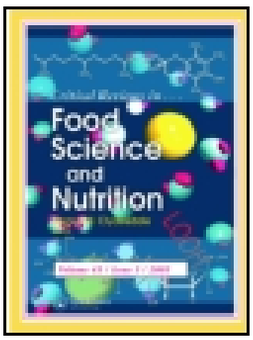

Critical Reviews in Food Science and Nutrition

ISSN: 1040-8398 (Print) 1549-7852 (Online) Journal homepage: http://www.tandfonline.com/loi/bfsn20

Validation of rapid descriptive sensory methods against conventional descriptive analyses: A systematic review

Lorena Andrade de Aguiar, Lauro Melo & Lívia de Lacerda de Oliveira

To cite this article: Lorena Andrade de Aguiar, Lauro Melo & Lívia de Lacerda de Oliveira (2018): Validation of rapid descriptive sensory methods against conventional descriptive analyses: A systematic review, Critical Reviews in Food Science and Nutrition, DOI: 10.1080/10408398.2018.1459468

To link to this article: https://doi.org/10.1080/10408398.2018.1459468

CRITICAL REVIEWS IN FOOD SCIENCE AND NUTRITION
https://doi.org/10.1080/10408398.2018.1459468

Taylor & Francis
Taylor & Francis Group

Validation of rapid descriptive sensory methods against conventional descriptive analyses: A systematic review

Lorena Andrade de Aguiar[a], Lauro Melo [a], and Lívia de Lacerda de Oliveira[b]

[a]College of Health Sciences, University of Brasilia, Campus Darcy Ribeiro, Asa Norte, Brasilia, Brazil; [b]Department of Biochemical Engineering, School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ABSTRACT

A major drawback of conventional descriptive profile (CDP) in sensory evaluation is the long time spent in panel training. Rapid descriptive methods (RDM) have increased significantly. Some of them have been compared with CDP for validation. In Health Sciences, systematic reviews (SR) are performed to evaluate validation of diagnostic tests in relation to a gold standard method. SR present a well-defined protocol to summarize research evidence and to evaluate the quality of the studies with determined criteria. We adapted SR protocol to evaluate the validation of RDM against CDP as satisfactory procedures to obtain food characterization. We used "Population Intervention Comparison Outcome Study – PICOS" framework to design the research in which "Population" was food/ beverages; "intervention" were RDM, "Comparison" was CDP as gold standard, "Outcome" was the ability of RDM to generate similar descriptive profiles in comparison with CDP and "Studies" was sensory descriptive analyses. The proportion of studies concluding for similarity of the RDM with CDP ranged from 0% to 100%. Low and moderate risk of bias were reached by 87% and 13% of the studies, respectively, supporting the conclusions of SR. RDM with semi-trained assessors and evaluation of individual attributes presented higher percentages of concordance with CDP.

KEYWORDS

Systematic review; conventional descriptive profile; rapid descriptive test; accuracy; sensory evaluation

1. Introduction

Among the sensory methods, descriptive analysis stands out as a powerful instrument to provide a description of qualitative and quantitative aspects of human perception (Lawless and Heymann 2010; Stone, 2012). Stone et al. (1974) developed Quantitative Descriptive Analysis (QDA) as a descriptive method to integrate flavor and texture aspects for a more complete profile. A major drawback of QDA stands in the long time spent for the training, besides the difficulty to validate the panel when the samples are not as standardized as processed food, such as when the goal is to characterize an exotic fruit whose production is still small-scaled (Pineli et al. 2015; Pineli et al. 2014). In this case, samples can be heterogeneous and impair repeatability along the time. In order to overcome the difficulties related to training process and resources, rapid descriptive methods have been developed. QDA can be considered a "gold standard" method to identify and quantify sensory attributes in food, beverages and other materials, once the skills of the assessors are validated statistically to assure ability to discriminate among treatments, repeatability and coherence with the panel (Stone and Sidel 2004). In spite of some adaptations from original QDA over the years, a method can provide a conventional descriptive profile (CDP) when the basis of assessors training and panel's statistical validation remains, as well as the quantitative description of sensory attributes (Silva et al. 2012; Valentin et al. 2012).

The use of rapid methods increased significantly in the last few years, ranging from methods based on the evaluation of individual attributes such as Free Choice Profile (FCP), Intensity Scales (IS), Check-all-that-apply questions (CATA), Flash Profiling (FP) and Paired Comparisons, or methods based on the evaluation of global differences: Sorting, Projective Mapping (PM) or Napping, to methods based on a free, global evaluation of the individual products: Open-ended questions (Varela and Ares 2012). These methods can use either semi-trained panel or untrained assessors (consumers) (Minim and Silva 2016) and are important tools for food development and quality control, with a wide application in Food Science and Technology.

