***Highlights (for review)**

- Semi-trained (ST) CATA was performed by consumers that had 1h reference training
- ST-CATA (N=37) was compared to consumer CATA (N=70) and descriptive analysis (DA)
- ST-CATA provided reliable and highly similar qualitative sample descriptions to DA
- Training increased the citation frequency of terms in ST- when compared to C-CATA
- Quantitative differences derived from CATA counts differed from DA's intensities

Check-All-That-Apply (CATA) with semi-trained assessors: sensory profiles closer to descriptive analysis or consumer elicited data?

Alexi N.[1,2*], Nanou E.[2], Lazo O.[3], Guerrero L.[3], Grigorakis, K.[2] & Byrne D.V.[1]

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

[2]Institute of Marine Biology, Biotechnology & Aquaculture, Hellenic Centre for Marine Research, Agios Kosmas Hellinikon, 16777, Athens, Greece

[3]IRTA-Food Technology, Spain

*Corresponding author.

Niki Alexi

Department of Food Science, Food Quality Perception & Society, Aarhus University, Kirstinebjergvej 10, 5792 Aarslev

& Institute of Marine Biology, Biotechnology & Aquaculture, Hellenic Centre for Marine Research, Agios Kosmas, 6777 Hellinikon, Athens

Tel.: +306972242284

Email: niki.alexi@food.au.dk

ABSTRACT

Check-All-That-Apply (CATA) is a simple and fast sensory profiling tool. Yet, its application has been mainly focused on consumer studies; the aim of this study was to evaluate the application of CATA with semi-trained (ST) individuals (N=37). ST individuals were consumers who underwent 1h of training with physical references on the definition of attributes included in the CATA ballot. ST-CATA results were compared, on a panel level, to Descriptive Analysis (DA) with trained panellists (N=8) and to CATA with consumers (N=70). Moreover, the effect of training was examined, to uncover training vs. method-related variations in CATA profiling.

ST-CATA and DA exhibited the highest similarity in sample configurations (94%) for two Multiple Factor Analysis factors. For all 3 factors, similarity was over 95% for all method combinations; however the *RV* coefficient between consumers and DA was marginally significant (*P*=0.08). The extent of explained sensory variations in ST-CATA was not negatively affected by the smaller panel size, compared to consumers' CATA. Training had a positive effect on attributes' citation frequency, identification of taste, flavour and complex attribute differences among samples. CATA results did not provide the same range of differences with DA, especially for texture.

Overall results support the validity of CATA with ST assessors and suggest its potential for industrial use, when a timely and cost-efficient description of products is required. Attention should be given though when a detailed quantitative profile of sample differences is required, since intensity is not well represented by CATA derived measurements due to the method constraints.

Keywords: training; CATA; fast method; consumers; descriptive analysis

1 Introduction[1]

Descriptive analysis (DA) has been the main sensory science tool to acquire detailed, reliable and reproducible data to describe the sensory profiles of food products. However, DA lacks cost- and time efficiency and therefore it can be largely unsustainable in practice for the industry in some cases (Byrne, O'Sullivan, Dijksterhuis, Bredie, & Martens, 2001; Murray, Delahunty, & Baxter, 2001; Valentin, Chollet, Lelievre, & Abdi, 2012). This led to the development of several fast sensory methods (Ares, 2015). Among them, Check-All-That-Apply (CATA) has gained popularity mainly due to its simple format, small cognitive effort requirements and rapid elicitation of sensory characteristics of the examined products from participants (Adams, Williams, Lancaster, & Foley, 2007; Ares, Varela, Rado, & Giménez, 2011; Meyners & Castura, 2014). In addition, CATA is a non-holistic method since it does not require a simultaneous evaluation of all samples, which makes it appropriate for large product sets and/or when monadic presentation order of samples is required (Ares, 2015).

However, low discrimination ability in product sets with subtle differences has also been reported for the CATA method (Ares et al., 2015; Reinbach, Giacalone, Ribeiro, Bredie, & Frøst, 2014). This is attributed to the dichotomous nature (0/1) of the CATA responses, which can lead to incapacity to reflect intensity differences for the same sensory attribute (Lazo, Claret, & Guerrero, 2016). Combined with its simple and rapid nature, the aforementioned limitation categorized CATA mainly as a sensory consumer research tool, appropriate when applied to a large set of participants (Ares, Tárrega, Izquierdo, & Jaeger, 2014; Varela & Ares, 2012). Specifically, the minimum recommendation for CATA is N = 60-80 consumers (Ares, Tárrega, et al.,

[1]TP: Trained panel; ST: Semi-trained; C: Consumer

2014; Varela & Ares, 2012), whereas when the numbers of untrained participants becomes <30 non-discriminant sample profiles emerge (Cruz et al., 2013). Yet, the large participants' number is not solely a characteristic of CATA, but a requirement for all consumer-based methods to ensure validity, due to the inconsistencies in measurements deriving among other reasons from lack of training (Hough, 1998). Specifically, training can be a meaningful measure to treat inconsistencies in descriptive sensory tests where scaling and definition of attributes is an issue.

The effect of a short training prior to evaluation has been examined for a rating variant of CATA (Rate-All-That-Apply, RATA), providing promising results in terms of validity and repeatability even with a low number of participants (N=11) (Giacalone & Hedelund, 2016). Similarly, the number of consumers required to acquire a reliable sensory profile using the CATA methodology might also expected to decrease after training, providing a viable alternative to cover the industry's need for time and cost-efficient methodologies. However, before advocating for method change, there is a need for evaluating the validity of the results obtained (Ares, 2015), since it is not known how training would affect the quality of CATA results from a smaller consumer panel than suggested by literature.

The current paper addresses this question by comparing a CATA performed by semi-trained assessors (consumers who underwent 1 h training on the attributes' definition via physical references, prior to CATA) both to DA (trained panellists who underwent 30h training with physical references on the definition and scaling of the attributes) and CATA performed by consumers (untrained participants that received a written definition of the attributes prior to evaluation). To facilitate this comparison the same vocabulary was used across the three methodologies. The aim of this comparison was not only to examine the degree of similarity between the three methods, but also to differentiate between method-related and training-related sources of variation as well as to uncover how written attribute definition vs. definition with physical references can affect the consumers' evaluation of different types of attributes. To achieve this, results of the three methodologies were compared in terms of:

- sensory attribute and product variation explained
- configuration and description of samples
- representation and quantification of differences, among samples, for the same sensory attribute

2 Materials and methods

2.1 Samples

Four different fish species, namely meagre (*Argyrosomus regius*), greater amberjack (*Seriola dumerili*), pikeperch (*Sander lucioperca*) and wreckfish (*Polyprion americanus*), were profiled using all 3 methodologies. To allow valid comparisons across methodologies, species rearing, origin and conditions, fish filleting and fillet storage (-20°C, vacuum-packed) were identical in all cases.

Moreover, preparation, cooking and serving of samples to participants were performed under similar conditions. Specifically, fish fillets were cut in 2 cm^3 cubes; cubes were placed inside individual containers covered with a lid, in which they were cooked at 110-115 °C for 20 min, in a convection oven. Samples were served to the participants within the same containers at approx. 60°C.

