with projected DoE factor levels, Middle: loading plots, sizes represent ticking number, attributes with p-values $>0.15$ are displayed in lighter colour, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component $2 \& 3$

Fig. S. 10 Smoothie: Perceptual space of trained profile panel, PCA with QDA data. Left: Score plots with projected DoE factor levels, Middle: loading plots, Right: Correlation of liking. Upper: Component 1 \& 2, Lower: Component 2 \& 3



## Test på brød

Erdu:ir jente
$\square \mathcal{R}_{\text {gutt }}$

Alder?


- 6 âr
- 7 ăr
- 8 âr
- 9 ăr

Hvor godt liker du brod E?



## Bread



## Smoothie




## Bread

Smoothie



CATA children ticking style groups



QDA trained panel





Bread

Smoothie







CA loadings










CA scores



CA loadings




CA loadings

















CA scores







CA scores



CA loadings




CA scores






## Credit Author Statement

Martina Galler: Ideation, investigation, formal analysis, visualization, writing of manuscript
Tormod Næs: Ideation, conceptualization, supervision, reviewing of manuscript Valérie L. Almli: Data curation, methodology, reviewing of manuscript
Paula Varela: Ideation, conceptualization, data curation, methodology, supervision, reviewing of manuscript