However, rapid methods can be evaluated to assure that the results are satisfactory despite of untrained panels and simplified procedures. Many studies have been carried out to compare rapid methods against CDP to validate them. Although some studies have also reviewed the rapid methods available in the literature (Varela and Ares 2012; Valentin et al. 2012), none of them have conducted a systematic review about the studies that tested the efficacy of various rapid tests by using conventional descriptive analysis as a "gold standard", as done for diagnostic tests in clinical research domains. There are several papers comparing "traditional" and "new" descriptive methods in an isolated way (Ares et al. 2015; Dehlholm et al. 2012; Richter et al. 2010; Dos Santos et al. 2015; Bruzzone, Ares, and Giménez 2012; Silva et al. 2013; Moussaoui and Varela 2010).

CONTACT Lívia de Lacerda de Oliveira liviadelacerda@gmail.com Campus Darcy Ribeiro, Asa Norte, 70910–900.

Supplemental data for this article can be accessed on the publisher's website.

The need for a systematic review has been found to compile all the knowledge that has not been analyzed yet together. According to Hemingway and Brereton (2009), systematic reviews (SR) have progressively substituted traditional narrative reviews and are recommended to summarize research evidence. Systematic reviews of diagnostic accuracy in health care fields are often carried out to evaluate test performance and impact based on all available evidences (Deeks 2001). In this type of review, studies that perform the experimental test and a "gold standard" diagnostic investigation to ascertain disease status are considered. Moreover, the quality and the risk of bias (RB) of these studies, related to aspects of design, methods of sample recruitment, execution of the tests, and the completeness of the study report, are considered as a procedure of systematic review protocols for diagnostic tests (Devillé et al. 2002).

A recent study of systematic review has been carried out by Lagast et al. (2017) regarding methods on consumer's emotions. However, they established a limitation on language (only papers written in English) and they did not perform an analysis of the risk of bias, an important step of systematic review to ensure the level of evidence. Likewise for systematic reviews on diagnostic tests, our goal is to adapt the protocol, designed for clinical field, to evaluate studies of validation of rapid descriptive tests against conventional descriptive analysis. For this purpose, we used Cochrane protocol as a reference (Leeflang et al. 2013). The aim was to evaluate if rapid descriptive tests can be used to replace conventional descriptive tests (considered here a "gold standard") with satisfactory efficacy to sensorially characterize food products.

This is a complete systematic review on the field of sensory analysis methodology and it is also a proposal of a full application of this type of review protocol from clinical research field to food and sensory sciences.

2. Materials and methods

This systematic review was reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist (Moher et al. 2009) and Guidance of European Food Safety Authority (EFSA 2010).

The "Population Intervention Comparison Outcome Study – PICOS (population, intervention, comparison, outcome, and type of study)" framework for diagnosis studies was used to design the research (Moher et al. 2009). "Population" was defined as food or beverages; "intervention" was the rapid descriptive methods, "Comparison" was made with conventional descriptive profile (CDP) as the gold standard, "Outcome" was the ability of rapid methods to generate similar descriptive profiles in comparison with the gold standard and "Studies (type of)" was the sensory descriptive analyses. Studies included in the review evaluated fast or rapid descriptive methods in comparison with CDP.

2.1. Protocol and registration

In the clinical field of health care studies, there is a Prospective Register of Systematic Reviews (PROSPERO), which does not apply to systematic reviews on food science. Therefore, we did not record the protocol in PROSPERO.

2.2. Eligibility criteria

2.2.1. Inclusion criteria

The present review included studies that evaluated fast or rapid descriptive methods in comparison with CDP, with no language or time restrictions. It comprised only experimental studies that used trained panel with quantitative description versus untrained (consumers) or semi-trained panels and any kind of descriptive evaluation.

2.2.2. Exclusion criteria

The following criteria were applied as exclusion criteria: 1) reviews, letters, conference abstracts, case reports; 2) other sensory analysis studies (affective and discriminative tests, instrumental-sensory correlations, threshold tests); 3) comparison between rapid methods or untrained panels without the gold standard (CDP); 4) using conventional descriptive tests without comparing with rapid descriptive method; 5) studies that were not related to sensory analysis and 6) rapid descriptive methods that used temporal approaches, such as temporal dominance of sensations or time-intensity analysis.

2.3. Information sources

Detailed individual search strategies for each of the following bibliographic databases were developed: SCIENCE DIRECT, SCOPUS, SPRINGERLINK, WEB OF SCIENCE and WILEY ONLINE LIBRARY. Grey literature search was carried out by using Google Scholar and Proquest. The end search date was across all databases and updated 27th September 2017. The references cited in the selected articles were also checked.

2.4. Search strategy

Appropriate truncation and word combinations were selected and adapted for each database search (Appendix A, electronic supplementary material). All references were managed by Endnote Web software version 3.1.1 and duplicate hits were removed.