The containers used for cooking and serving of samples were transparent glass jars for descriptive analysis and black ceramic jars of the same dimensions for the semi-trained and consumer analyses. Black containers were used instead of transparent ones, since they facilitated the evaluation of appearance attributes, due to contrast effects, which was crucial in the case of partly trained or untrained individuals.

2.2 Sensory vocabulary development

The development and selection of the 22 sensory attributes used in the DA was performed by the trained panel (TP); the process of the attribute selection for TP-DA can be found in Lazo et al. (2017). The list of sensory attributes used in the semi-trained CATA (ST-CATA) and consumers CATA (C-CATA) methodologies was similar to the one used in TP-DA with the exception of 3 appearance attributes which were split into two separate terms, corresponding to the anchors of the intensity scale (Table 1). This resulted in a 25 sensory attribute list used in both ST-CATA and C-CATA (Table 1).

The TP-DA sensory attribute list was translated from Spanish to English by experts in fish sensory evaluation in the Institute of Agricultural-Alimentary Research and Technology (IRTA) in Spain and then from English to Greek by experts in fish sensory evaluation in the Hellenic Centre for Marine Research (HCMR), in Greece. To ensure linguistic equivalence between Spanish and Greek, the definition of attributes, as specified by the TP, was used as an additional measure to ensure that the terms used reflected the same meaning between languages.

It should be mentioned that using the same sensory attributes across a trained and a consumer panel is not a common practice, since for consumer studies simpler and more generic terms are recommended, whereas technical, complex and specific terms, that are usually evaluated with a TP, are avoided (Giacalone, Bredie, & Frøst, 2013; Moussaoui & Varela, 2010; Van Trijp & Schifferstein, 1995). However, the decision to retain the same set of sensory terms across all 3 studies was made to explore the consumer performance on such a list of terms, and evaluate how this performance can change when untrained individuals receive a physical reference additionally to the written or no definition they commonly receive.

2.3 Sensory evaluation

The ST-CATA training, was divided in two half-hour sessions and focused on providing a clear definition of the attribute list, used in the CATA ballot. The first session included training on aroma and flavour attributes by the use of physical references (Table 1). The second training session was dedicated to the appearance and texture modalities. Attribute definition was achieved by images and physical references for appearance and texture attributes, respectively (Table 1). During the training sessions participants were also provided with a spreadsheet, which included a general definition of the attributes. Participants were advised to keep notes on the spreadsheet, corresponding to their own attribute perception when examining the references. General instructions on CATA task and tasting process were also given to participants during training.

The CATA ballot was divided in sensory modalities for which presentation followed the ‘dynamics of sensory perception’ (odour, appearance, taste/flavour and texture) to reduce the cognitive effort required by participants (Ares & Jaeger, 2013; Ares et al., 2013). Within modalities, attributes appeared in a fixed order across assessors, since participants were already familiar with the list through training. The ballot was filled in by hand and the CATA task was performed as described in Adams et al. (2007).

Sensory evaluation was performed in isolated sensory testing booths (ISO, 2007). The definition spreadsheet with the personal notes was available to participants during the whole process. All samples were evaluated in one session (approx. duration 45 min). Participants tasted each sample once (no replicates) and in a monadic sequence. Samples were blind-labelled with a three-digit code and the serving order of samples was randomized and balanced to account for first order and carry-over effects (MacFie, Bratchell, Greenhoff, & Vallis, 1989). Mineral water was provided to assessors to cleanse their palates between samples.

The semi-trained study was conducted in HCMR in Athens, Greece. The semi-trained panel consisted of 37 consumers, mainly HCMR employees, 22 to 60 years of age, with no previous experience in sensory profiling. A prerequisite for participation was that subjects were consumers of fish and/or fish products.

An overview of the three methodologies with all essential information needed for comparison is given in Table 2. A more detailed description of C-CATA evaluation process and TP-DA training conditions, reference material and evaluation, if of interest to the reader, is provided in (Alexi, Byrne, Nanou, & Grigorakis, 2017; Lazo et al., 2017)

2.4 Data analyses

2.4.1 Explained variances via Discriminant Partial Least square regression

Discriminant Partial Least Square Regression (D-PLSR) was employed to compare the amount of sensory and product variation between the methodologies. The partial least square regression approach is appropriate for analysis of both intensity and frequency (0/1) sensory data originating from DA and CATA, respectively (Giacalone et al., 2013; Martens, Bredie, & Martens, 2000; Reinbach et al., 2014; Rinnan, Giacalone, & Frøst, 2015). Three separate D-PLSR models were calculated, one for each methodology discussed herein. D-PLSR models were performed with predictor matrix {X} = the sensory variables, (attributes), and response matrix {Y} = the four fish samples. For the X matrix, input data for DA were the intensity measurements acquired per attribute, whereas for the two CATA datasets were the 0/1 measurements obtained by semi-trained subjects and consumers, respectively. The Y matrix was composed by 0/1 measurements used as indicators of the fish sample evaluated in each case. Cross validation of the models was performed by excluding one measurement at a time (full-cross validation) and one participant at a time (four consecutive measurements corresponding to the different samples) for DA and CATA data, respectively (Giacalone et al., 2013; Rinnan et al., 2015). D-PLSR models were performed in Unscrambler X® software, version 10.3 (CAMO, ASA, Norway).

2.4.2 Multiple Factor analysis and Regression Vector coefficients

In order to assess the configuration similarity of samples between sensory methodologies, Multiple Factor Analysis (MFA) was employed. For the MFA analysis three separate matrices were constructed: two frequency matrices corresponding to ST- and C-CATA; one matrix with average intensity ratings for the TP-DA. Each of the three matrices served as an individual group for performing the MFA analysis. To acquire a quantitative measure of proximity between sample configurations the Regression Vector (RV) coefficient was calculated for the first two and all three dimensions, for all possible combinations between methodologies (Robert & Escoufier, 1976). RV values range between 0 and 1, with 1 indicating the highest similarity between configurations obtained between two matrices. MFA analysis and RV coefficient calculations were carried out in the XLSTAT® software, 2016 (Addinsoft™).

2.4.3 Discrimination and sensory profiles of samples

For DA, determination of significant ($P < 0.05$) sample differences and pairwise comparisons between samples were performed with the SensMixed package (Kuznetsova, Brockhoff, & Christensen, 2013) in R 3.2.3. A mixed ANOVA model with interactions (fixed effect: samples; random effects: assessor, replica) was used. Estimation of significance was achieved by sequential elimination of non-significant random effects, following a procedure proposed by Kuznetsova, Christensen, Bavay, and Brockhoff (2015). For CATA datasets, Cochran's Q test, a non-parametric statistical test for estimating significance when the response variable is binary, was employed (Varela & Ares, 2012). For the construction of sensory maps, only significant attributes were included. Sensory maps were obtained via Principal Component Analysis (PCA) and Correspondence Analysis (CA) for DA and CATA datasets, respectively. Cochran's Q test, PCA and CA were performed in XLSTAT® software, 2016 (Addinsoft™).