2.5. Study selection

The selection was completed in 2 phases. In phase 1, two reviewers (LAA; L de L de O) independently reviewed the titles and abstracts of all identified electronic database citations. Articles that did not appear to meet the inclusion criteria were discarded. In phase 2, the same reviewers applied the inclusion criteria to the full text of the articles. The reference list of selected studies was critically assessed by the examiners (LAA, L de L de O). Any disagreement in first or second phase was resolved by discussion until an agreement among the reviewers was attained. When they did not reach a consensus, a third author (LM) became involved and was required to make a final decision. LM was considered the expert on sensory analysis to whom any doubt about sensory methods was addressed.

2.6. Data collection process

The following characteristics (Table 1) were collected: authors and year of publication, rapid method compared with a CDP,

Table 1. Characteristics of eligible studies.

Authors, year	RDM compared with CDP	Number of assessors	Assessors in RDM	Samples	Statistical analysis	Statistical criteria to compare conventional and rapid methods	Ability of RDM to generate similar descriptive profiles in comparison with CDP.
Rodrigue et al. (2000)	RDA	CDP: 8RDA: 20	Untrained	Types of corn	CDP: mixed model of ANOVA followed by Fisher's LSD; Multivariate analysis and GPA.RDA: Friedman test followed by Fisher's LSD; Multivariate analysis and GPA.	A) Levels of discrimination among products on each attribute between the multiple comparisons tests (Friedman/LSD and ANOVA/LSD)B) Consensus among assessors – GPA,C) Visual inspection of the maps	Yes.A)RDA panel discriminated more among products than the trained panel did, due to the use of a comparative test. CDP panel differentiated more products on three flavors and in mouth texture attributesB)In both cases, the consensus among judges for most attributes was satisfactory. Discrepancies observed with the untrained panel were not unexpected. A)Similar, only two non-concordant attributes.
Dairou and Sieffermann (2002)	FP	CDP: 13FP: 8	Untrained assessors with previous experience in sensory analysis methodology.	Jams	CDP: ANOVA, PCA and CVAFP: ANOVA, CVA, GPA and cluster analysis.	GPA by using coordinates from products after FP, GPA and mean conventional profile data were used as input to afurther GPA.A)Comparison of terms created by both panels.B)Panel performance: reproducibility and discrimination ability by to way ANOVA and CVA (% of good attribution).C)Comparison of panels performances: Kolmogorov-Smirnov test and Mann-Whitney test on data (% of good re-attribution) issued from CVA.D)Comparison of space products obtained by the 2 procedures.E)Comparison of information percentage principal components.F)Comparison of products clusters.	Yes. The two procedures produced similar information both in respect to the way the jams grouped and differed from another.A)Terms created are globally similar.B)Both consistent. The CDP panel was more discriminatingC)The assessors of CDP appeared to be more repeatable and discriminant than the FP.D)The 2 procedures produced similar information both in respect to the way the jams grouped and differed from another.E)The axes represented the same percentage of information in the 2 studies.F)Extremely similar clusters of products were formed by two methods.
Delarue and Sieffermann, (2004)	FP	Study1CDP: 10FP: 9Study2CDP:9FP: 10	Untrained assessors, experienced in sensory evaluation	Study 1Strawberry blended fruit yoghurtsStudy 2Apricot fresh cheeses	CDP: PCA and GPAFP: GPA	A)Comparison of sensory maps: PCA and GPA, visual inspection; Normalized RV coefficientB)Quality of product discrimination: CVA calculation of the F statistic (Rao's approximation), on inter-group Malahanobis distances and on the percentage of correct reattribution to the various groups of products.C)Semantic comparison.D) Description of the products.	Partially, depends on the product.A)Study 1: Similar structures of PCA plots. GPA plots perfectly matched Significant NRV coefficient for the matrices of samples coordinates.Study 2- different PCA plots. Good correspondence for GPA plots. Non-significant NRV coefficient.B)Study 1: CVA- distance matrix, dissimilarity of products pairwise, similarity of identical products by different methods. 100% of observations re-attribution, profiling method made no difference. Regarding the structure of product space. Higher F-approximation (discrimination power) for FP.Study 2: some products distinctly positioned, differing products discriminated and correctly reattributed to their corresponding clusters. Better discriminating power for FP.C)Study 1: more descriptors for FP. Attributes from CDP were present in FP. Many of the same descriptive characteristics were obtained with both methods. Many information brought to light by FP. Study 2: two attributes of CDP not used in FP. Richer vocabulary employed in FP.

(*Continued on next page*)

Table 1. (*Continued*).

Authors, year	RDM compared with CDP	Number of assessors	Assessors in RDM	Samples	Statistical analysis	Statistical criteria to compare conventional and rapid methods	Ability of RDM to generate similar descriptive profiles in comparison with CDP.
Cartier et al. (2006)	Sorting	CDP: 12Sorting: 24	Untrained assessors with previous experience in sensory analysis of breakfast cereal and CDP orUntrained regular consumers of breakfast cereals.	Breakfast cereals	CDP: two-way mixed-model of ANOVA, Fisher LSD and PCA.Sorting: MDS	A)Proximity of the two factorial configurations: RV coefficientB) Percentage of variance explained by the first two axes in the mapsC) Analysis of sample groups: k means clusteringD)Significant attributes	PartiallyA)RV coefficient – 0.31 (moderate) B)Slightly higher for CDPC)Sorting with trained assessors formed groups very similar to CDP. Grouping in sorting with untrained subjects differed slightly, by few samples.D)Products were described by more attributes in CDP. Differences in description were observed.
Kennedy and Heymann (2009)	PM	Study 1CDP: 9PM: 9Study 2CDP: 9PM: 9Study 3CDP: 11CPD: 8PM: 8	mix of experienced panelists and naïve consumers, none with previous sensory experience with chocolate	Dark chocolates	CDP: PCA.PM: MFA	MFAPCAA)Similarity between the sensory spaces: RV	Yes.A)Product spaces obtained from projective mapping show good correlation to product spaces from descriptive analysis. The RV coefficient for the matrices of samples coordinates between CDP versus PM for was more than 0.8 (strong).
Moussaoui and Varela (2010)	PMFPRGMSorting	CDP: 11PM: 24FP: 24RGM: 24Sorting: 24	untrained consumers	Hot beverages	CDP: ANOVA and Tukey's test, PCAPM: MFASorting: MDSFP and RGM: GPA	A)Grouping – HCAB)Correlation between RDM maps and CDP map: RV coefficient.C)Generated vocabularyD) Sample discrimination and repeatability: visual inspection of the maps	No for sorting and PM; partially for FP and RGM.A)RGM, FP and CDP formed the same clusters of samplesB)FP = RV coefficient – 0.91; RGM: RV coefficient – 0.88. Significance test not mentioned, and not presented for the other methods.C)FP and RGM: more suitable to generate a rich vocabulary. When describing mouthfeel (e.g. creaminess) the consumers' description was not enough detailed or consensual.D)Sorting and projective mapping performed poorly in terms of product discrimination and repeatability.
Richter et al. (2010)	RDAFCP	CDP: 12FCP: 14RDA: 21	untrained	Chocolate pudding	CDP: ANOVA, Tukey and GPA. FCP: GPARDA: Friedman test– to evaluate sample differences for each attribute – and GPA.	GPAA)Comparison among panel's.B) Configuration of samples consensus in comparison between CDP and FCP.C)Comparison of percentage of information on principal components.D)Comparison of samples distribution.	Yes, for RDA and partially for FCP.A) Analysis of the general configuration and residual variance of assessors of each panel demonstrated that there was consensus in the three panels. RDA panel showed less dispersion and a low residual variance.B)Both panels presented good repeatability for all samples.C)For the two first dimensions: CDP – 46%; FCP – 38% and RDA – 78%.D)Comparing the samples distribution in RDA with FCP and CDP similarities were observed in configuration (primarily in dimension 1) and in the attributes more important to discrimination.Both CDP and RDA assessors used aroma and texture descriptors more consensually than the FCP panel.