2.4.4 Attribute ranges and citation frequencies

For each of the three methodologies, the normalized difference between minimum and maximum (normalized maximum range) in ratings of an attribute across samples was calculated. For DA, the normalized maximum range was calculated using the following equation:

$$Normalized\ maximum\ range\ (\%) = \frac{(Max_{attr.K} - Min_{attr.K})}{10} * 100\% \qquad (1)$$

where, $Max_{attr.K}$ is the maximum average panel value for attribute K and $Min_{attr.K}$ is the minimum average panel value for attribute K. Their difference is divided by 10, since a 10 cm-linear semi-structured scale was used for the evaluation of each attribute in DA (Lazo et al., 2017). For CATA, the normalized maximum range was calculated using the following equation:

$$Normalized\ maximum\ range\ (\%) = \frac{(Max_{attr.K} - Min_{attr.K})}{N} * 100\% \qquad (2)$$

where, $Max_{attr.K}$ is the maximum CATA count for attribute K, $Min_{attr.K}$ is minimum CATA count for attribute K and N is the total number of participants. For CATA, the normalized total citation frequency of an attribute was also calculated using the following equation:

$$Total\ citation\ frequency\ (\%) = \frac{SumCATAcounts_{attr.K}}{4 * N} * 100\% \qquad (3)$$

where, $SumCATAcounts_{attr.K}$ is the sum of CATA counts for all samples, N is the total number of participants multiplied by the number of samples (4).

3 Results

The results focus on underlining similarities and differences among methodologies in terms of explained sensory variability, configurational congruence, sample discrimination and description as well as quantification of differences between samples by each methodology. This will lead to an evaluation of the validity of results obtained by the ST-CATA variation, compared to both TP-DA and C-CATA obtained data. No detailed results on individual attributes measurements are included, since profiling of samples is out of the scope of the current study.

3.1 Explained variances via Discriminant Partial Least square regression

DA explained higher attribute and product variation than the CATA methods. The variation explained by the two CATA methods was similar. According to cross-validation results of D-PLSR, the optimum number of components was 4 for TP-DA and 3 for both CATA methods. The cumulative validated explained variance for the optimum number of components was: for the sensory attributes {X}, 50%, 29% and 25% for TP-DA, ST-CATA and C-CATA, respectively. Retaining the same number of components, explained product variation {Y} was 55%, 32% and 30% for TP-DA, ST-CATA and C-CATA, respectively.

3.2 Multiple Factor analysis and Regression Vector coefficients

The explained variance of MFA F1, F2 and F3 was 47.9%, 30% and 22.1 %, respectively (Figure 1). Contribution of groups (TP-DA; ST-CATA; C-CATA) in MFA was equal (approx. 33%) for the first two factors (F). For F3, C-CATA had the highest contribution, accounting to 43%, whereas TP-DA and ST- CATA accounted for 29% and 28%, respectively.

RV coefficients indicated a higher degree of similarity in samples' configurations between ST-CATA with TP-DA. Calculated RV coefficients for the first two MFA factors (77.9%) between methodologies were: 0.94 ($P < 0.001$) between TP-DA and ST-CATA; 0.79 ($P = 0.25$) between TP-DA and C-CATA; and 0.77 ($P = 0.33$) between ST-CATA and C-CATA.

For all 3 MFA factors, calculated RV coefficients revealed over 95% similarity in samples' configurations for all possible methodology combinations. Explicitly, the highest RV coefficient was 0.98 and was found between the C-CATA and TP-DA, yet this result was not significant since it only presented a trend ($P = 0.08$). The RV coefficient between ST- CATA with TP-DA and C-CATA was 0.95 ($P = 0.04$), for both combinations.

3.3 Discrimination and sensory profiles of samples

Sensory profiles and relative sample configurations were more similar between TP-DA and ST-CATA (Figure 2). Specifically, the sensory map of F1 vs. F2 of DA was very close to the map F1 vs. F3 of ST-CATA and vice versa (Figure 2). On the other hand, the C-CATA sensory maps differentiated from both TP-DA and ST-CATA, mostly in relation to meagre and greater amberjack (Figure 2).

The differentiation of C-CATA results for meagre and greater amberjack, when compared to both TP-DA and ST-CATA, can be attributed to several reasons. Firstly, according to C-CATA results, meagre and greater amberjack had a higher CATA count for seafood odour, seafood flavour and butter flavour when compared to the rest of the samples (Alexi et al., 2017). This resulted in these two samples being discriminated from pikeperch, which acquired a significantly lower CATA count for the respective attributes (Alexi et al., 2017) (Figure 2; Table 3). However, the same attributes were found non-discriminant ($P > 0.05$) between species in general, in both TP-DA and ST-CATA (Table 3). Moreover, whereas TP-DA and ST-CATA agreed that greater amberjack displayed a distinct sourness, when compared to the rest of the samples, for C-CATA sour taste had a really low CATA count and was not discriminant ($P > 0.05$) between samples (Table 3; Table 4). Additionally, turbid exudate, which discriminated greater amberjack from meagre in TP-DA and ST-CATA, was found insignificant for discrimination among samples in C-CATA (Figure 2; Table 3). Similarly to turbid exudate, laminar structure was also found insignificant for in-between species discrimination in C-CATA, whereas it was found discriminant in ST-CATA and TP-DA. However it should be mentioned that ST-CATA and TP-DA did not provide the exact same sample subgrouping for the aforementioned attribute (Table 3).

With respect to ST-CATA variations from TP-DA and C-CATA, it involved mainly attributes belonging to the odour and texture modalities (Table 3). For the odour modality, ¾ attributes varied for the ST. Specifically, the ST panel found sardine aroma discriminant, whereas no significant variations in this attribute existed according to TP-DA and C-CATA results. Moreover, TP-DA and C-CATA agreed on the discrimination and evaluation of butter and earthy odour. On the other hand, ST-CATA results indicated that the aforementioned attributes were insignificant for between species discrimination (Table 3). Regarding texture, ST-CATA disagreed with both TP-DA and C-CATA in its evaluation of juiciness, since this attribute was found non-discriminant for ST, whereas significant variations between species existed according to TP-DA and C-CATA (Table 3). Pasty was the second texture attribute that varied between methods since it was found non-significant in TP-DA, whereas it was significant for both ST- and C-CATA. However, the evaluation of the aforementioned attribute varied even between ST- and C-CATA, indicating a general non-agreement between methods. Despite the mentioned variations, the results for pikeperch and wreckfish were not majorly affected, since the samples' sensory profiles were relatively similar across all three methodologies (Figure 2; Table 3).

3.4 Attribute ranges and citation frequencies

To investigate whether the CATA task provided the same range of differences between samples with DA intensity measurements, the normalized maximum range between samples for an attribute was calculated in all 3 methodologies (eq. 1-2). Moreover, to examine the effect of training on the CATA task, the total citation frequency (%) of the attributes was calculated for both ST- and C-CATA (eq. 3). According to TP-DA results, the biggest intensity differences, within the same attribute, existed within the appearance, taste, texture modalities. Specifically, only 5 attributes had a normalised maximum intensity range greater than 30%, 3 of which belonged to appearance modality (Table 4).