Albert et al. (2011)	FPPM	CDP: 10FP: 10PM: 20	FP: semi-trained assessorsPM: untrained consumers	Fish nuggets	CDP: ANOVA and PCA using the Pearson correlation. FP: GPAPM: MFA	A)Comparison of terms.B)Similarity between the sensory spaces determined by the evaluations of the three panels – visual inspection and coefficient RV.C)product clusters as perceived by the three Panels HCA.	Yes, for FP and PMA)CDP terms were less numerous, but more specific to the food category.B)Relative positioning of the samples in the three sensory spaces was similar. RV coefficient for CDP vs FP was 0.856 and for CDP vs PM was 0.816. Significance of Rv not mentioned.C)All the methods grouped the samples in three clusters, but using different criteria. Grouping in PM was closer to CDP.
Ares et al. (2011)	IS	CDP: 9IS: 86	Panel A: expertsPanel B: specialized expert assessors	Milk Desserts	CDP and IS: mix aaxxx of ANOVA model and Tukey. PCA and Pearson Correlation.	A)Global Panel Performance: mix linear of ANOVA model and Tukey test.B) Individual Assessors performance: agreement with the rest of the panel, variability in sample's evaluation discriminative capacity and reproducibility.	Partially.A)The discriminative capacity of the consumer and trained assessors panel was very similar for three out of five attributes. The results of Tukey test suggest that they were able to discriminate the desserts in the same way.B)Correlation coefficients between individual and average scores for the IS panel were low, ranging from 0.705 to 0.044. High variability in consumers' evaluation of the samples observed in the PCA (consumers distributed along the first two principal components). Low percentage of consumers who gave scores that discriminated between samples for three out of five samples. Percentage of consumers who showed a standard deviation of the residuals higher than 1.96 from 13% to 22% for the five attributes
Bruzzone et al. (2012)	CATAIS	CDP: 9CATA: 100IS: 100	untrained	Milk desserts	CDP: ANOVA, significant differences were calculated using Tukey's test. PCA– correlation matrix of the attributes scores.CATA: Friedman's test; MFA using overall liking as supplementary variable.IS: Mean, standard deviation, minimum and maximum score; ANOVA, Tukey's test; PCA.	A)Variability in assessors or consumer's evaluations.B)Coefficients for the correlation between textural attributes evaluated by consumers and trained assessors.C)Visual inspection of PCA representation of CDP and IS.D)Consensus representation of the methodologies with MFA.	Yes, for IS and CATAA)Trained assessors and consumers' intensity scores to discriminate between samples were similar. Larger number of consumers compared with the trained assessors compensated the larger variability and lack of training of the former panel.B) Good correlations, except for one attribute.C)Similar PCA configuration for CDP and IS.D)Proximity of the methods in the first two dimensions of the MFA: very similar information.
Dehlholm et al. (2012)	FMSGN/UFPPN/ UFPFP	CDP: 17FMS: 23GN/UFP: 18PN/UFP: 18FP: 10	Panel A: expert assessorsPanel B: specialized expert assessors	Liver pâté	CDP: ANOVA and PCARapid: MFA	A)Comparisons of overall sample configurations: RV coefficient, with significance testB)Differences in uncertainty between methods: 95% confidence ellipses.C)Analysis of sensory attributes used in the different methods for their significance of discriminating the samples: CDP – ANOVA; Rapid – frequency analysis.	Yes, for FP, FMS, PN/UFP and partially for GN/UFP.A)Panel A: RV all significantly and close to the CDP (0.86 to 0.74, strong). Panel B: GN, RV 0.48 and not configurationally similar to CDP. FMS, RV between 0.60 and 0.75 according to choice of dimensions. PN, RV, 0.74.B) GN: larger ellipsesC)Large difference in number of attributes, from CDP – 20 terms to FP – 189 terms. Vocabulary was not completely similar.

(*Continued on next page*)

Table 1. (*Continued*).

Authors, year	RDM compared with CDP	Number of assessors	Assessors in RDM	Samples	Statistical analysis	Statistical criteria to compare conventional and rapid methods	Ability of RDM to generate similar descriptive profiles in comparison with CDP.
Silva et al. (2012)	ODP	CDP: 9ODP: 14	Semi-trained	Light cheese curds	CDP: ANOVA (sample and repetition) and PCA.ODP: ANOVA with two sources of variation (sample and judge) and sample X judge interaction. In the case of significant, the principal effect (samples) was recalculated using the interaction mean square as the denominator and PCA.	A)Comparison between the descriptive methods: ODP and CP: ANOVAB) Comparison between the graphical representations of the descriptive data – PCA.C)Correlation analysis (Pearson) and Student t-test for paired samples.D)Correlation of the descriptive methodologies with the instrumental data	Yes.A)The effect product x assessor was significant for all attributes in CDP but for only one attribute in ODPB)In PCA, the first two components explained 98% of the variation of the data. Similar spatial configuration of PCA maps.C)Sensory description obtained from ODP and CDP present a significant correlation ($p < 0.10$) in relation to all attributes. Results of sensory description of ODP and CDP did not differ by the t-test for paired samples ($p > 0.10$)D)Existence of the same correlations suggested by the sensory data of CDP x instrumental data and ODP x instrumental data.
Jiménez et al. (2013)	FPPM	CDP: 8FP: 30PM:30	FP: semi-trainedPM: untrained	Potato puree	CDP: two-way mixed-model of ANOVA, Fisher and PCA– Pearson's correlation.FP: GPAPM: MFAAll methods- HCA	A)Samples clustering: HCACorrelation with CDP: MFA and RV coefficient	Yes for FP and partially for PMA)Relative positioning of both samples and attributes quite comparable. Sample clustering differed slightly between CDP and RDM.B)Rv coefficient between QDA versus FP, 0.92; QDA versus PM, 0.90 (significance test not mentioned), however, in PM untrained assessors' perception of "firm/ firmness" was not well correlated to the trained panel assessment.
Silva et al. (2013)	RDAODP	CDP: 10RDA: 15ODP: 16	Semi trained	Chocolate formulations	CDP and ODP: ANOVA. F test by using MS interaction. Duncan's multiple range test.RDA: Friedman test.	A)Comparison between the graphical representations: correlations of the attributes with the principal components and the similarity in the formation of product groupsB) Discrimination power of the treatments (CDP x ODP): p-value of samples (ANOVA).C) Effect of interaction and use of the unstructured scale (CDP x ODP): existence of interaction between the judges and formulations by ANOVA; graphical evaluation – evaluations of each judge for each formulation in relation to the individual sensory attributes; frequency analysis of scores.D)Correlation between sensory and instrumental measurements: Pearson correlation (CDP, ODP) Spearman correlation (RDA).	Yes, for ODP and RDAA)Similar descriptive maps for the three methods.B)Similar power of F test (CDP and ODP). ODP showed greater variance among the chocolate formulations and less variance due to interaction when compared to the CP. Similar estimate of residual variance due to the random error. Higher sample mean square for ODP. Similar discrimination power for ODP and CDP.C)From F-test it was found that interaction between the formulations and assessors was significant ($p < 0.10$) for all attributes in the ODP (except for one) and CDP. In the chi-square test a significant difference was observed between the distributions ($p < 0.10$) for all sensory attributes indicating that CDP and ODP showed differentiated utilization of the unstructured scale portions.D)RDA > ODP >CDP.

Torri et al. (2013)	PM	CDP: 9PM: 81	wine experts and untrained consumers	Wine	CDP: three-way ANOVA mixed model, sample as fixed factor; LSD Fisher's test. PCA and GPA.PM: GPA	A)Relationship between the products' coordinates from the two first dimensions of consensus maps	No.A)Different sample configuration
Cardinal et al. (2014)	FS	CDP: 11FS: 45	Untrained from staff and students	Salmon by-products	CDP: ANOVA, Duncan multiple range test and PCAFS: MDS	Comparison of the two sensory maps:A) RV coefficient.B)Confidence ellipsesC)Descriptors	Yes.A)The RV coefficient of sample configurations for CDP versus FMS was strong (0.81, significance test not mentioned)B)CDP: confidence ellipses of more samples were separated, highlighting more differencesC)Slight differences. CDP highlighted a specific odor for one sample
Mielby et al. (2014)	Sorting;PM	CDP: 11Sorting: 11PM: 11	prior experience with general descriptive analysis	Pictures of fruit and vegetable mixes	CDP: Mixed three-way ANOVA and PCA.PM: MFA.Sorting: MDS and DISTATIS.	A)Visual inspection of the mapsB) Differences between replicates	Yes, for PM and sorting.A)Similar results of PM, sorting and CDP.B)Differences were found between replicates for RDM, possibly holistic product evaluation leads to change in assessors' evaluation criteria
Ares et al. (2015)	CATA	Study 1CDP: 10CATA: 199Study 2CDP: 11CATA: 431Study 3CDP: 10CATA: 102Study 4CDP: 10CATA: 134	consumers	Study 1Milk dessertsStudy 2Orange flavoured powered drinksStudy 3Raspberry coulisStudy 4White wine	CDP: ANOVA, Tukey and PCA. CATA: Cochran's Q test and CA	A)Percentage of CDP descriptors for which conclusions regarding differences among samples differed in CDP and CATAB)RV between sample configurations based on the first two dimensions of PCA/CAC) Visual inspection of the maps	Partially.A)The percentage varied from 0 to 25% depending on the productB)All Rv coefficients of samples configurations were significant and strongC)Many differences between sample configurations from CDP and CATA questions. Discriminative ability of CATA questions with consumers was lower than that of CDP with trained assessors for some attributes.
Dos Santos et al. (2015)	CATAFL	CDP: 18CATA: 106FL: 106	Consumers	Dry fermented sausages	CDP: one -way of ANOVA and PCA using Pearson correlation matrix.CATA: CA considering chi-square distance and MDA.Free listing: Smith's salience index (Sj) and PCA.	A)HCA- visual inspection of the dendograms.	No, for both.CATA and FL did not exhibit similar configuration compared with CDP.A)Different clusters were formed according to the method applied, which is related to cognitive strategies used by consumers in the evaluation of the samples in relation to the trained panel.
Oppermann et al. (2017)	RATA	CDP: 11RATA: 80	Untrained participants	Emulsions	CDP: ANOVA with Tukey's Honest Significant Difference test (HSD) and PCA.RATA: two-way ANOVA with Tukey's HSD. PCA and Cochran's Q test.	A)MFA – configurational similarity of products space obtained.B)Rv coefficients: significances of the Rv were calculated with Pearson type.	Yes.A)Similar configurational between maps for CDP and RATA.B)High values of RV indicated good agreement between sample configurations. RATA frequency vs CDP Rv = 0.778 ($p = 0.002$) and RATA intensity Rv = 0.663 ($p = 0.005$).
Moelich et al. (2017)	Partial and Global PM	CDP: 9PM: 10	Trained assessors	Honeybush infusions	CDP: ANOVA; Shapiro-Wilk test. PCA.PM: MFA	A)Rv coefficients was calculated to determine the similarity between the PCA (CDP) and MFA (PM)	Yes.A)The Rv coefficients between the variations of PM and CDP was high ($\geq$0.86). Global PM showed the highest RV coefficient (0.90).
Horita et al. (2017)	PPM	CDP: 10PPM: 50	Naive consumers	Low sodium sausages	CDP: ANOVA with Tukey's. PCA. HCAPPM: MFA and HCA.	A)MFA used to compare the sensory methodologies, assessing the positioning of products in a bidimensional map.B)Rv used to analyze similarities between samples of both methodologies.	Yes.A)The positions of samples in two dimensional space shows that both assessors (trained and naive) were capable of discriminating samples.B) The RV was 0.649 ($p<0.001$), close to the minimum valou recommended as a cut-off point among samples configuration.