Examining the results of attributes per sensory modality, more similarities existed for odour between C-CATA and TP-DA, with 3 out of 4 attributes (butter, sardine and earthy) exhibiting the same range. On the other hand ST-CATA did not agree with either of the aforementioned methods (Table 4). For the appearance modality, comparisons of ranges were not applicable for 3/6 attributes, since they were not common between CATA and DA methods (Table 1). For the taste and flavour modality, more similarities were found between TP-DA and ST-CATA (Table 4). Specifically, sour taste was found as highly discriminant for greater amberjack (Table 3; Table 4) and the maximum difference (eq. 1-2) between samples measured was approximately 30% in both TP-DA and ST-CATA. Yet, C-CATA was not in line with the other two methods, since variation among samples for sourness was minor and insignificant (Table 4). Moreover, according to C-CATA, two flavour attributes, butter and seafood, were found to significantly differ ($P < 0.05$), a result which did not agree with TP-DA and ST-CATA (Table 3; Table 4). These attributes displayed a bigger maximum range (eq. 1-2) between samples in C-CATA than they did in the other two methods. However, the actual difference between the C-CATA method with TP-DA (3.1%) and ST-CATA (2.2%) was relatively small for the butter flavour (Table 4). For texture, the majority of attributes displayed magnified ranges between samples (eq. 1-2) in both CATA methods when compared to DA ones (Table 4).

Regarding the attributes' total citation frequency in ST-CATA, with the exception of seafood odour and flavour, it was either similar (±5%) or higher than in C-CATA. Specifically, 1/4 odour, 6/9 appearance, 2/2 taste, 1/4 flavour and 5/6 texture attributes had a 10% increase or more in their citation frequency in ST-CATA, when compared with C-CATA (Table 4). However, the increase in citation frequency did not translate into a better discrimination, since the majority of the aforementioned attributes were discriminant in both ST- and C-CATA (Table 3). The only attributes that had an increased citation frequency in ST-CATA and were only discriminant in ST when compared to C-CATA were sour taste, turbid exudate and laminar structure (Table 3; Table 4).

4 Discussion

4.1 Level of explained sensory variations per method

A multivariate technique (PLSR) was chosen to compare between the explained variances of three methodologies, since it is capable of separating the information from the noise within a sensory dataset (Rinnan et al., 2015). Noise is a common limitation of datasets originating from untrained or partly trained individuals, yet hard to detect in the absence of replicated measurements (Valentin et al., 2012; Varela & Ares, 2012). However, including noise as meaningful information, can lead to overestimation of the degree of explained variations. Thus, the PLSR approach is highly appropriate for comparing the explained variances between methodologies, since no replicates were used for ST-CATA and C-CATA.

The proportion of explained sensory variation in TP-DA was higher, when compared to ST and C-CATA. This complies with the literature, where loss of quantitative information has been mentioned as one of the main limitations of CATA compared to DA, especially when highly similar products are profiled (Dooley, Lee, & Meullenet, 2010; Giacalone et al., 2013; Varela & Ares, 2012). This loss is mainly attributed to the CATA constraints in evaluating intensity differences for the same sensory attribute (Lazo et al., 2016). Besides the method constraints, noise in measurements due to lack or limited training of subjects is addressed via increase in panel size (Ares, Tárrega, et al., 2014; Varela & Ares, 2012). Thus, a small panel size would be expected to lead to less meaningful, more noisy and unstable results. Yet, the reduction of the ST-CATA panel size in half, when compared to C-CATA, did not result in additional loss of sensory information. On the opposite, ST-CATA had a small increase in its explained variances when compared to C-CATA. This is an indication of the positive effect of training on the required panel size for CATA.

This positive effect is partly reflected in the higher citation frequency of some attributes in ST-CATA (Table 4). Yet, a mere increase in the citation frequencies alone is not adequate by itself to justify the similar explained variances. Indeed, only three attributes with increased citation frequencies were discriminant in ST- and not in C-CATA (Table 3; Table 4). Other attributes were either discriminant in both ST and C-CATA, or the opposite applied (Table 3; Table 4). However, taking into account that the measurements gathered were much lower for the ST-CATA, due to the smaller panel size, there is an indication that the results obtained via ST-CATA were also more consistent than those of C-CATA, balancing out the potential negative effect of panel size decrease.

4.2 Similarity in samples' configurations across methods

To compare the configurational similarity of samples between methods, the RV coefficients were calculated for both 2 (78% retained variability) and 3 (100% retained variability) MFA factors. For the first two MFA factors a higher degree of similarity in samples configurations was evident between ST-CATA with TP-DA. Taking into account all 3 MFA factors, samples' configurations were similar between all three methodologies (RV $\geq$ 0.95). Yet, drawing our conclusions based on all 3 MFA factors has several limitations. Firstly, the contribution of C-CATA on F3 of MFA was approximately 50%, limiting the contributions of the two other methodologies. Moreover, the explicitly high RV coefficient between TP-DA and C-CATA was found insignificant, since it presented only a trend (P = 0.08).

Thus, the results so far indicate that the highest similarity existed between TP-DA and ST-CATA. This is furthermore supported by the individual sensory maps of the methodologies presented in Figure 2. These findings reveal that not only the training in ST-CATA had a positive effect on the explained variability via the reduction of noise, but it also altered the performance of consumers, bringing results closer to TP-DA, when compared to C-CATA. Among other reasons, training could have improved the ST- performance due to the use of physical references to define the technical terms included in the ballot (namely: large/little exudate, turbid/transparent exudate, 'laminar structure', crumbliness, and teeth adherence). However, since technical terms are not commonly used with untrained subjects, the effect of training in the evaluation of common and technical terms is separately discussed hereafter (Van Trijp & Schifferstein, 1995).

4.3 Samples' description: a comparison between methods

Since both ST and C panels used the same evaluation method and received the same written attribute definitions, the current inability of C-CATA to provide similar profiles to TP-DA and ST-CATA for the species meagre and greater amberjack could be attributed to the lack of training with physical references. Examining the description of the aforementioned samples more closely, the variation in C-CATA profiling can be attributed to various reasons (Table 3).

Firstly, consumers exhibited an inability to identify differences in common attributes belonging to the taste and flavour modalities. Specifically, sour taste which was a highly discriminant attribute for greater amberjack, according to both TP-DA and ST-CATA, did not vary significantly between samples in C-CATA. The lack of sourness discrimination can be connected to the reluctance of the C-panel in using the term, which was expressed in its very low citation frequency (2.1%) and can be attributed to the a confusion of the adjectives sour-bitter (O'Mahony, Goldenberg, Stedmon, & Alford, 1979). Indeed, inability of consumers to identify differences in a discriminant taste attribute has been also reported for bitter in beers (Giacalone et al., 2013). According to O'Mahony et al. (1979), this confusion can be surpassed after clarification of the term via definition with reference standards. Indeed, training with physical references increased the sensitivity of the ST on this taste attribute. This is indicating that even for terms that are common, such as sourness, a definition and the context in which an attribute is evaluated is crucial for correct evaluation. Moreover, the short familiarization with attributes' definition via reference material increased the ST panel's capacity in identifying flavour variations similarly to TP-DA (Figure 2;

Table 3). On the other hand, C-CATA attributed flavour characteristics to meagre and greater amberjack, which were non-discriminant for the samples, leading to a different pattern of sample associations in general, when compared to ST-CATA and TP-DA (Figure 2).

The training with physical references that preceded ST-CATA had also a positive effect on discrimination and identification of more complex technical attributes, turbid exudate and 'laminar structure'(Figure 2; Table 3). On the contrary, these two attributes had amongst the lowest citation frequencies within the appearance modality and were found insignificant for discrimination among the species in C-CATA. The difficulties consumers face in the evaluation of complex attributes has been previously described, suggesting the need of physical references for identification of such terms (Ares et al., 2015; Giacalone et al., 2013; Moussaoui & Varela, 2010). Indeed, specialized terminology is not appropriate for consumers, since they need to relate to the attribute they are evaluating (Van Trijp & Schifferstein, 1995). Thus, training not only increased the citation frequency of ST-CATA attributes, but it did so in a meaningful way, according to TP-DA. However, some additional training may have been required, especially on the 'laminar structure" attribute, since even the ST-panel did not acquire the exact same subgroupings with TP-DA. The positive effect of short training prior to evaluation has been shown also for other fast methodologies, which usually include no training step, such as napping (Liu, Grønbeck, Di Monaco, Giacalone, & Bredie, 2016).

On the other hand, untrained consumers faced difficulties in recognizing and evaluating attributes that they were unfamiliar with or uncertain of. The only sensory modalities which were not affected by the lack of training in C-CATA were odour and texture. Texture has been found as the most discriminant modality for consumers in fish, in the absence of off-odour/ flavours, explaining the high citation frequencies and good performance of C-CATA subjects even for technical attributes (teeth adherence and crumbliness) belonging to this modality (Nielsen, Hyldig, & Larsen, 2002; Wesson, Lindsay, & Stuiber, 1979). Pasty texture, created an exception since the results for this attribute varied in general across all 3 methods, indicating a general difficulty in its evaluation, independent of the type of definition provided to consumers (written definition vs. references) (Table 4). Thus in overall the written definition the C-panel received was adequate to acquire a correct discrimination among species for texture in the majority of cases. This is indicating that the training required prior to the evaluation of a specific attribute depended on several factors besides complexity, including sensory modality and the context in which an attribute is evaluated.

Whereas ST-CATA and TP-DA shared a high degree of similarity in samples' description and configuration, the quantitative differences between samples for the same attribute (normalized maximum range, eq. 1-2), was altered in both ST-CATA and C-CATA (Table 4). This was true specifically in appearance and texture modalities, where the differences between samples were larger with the CATA method, when compared to TP-DA. The fact that CATA frequencies cannot substitute for DA measurements, but consist only a relative measure of intensity has been described previously (Ares et al., 2015). Current results indicate that this seems to be connected rather to the method of evaluation than to the training that panel received. Yet, as similarly suggested for a rating variant of CATA, Rate-All-That-Apply (RATA), the comprehension of the CATA task by participants and how it affects the resulting measurements, should also be investigated (Oppermann, de Graaf, Scholten, Stieger, & Piqueras-Fiszman, 2017). Moreover, specifically for the evaluation of appearance attributes, it should be taken into account that the containers used in ST- and C- CATA created a higher contrast (transparent for TP-DA and black for ST-/C-CATA). This may partly explain the differences found in this modality between TP-DA and ST-/C-CATA, especially in terms of maximum range of difference between samples (Table 4).

Additionally, it is important to mention that several factors such as linguistic equivalence, data collection style and differences in response style of participants can affect cross-cultural comparisons (Ares, 2016). However, in the current study several measures were taken to reduce variations due to cross-cultural differences. These measures included: expert translation of terms along with the use of the same definitions; use of similar reference material between the ST-CATA and TP-DA; and randomization of the list in the ballot for C-CATA (Table 2). Besides, the Spanish study involved a trained panel (TP-DA) and training with use of reference scales for attribute quantification. Moreover, since the studies that involved semi-trained and untrained participants were both conducted in Greece, no major cultural interference is expected. It should be mentioned, though that while no data on the educational level of participants were gathered, the fact that the majority of the ST panel were HCMR employees could have resulted in an additional advantage, due their possible familiarity with the definition of some technical terms common for describing fish.

Finally, we want to underline that whereas evaluating hot served fish samples can be considered as a complex task, more research is needed into the possible effect of large sample sets, with variable levels of complexity on the outcomes of the method.

5 Conclusions

One hour training changed the performance of consumer subjects bringing the results of CATA closer to descriptive analysis (DA) than CATA elicited data from untrained individuals. Specifically, a high configurational similarity of samples as well as a similar sample description existed between DA and semi-trained CATA. Consumer CATA also shared an overall high configurational similarity with DA, yet they differed in qualitative description of some of the samples. The consumer differentiation was mainly attributed to variations in flavour description of samples, insensitivity in taste differences and difficulties in evaluation and discrimination of more complex appearance attributes when compared to the semi-trained and trained panels. Moreover, training increased the citation frequency of the majority of CATA ballot terms which can be a useful measure to increase the amount of overall answers (ticks) gathered in panels that have a low amount of participants. Thus, the introduction of a short training not only increased the similarity of results to DA but also lowered the amount of participants required to acquire a reliable sensory profile from CATA. This is suggesting that the semi-trained CATA variation is a valuable research tool when a trained panel cannot be sustained and a reliable, time- and cost-efficient sensory profiling of samples in needed. Yet, it should be noted that whereas the profiling of the samples is really similar, CATA derived sample differences for the same sensory attribute should be carefully interpreted, since they do not always represent intensity differences.

Acknowledgments

All tasks were performed within the frames of DIVERSIFY project. This project is funded from the European Union's Seventh Framework Program for research, technological development and demonstration (KBBE-2013-07 single stage, GA 603121, DIVERSIFY). We also acknowledge the financial support from Department of Food Science, Aarhus University, Denmark.

References

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In, *7th Pangborn sensory science symposium*.

Alexi, N., Byrne, D. V., Nanou, E., & Grigorakis, K. (2017). Investigation of sensory profiles and hedonic drivers of emerging aquaculture fish species. *Journal of Science of Food and Agriculture, Accepted*.

Ares, G. (2015). Methodological challenges in sensory characterization. *Current Opinion in Food Science, 3*, 1-5.

Ares, G. (2016). Methodological issues in cross-cultural sensory and consumer research. *Food Quality and Preference*http://dx.doi.org/10.1016/j.foodqual.2016.10.007.

Ares, G., Antúnez, L., Bruzzone, F., Vidal, L., Giménez, A., Pineau, B., et al. (2015). Comparison of sensory product profiles generated by trained assessors and consumers using CATA questions: Four case studies with complex and/or similar samples. *Food Quality and Preference*, *45*, 75-86.

Ares, G., Antúnez, L., Giménez, A., Roigard, C. M., Pineau, B., Hunter, D. C., et al. (2014). Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers. *Food Quality and Preference, 36*, 111-121.

Ares, G., Etchemendy, E., Antúnez, L., Vidal, L., Giménez, A., & Jaeger, S. R. (2014). Visual attention by consumers to check-all-that-apply questions: Insights to support methodological development. *Food Quality and Preference*, *32*, 210-220.

Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference, 28*(1), 141-153.

Ares, G., Jaeger, S. R., Bava, C. M., Chheang, S. L., Jin, D., Gimenez, A., et al. (2013). CATA questions for sensory product characterization: Raising awareness of biases. *Food Quality and Preference, 30*(2), 114-127.

Ares, G., Tárrega, A., Izquierdo, L., & Jaeger, S. R. (2014). Investigation of the number of consumers necessary to obtain stable sample and descriptor configurations from check-all-that-apply (CATA) questions. *Food Quality and Preference*, *31*, 135-141.

Ares, G., Varela, P., Rado, G., & Giménez, A. (2011). Are consumer profiling techniques equivalent for some product categories? The case of orange-flavoured powdered drinks. *International journal of food science & technology*, *46*(8), 1600-1608.

Byrne, D. V., O'Sullivan, M. G., Dijksterhuis, G. B., Bredie, W. L. P., & Martens, M. (2001). Sensory panel consistency during development of a vocabulary for warmed-over flavour. *Food Quality and Preference, 12*(3), 171-187.

Cruz, A., Cadena, R., Castro, W., Esmerino, E., Rodrigues, J., Gaze, L., et al. (2013). Consumer perception of probiotic yogurt: Performance of check all that apply (CATA), projective mapping, sorting and intensity scale. *Food Research International*, *54*(1), 601-610.

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.

Giacalone, D., Bredie, W. L., & Frøst, M. B. (2013). “All-In-One Test”(AI1): A rapid and easily applicable approach to consumer product testing. *Food Quality and Preference, 27*(2), 108-119.

Giacalone, D., & Hedelund, P. I. (2016). Rate-all-that-apply (RATA) with semi-trained assessors: An investigation of the method reproducibility at assessor-, attribute-and panel-level. *Food Quality and Preference, 51*, 65-71.

Hough, G. (1998). Experts versus consumers: a critique. *Journal of sensory studies, 13*(3), 285-289.

ISO. (2007). ISO Standard 8589:2007 Sensory analysis -- General guidance for the design of test rooms. In.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2013). SensMixed: Mixed effects modelling for sensory and consumer data. In, *R package version*.

Kuznetsova, A., Christensen, R. H., Bavay, C., & Brockhoff, P. B. (2015). Automated mixed ANOVA modeling of sensory and consumer data. *Food Quality and Preference, 40*, 31-38.

Lazo, O., Claret, A., & Guerrero, L. (2016). A Comparison of Two Methods for Generating Descriptive Attributes with Trained Assessors: Check-All-That-Apply (CATA) vs. Free Choice Profiling (FCP). *Journal of sensory studies, 31*(2), 163-176.

Lazo, O., Guerrero, L., Alexi, N., Grigorakis, K., Claret, A., Pérez, J. A., et al. (2017). Sensory characterization, physico-chemical properties and somatic yields of five emerging fish species. *Food Research International*https://doi.org/10.1016/j.foodres.2017.07.023.

Liu, J., Grønbeck, M. S., Di Monaco, R., Giacalone, D., & Bredie, W. L. (2016). Performance of Flash Profile and Napping with and without training for describing small sensory differences in a model wine. *Food Quality and Preference, 48*, 41-49.

MacFie, H. J., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of sensory studies, 4*(2), 129-148.

Martens, M., Bredie, W. L. P., & Martens, H. (2000). Sensory profiling data studied by partial least squares regression. *Food Quality and Preference, 11*(1–2), 147-149.

Meyners, M., & Castura, J. C. (2014). Check-All-That-Apply Questions. In, *Novel Techniques in Sensory Characterization and Consumer Profiling*: CRC Press.

Moussaoui, K. A., & Varela, P. (2010). Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference, 21*(8), 1088-1099.

Murray, J., Delahunty, C., & Baxter, I. (2001). Descriptive sensory analysis: past, present and future. *Food Research International, 34*(6), 461-471.

Nielsen, J., Hyldig, G., & Larsen, E. (2002). 'Eating Quality'of Fish—A Review. *Journal of aquatic food product technology, 11*(3-4), 125-141.

O'Mahony, M., Goldenberg, M., Stedmon, J., & Alford, J. (1979). Confusion in the use of the taste adjectives 'sour' and 'bitter'. *Chemical Senses, 4*(4), 301-318.

Oppermann, A., de Graaf, C., Scholten, E., Stieger, M., & Piqueras-Fiszman, B. (2017). Comparison of Rate-All-That-Apply (RATA) and Descriptive sensory Analysis (DA) of model double emulsions with subtle perceptual differences. *Food Quality and Preference, 56*, 55-68.

Reinbach, H. C., Giacalone, D., Ribeiro, L. M., Bredie, W. L., & Frøst, M. B. (2014). Comparison of three sensory profiling methods based on consumer perception:

CATA, CATA with intensity and Napping®. *Food Quality and Preference, 32*, 160-166.

Rinnan, Å., Giacalone, D., & Frøst, M. B. (2015). Check-all-that-apply data analysed by Partial Least Squares regression. *Food Quality and Preference, 42,* 146-153.

Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied statistics*, 257-265.

Valentin, D., Chollet, S., Lelievre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International journal of food science & technology, 47*(8), 1563-1578.

Van Trijp, H., & Schifferstein, H. (1995). Sensory analysis in marketing practice: comparison and integration. *Journal of sensory studies, 10,* 127-147.

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.

Wesson, J., Lindsay, R. C., & Stuiber, D. A. (1979). Discrimination of fish and seafood quality by consumer populations. *Journal of Food Science, 44*(3), 878-882.

List of figure captions

Figure 1: Consensus MFA map with superimposed partial points from different methods (●) on the consensus MFA point (■). TP-DA: Descriptive analysis with a trained panel; ST-CATA: CATA evaluation with a semi-trained panel; C-CATA: CATA evaluation with consumers. Total retained variability for all 3 factors = 100%.

Figure 2: Principal Component analysis (A) and Correspondence Analysis (B, C) plots illustrating fish samples (●) and significant (P <0.05) sensory attributes (◆) for Descriptive analysis with a trained panel (A), CATA with a semi-trained panel (B) and CATA with a consumer panel (C); suffixes: Od. (Odour), Fl. (flavour). The C plot is taken from Alexi et al. (2017)

Tables:

Table 1: The 22 and 25 attribute lists used in descriptive analysis (DA) and CATA ballots, respectively, along with attribute references provided to semi-trained participants during training; references are given per serving where applicable. The references used in DA are available in Lazo et al. (2017)