(*Continued on next page*)

Table 1. (*Continued*).

Authors, year	RDM compared with CDP	Number of assessors	Assessors in RDM	Samples	Statistical analysis	Statistical criteria to compare conventional and rapid methods	Ability of RDM to generate similar descriptive profiles in comparison with CDP.
Pickup, Bremer, Peng (2017)	Napping®-UFP	CDP: 8Napping/UFP: 72	Untrained and with no experience in any sensory experiment	Apple cultivars	CDP: ANOVA with Tukey's post hoc and PCA. Napping/UFP: MFA	A)RV coefficient was employed to indicate resemblance between product maps generated and Pearson's correlations.	No.A)The comparisons of sensory profiles with the RV coefficient were weakly correlated – RV = 0.425, P = 0.077. CDP and Napping®-UFP separated the samples based on distinct mechanisms. In this study Napping®-UFP and DA should be regarded as complementary methods, combination of which lead to comprehensive sensory profiles of products.

Statistical analysis: ANOVA: Analysis of Variance; **CA**: Correspondent analysis; **CVA**: Canonical Variate Analysis; **GPA**: Generalized Procrustes Analysis; **HCA**: Hierarquical Cluster Analysis; **PCA**: Principal Components Analysis; **MANOVA**: Multiple Analysis of Variance; **MDA**: Multidimensional alignment; **MDS**: Multidimensional Scaling; **MFA**: Multiple Factor Analysis; **RV**: Regression Vector; **Sj**: Smith's salience index.

Method: CATA: Check all that apply; **CDP**: conventional descriptive profile; **FCP**: Free Choice Profiling; **FL**: Free Listing; **FMS**: Free Multiple Sorting; **FP**: Flash Profiling; **FS**: Free Sorting; **FST**: Free Sorting Task; **GN**: Global napping; **IS**: Intensity Scales; **ODP**: Optimized Descriptive Profile; **PM**: Projective Mapping; **PN**: Partial Napping; **PPM:** Polarized Projective Mapping; **PSP**: Polarized sensory positioning; **RATA: Rate all that apply; RDA**: Ranking descriptive analysis; **RDM**: Rapid Descriptive Method **RGM**: Repertory Grid Method; **UFP**: Ultra Flash Profiling.

number of assessors, assessors in RDM, samples, statistical analysis, statistical criteria to compare CDP with rapid methods, ability of RDM to generate similar descriptive profiles in comparison with CDP. To ensure consistency across reviewers, calibration exercises were conducted before starting the review. Reviewers resolved disagreements by discussion, and the third author (2R) adjudicated unresolved disagreements.

2.7. Risk of bias (RB)

In the clinical field of health care studies, there are Newcastle-Ottawa Quality Assessment Scale, Risk of bias assessed by Meta Analysis of Statistics Assessment and Review Instrument (MAStARI), Cochrane instrument and others, which generally does not apply to systematic reviews on food science. A specific instrument was created for this study. The RB assessment was based on the conditions stated by Stone et. al. (1974) for Quantitative Descriptive Analysis and on general rules related to correct sensory evaluation procedures, from data collection to statistical analyses (Meilgaard et al. 2006; Lawless and Heymann 2010; Stone and Sidel 2004; Valentin et al. 2012; Delarue, Lawlor, and Rogeaux 2014; Varela and Ares 2014). Selected criteria for RB assessment were: randomized coded (blind) samples; appropriate number of samples per session (based on possible sensory and mental fatigue); balanced presentation of samples for both tests; recruitment and selection of assessors; appropriate number of assessors for CDP; reported statistical validation of trained panel; appropriate number of samples to the applied data analysis and vice versa (based on verifying whether the studies used multivariate statistical techniques with the appropriate number of samples); appropriate statistical method in relation to type of data; percentage of explanation of CDP maps (higher than 70%, according to Favero et al. 2009); and RV coefficient between rapid descriptive test and CDP. Likewise for the analysis with Pearson correlation, $1 \geq r \geq 0.90$ was considered a very strong correlation; $0.90 > r \geq 0.6$, a strong correlation; $0.6 > r \geq 0.3$, a moderate correlation and $0.3 > r \geq 0.0$, a weak correlation. Evaluation of the statistical significance of RV was also checked. Two reviewers (LAA and L de L de O) performed RB analysis of the studies. Reviewers resolved disagreements by discussion, and the third reviewer (LM) adjudicated unresolved disagreements. Each criterion received yes (Y), no (N), unclear (U), or not applicable (NA) in the evaluation of the studies. RB was categorized as High (H) when the study reached up to 49% score "yes", Moderate (M) when the study reached 50% to 69% score "yes", and Low (L) when the study reached more than 70% score "yes" (Higgins and Green, 2011) (Table 2).