DA attribute list	CATA ballot attributes	Attribute references for semi-trained panel
1. Butter odour	1. Butter odour	10g halibut[1] paste mixed with 1g melted butter
2. Seafood odour	2. Seafood odour	Content of one red crab claw[1]
3. Sardine odour	3. Sardine odour	10g of gilthead sea bream[1] paste mixed with1ml of cod liver oil
4. Earthy odour	4. Earthy odour	Wet soil
5. Colour		Tissue appearance
	5. Brown colour[2]	Images with brown colour gradients (light brown; dark beige)
	6. White colour[2]	Images with white colour gradients (white, ivory, beige)
6. Colour uniformity	7. Colour uniformity	Tissue appearance: image of Cat fish *Ictalurus furcatus*[1]
7. Exudate		Exudate amount
	8. Little/No exudate[2]	Image of 0.2ml of turbid solution[3] in a container with a cooked piece of sword fish (4cm x 3cm)
	9. Large Exudate[2]	Image of 5ml of turbid solution[3] in a container with a cooked piece of sword fish (4cm x 3cm)
8. Turbidity		Exudate turbidity
	10. Turbid[2]	Image of 10% milk in water
	11. Transparent[2]	Image of plain water
9. Fat droplets	12. Fat droplets	Exudate appearance: Image of 15 µl of oleoresin colorant diluted in a turbid solution[3]
10. laminar structure	13. laminar structure	Tissue appearance: image of tuna[1]
11. Sour taste	14. Sour taste	10gr of gilthead sea bream[1] paste mixed with 1ml citric acid solution 1:10
12. Bitter taste	15. Bitter taste	10g of gilthead sea bream[1] paste + 1ml proteolitic enzyme[4] solution (1:1)

13. Butter flavour	16. Butter flavour	10g halibut[1] paste mixed with 1gr melted butter
14. Seafood flavour	17. Seafood flavour	Content of one red crab claw[1]
15. Boiled vegetable flavour	18. Boiled vegetable flavour	0.6g of gilthead sea bream[1] paste mixed with 0.4g of boiled green beans + potato (1:1)
16. Earthy flavour	19. Earthy	Wet soil
17. Firmness	20. Firm	Canned mackerel
18. Crumbliness	21. Crumbly	Halibut[1]
19. Juiciness	22. Juicy	Salmon cooked for 15 min in 110°C
20. Chewiness	23. Chewy	Swordfish[1]
21. Pastiness	24. Pasty	Salmon cooked for 15 min in 110°C
22. Teeth adherence	25. Teeth adherence	Salmon[1]

[1]Cooked in oven for 20 min in 110°C

[2]Attributes used in CATA ballots as replacement of original DA attributes, colour; exudate and turbidity

[3]30% milk in water

[4]Enzyme mix Delvolase DSM

Table 2: Summary of three sensory methodologies; a detailed description of TP-DA and C-CATA can be found in Lazo et al. (2017) and Alexi et al. (2017), respectively

	Trained Panel (TP)	Semi trained panel (ST)	Consumer panel (C)
Method	Descriptive analysis (DA)	Check-all-that-apply (CATA)	Check-all-that-apply (CATA)
Number of assessors	8	37	70
Previous experience of assessors	2-3 years of experience in sensory profiling	None	None
Vocabulary development	Yes	No[1]	No[1]
Number of attributes	22	25[2]	25[2]
Training duration	30 hrs.	1 h.	No training
References	Physical references corresponding to 3 parts of the scale (low, medium, high)	Physical references	No references
Instructions of CATA task	-	During training[3]	Prior to evaluation[3]
Attribute definitions	-	Written definitions[4]	Written definitions[4]
Attribute order in CATA	-	Fixed within modalities	Randomized within modalities[5]
Sample presentation	Randomized; monadic	Randomized; monadic	Randomized; monadic
Sample evaluation	5 replicates (5 sessions)	No replicates (1 session)	No replicates (1 session)
Total duration	Approx. 35hrs.	Approx. 2hrs	Approx. 1hr
Institute, country	IRTA, Spain	HCMR, Greece	Greece

[1]The sensory vocabulary was the same across both the semi-trained and consumer CATA ballot and was an adaptation of the one used in DA

[2]The total number of attributes differs between DA and CATA, since for 3 of the appearance attributes the end points of the scale in DA were used as separate individual CATA attributes; all other attributes were exactly the same as in DA

[3]The same instructions were given both to semi-trained and consumer panels

[4]The same definitions for all attributes, with the exception of sour and bitter which were considered self-explanatory, were given both to semi-trained and consumer panels. Definitions are available in Alexi et al. (2017)

[5](Ares, Antúnez, et al., 2014; Ares, Etchemendy, et al., 2014; Ares & Jaeger, 2013)

Table 3: Significance level and pairwise comparisons for sensory terms, which were significantly different (*P*-Level < 0.05) in at least one out of three methodologies. Within an attribute and method different letters denote significant differences among the samples (*P* <0.05). For methods, TP-DA, ST-CATA and C-CATA stands for Trained Panel –Descriptive Analysis, Semi Trained panel –Check-All-That-Apply and Consumer –Check-All-That-Apply. For samples GA, M, P and W stand for greater amberjack, meagre, pikeperch and wreckfish, respectively.

		P^2	Post-hoc[3]						P^2	Post-hoc[3]			
Attributes	**Method**	**Level**	**GA**	**M**	**P**	**W**	**Attributes**	**Method**	**Level**	**GA**	**M**	**P**	**W**
Butter	TP-DA	**	ab	a	b	a	laminar	TP-DA	**	a	b	b	b
odour	ST-CATA	ns					structure	ST-CATA	**	ab	b	b	a
	C-CATA	*	a	ab	b	a		C-CATA	ns				
Seafood	TP-DA	ns					Sour taste	TP-DA	***	a	b	b	b
odour	ST-CATA	ns						ST-CATA	**	a	b	b	b
	C-CATA	*	a	a	b	ab		C-CATA	ns				
Sardine	TP-DA	ns					Butter	TP-DA	ns				
odour	ST-CATA	**	ab	b	b	a	flavour	ST-CATA	ns				
	C-CATA	ns						C-CATA	*	ab	a	b	ab
Earthy	TP-DA	***	b	b	a	b	Seafood	TP-DA	ns				
odour	ST-CATA	ns					flavour	ST-CATA	ns				
	C-CATA	*	ab	ab	a	b		C-CATA	**	a	a	b	ab
Colour	TP-DA	***	b	b	c	a	Earthy	TP-DA	*	bc	c	a	bc
Brown	ST-CATA	***	b	a	b	a	flavour	ST-CATA	**	b	b	a	b
colour[1]	C-CATA	**	ab	a	b	ab		C-CATA	**	ab	b	a	ab
White	ST-CATA	***	a	b	a	b	Firm	TP-DA	***	b	c	c	a
colour[1]	C-CATA	**	ab	b	a	b	texture	ST-CATA	***	b	b	b	a
Colour	TP-DA	***	a	b	a	b		C-CATA	***	b	b	b	a
uniformity	ST-CATA	***	a	b	a	b	Chewy	TP-DA	***	a	b	b	a
	C-CATA	**	ab	b	a	b	texture	ST-CATA	*	b	b	b	a
Exudate	TP-DA	***	b	c	d	a		C-CATA	***	ab	bc	c	a
Large	ST-CATA	**	b	ab	b	a	Juicy	TP-DA	*	ab	a	ab	b
exudate[1]	C-CATA	***	b	b	b	a	texture	ST-CATA	ns				
Little/No	ST-CATA	**	a	ab	a	b		C-CATA	***	a	a	a	b

exudate[1]	C-CATA	***	ab	a	a	b	Crumbly	TP-DA	***	b	a	a	c
Turbidity	TP-DA	***	c	b	a	b		ST-CATA	**	ab	ab	a	b
Turbid	ST-CATA	**	b	a	ab	a		C-CATA	***	bc	a	ab	c
exudate[1]	C-CATA	ns					Pasty	TP-DA	ns				
Transparent	ST-CATA	**	a	b	ab	b	texture	ST-CATA	**	a	a	ab	b
exudate[1]	C-CATA	***	a	ab	b	a		C-CATA	***	b	ab	a	b
Fat droplets	TP-DA	***	a	a	b	a	Teeth	TP-DA	***	a	b	b	b
	ST-CATA	***	a	a	b	a	adherence	ST-CATA	***	a	b	b	b
	C-CATA	***	a	a	b	a		C-CATA	***	a	b	b	b

[1]Alternate attributes used in the CATA ballots as replacement of colour, exudate and turbidity of DA

[2]For DA significance was obtained by Mixed model ANOVA in the SensMixed package in R, for CATA by Conchran's Q test in XLSTAT®; ns: non-significant; *p<0.05; **p<0.01; ***p<0.001

[3]For DA post hoc groups were computed by pairwise comparisons in SensMixed package in R and for CATA data using the McNemar (Bonferroni) approach in XLSTAT®

Table 4: Normalized maximum range of difference between samples (%) for all attributes used in the 3 different methodologies and citation frequencies (%) for attributes used in the CATA-ballots. TP-DA, ST-CATA and C-CATA stands for Trained Panel –Descriptive Analysis, Semi Trained panel –Check-All-That-Apply and Consumer –Check-All-That-Apply

Attributes	Normalized maximum range (%)[2]			Citation frequency (%)[3]	
	TP-DA	ST-CATA	C-CATA	ST-CATA	C-CATA
Butter odour	20.5	8.1	18.6	48	30
Seafood odour	6	13.5	22.9	45.3	51.1
Sardine odour	11.9	24.3	10	13.5	15.7
Earthy odour	17.6	10.8	18.6	12.8	15.7
Colour	25.4				
White[1]		40.5	27.1	72.3	72.5
Brown[1]		37.8	20	23	13.6
Colour uniformity	15.9	45.9	27.1	55.4	43.2
Exudate	36.1				
Large[1]		45.9	51.4	52	28.9
Little/No[1]		37.8	41.4	43.2	41.8
Turbidity	66.3				
Turbid[1]		29.7	17.1	40.5	17.1
Transparent[1]		35.1	34.3	52.7	38.2
Fat droplets	51	54.1	41.4	52.7	36.1
laminar structure	18	29.7	12.9	35.1	20.7
Sour taste	33.9	29.7	2.9	20.3	2.1
Bitter taste	2.3	8.1	5.7	20.3	6.1
Butter flavour	12.6	13.5	15.7	35.8	22.5
Seafood flavour	8.1	18.9	25.7	33.8	47.9
Boiled vegetable flavour	11.4	2.7	7.1	19.6	15.7
Earthy flavour	21	32.4	15.7	16.9	21.8
Firm texture	23	56.8	52.9	37.2	19.6
Crumbly texture	26.4	32.4	34.3	51.4	37.1
Juicy texture	11.3	18.9	40	54.1	28.2
Chewy texture	26.1	43.2	37.1	42.6	34.6
Pasty texture	9.3	32.4	18.6	25.7	13.6
Teeth adherence	38.9	56.8	44.3	44.6	33.2

[1]Alternate attributes used in the CATA ballots as replacement of colour, exudate and turbidity of DA

[2]For DA, maximum range was calculated according to equation 1 and for CATA using equation 2

[3]Total citation frequency was calculated using equation 3

Figure 1

Click here to download high resolution image

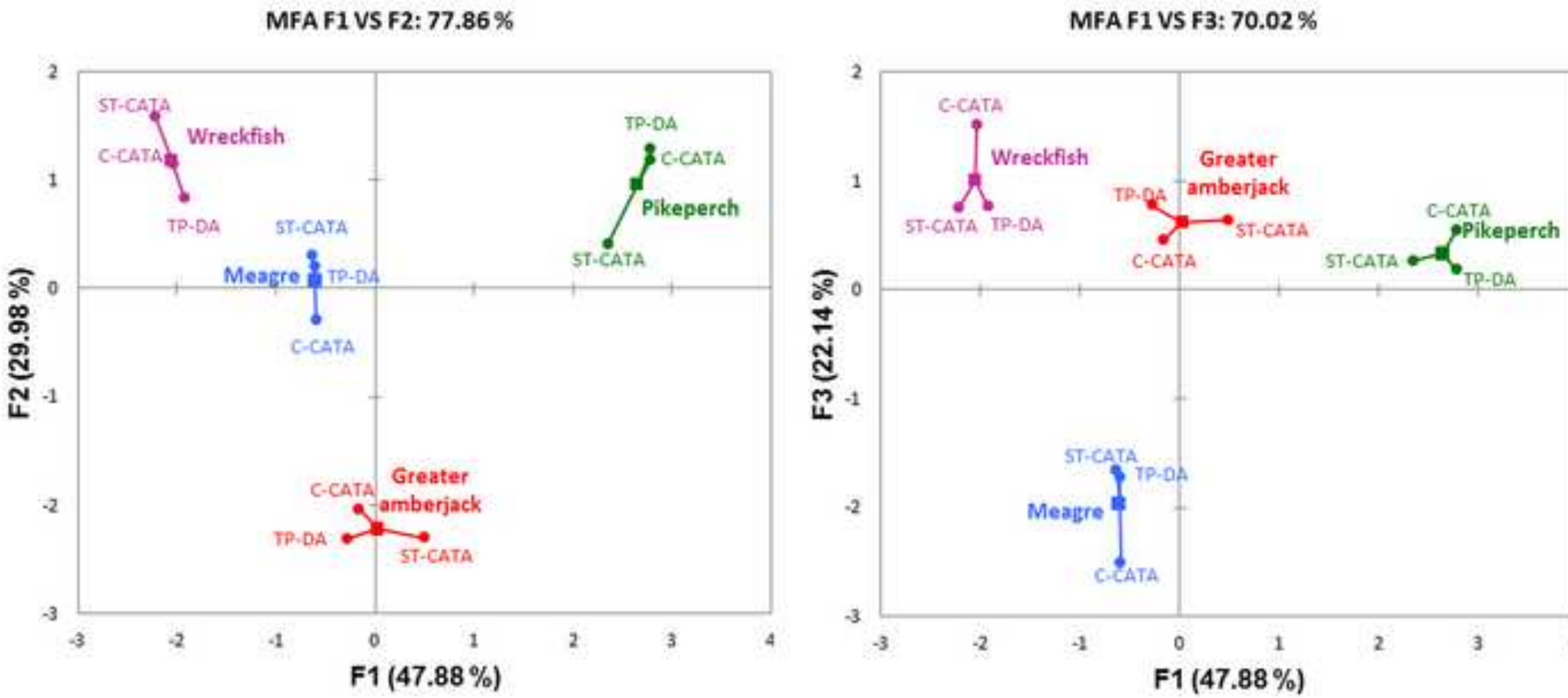

Figure 2

Click here to download high resolution image