2.8. Summarized requirements of rapid descriptive methods

To present rapid methods that are part of this study, the summary measures were described in Table 3. This table described the procedures and statistical analyses to each rapid descriptive method, based on the literature of sensory analysis (Minim and Silva 2016; Simiqueli et al. 2015; Varela and Ares 2014; Delarue et al. 2014; Valentin et al. 2012; Hough and Ferraris 2010; 2010; Jack and Piggott 1991) and is a reference to analyze Table 1 (studies' data extraction).

Table 2. Risk of bias assessment.

Paper	Risk of Bias
Rodrigue et al. (2000)	M
Dairou and Sieffermann (2002)	L
Delarue et al. (2004)	M
Cartier et al. (2006)	L
Kennedy et al. (2009)	L
Moussaoui e Varela (2010)	L
Richter et al. (2010)	L
Albert et al. (2011)	L
Ares et al. (2011)	L
Bruzzone et al. (2012)	L
Silva et al. (2012)	L
Dehlholm et al. (2012)	L
Jiménez et al. (2013)	L
Silva et al. (2013)	L
Torri et al. (2013)	L
Cardinal et al. (2014)	L
Mielby et al. (2014)	M
Ares et al. (2015)	L
Dos Santos et al. (2015)	L
Oppermann et al. (2017)	L
Moelich et al. (2017)	L
Horita et al. (2017)	L
Pickup, Bremer, Peng (2017)	L

3. Results and discussion

Table 1 presents eligible studies, their characteristics and the ability of the RDM to generate similar descriptive profiles in comparison with CDP. Initially, 1168 studies were identified across the electronic databases, remaining 1099 after removing duplications. After a comprehensive evaluation of the abstracts in phase 1, 61 articles, deemed potentially useful, according to inclusion and exclusion criteria, were selected for phase 2 assessment. In addition, 12 articles were identified from additional search, Google Scholar (n = 11) and ProQuest (n = 1). Of these 61 remaining studies, a total of 42 were subsequently excluded (Appendix B, electronic supplementary material, reasons for exclusion) and 4 studies extracted from reference lists were added. Only 23 studies were ultimately retained for this systematic review. A flowchart of the identification, inclusion, and exclusion process is shown in Figure 1.

The years of publication of the 23 eligible studies in the systematic review are from 2000 to 2017, from 13 different countries: Canada (1), France (4), Switzerland (1), United States (2), United Kingdom (1), Brazil (5), Spain (2), Uruguay (3), Denmark (1), Italy (1), New Zealand (2), South Africa (1) and Netherlands (1). Thirteen RDM were tested in comparison with CDP: Projective Mapping or Global or Partial Napping or Napping, sometimes associated with Ultra Flash Profiling or to further description of sample groups to generate the analyzed attributes, with nine studies; Flash Profiling with six studies, Sorting (including Free Multiple Sorting, Free Sorting Task) with five studies; Check all that apply, Ranking Descriptive Analysis, three studies each; Intensity Scales and Optimized Descriptive Profile, two studies each; Free Choice Profiling, Free Listing, Repertory Grid, Polarized Projective Mapping and Rate-all-that-apply, with one study each. Although Valentin et al. (2012) have claimed Free Sorting to be the most popular alternative to CDP, it was observed that other studies have more frequently validated other methods against the gold standard. Besides, it is worth noting that one of the most ancient

Table 3. Summarized requirements of rapid descriptive tests.

Rapid descriptive tests	Procedure	Statistical Analysis
Ranking Descriptive Analysis (RDA) (Minim and Silva 2016)	Subjects are not trained with the reference samples and use ranking instead of intensity scales to compare samples simultaneously.	Friedman and Kramer's tests and LSD and Christensen's multiple comparison tests
Optimized Descriptive Profile (ODP) (Simiqueli et al. 2015)	Similar to CDP. However, panelists are not trained and they evaluate all the samples simultaneously; one attribute per session, and with the presence of references of the intensity scales.	ANOVA, Tukey, LSD 's multiple comparison tests, PCA.
Intensity Scales (IS)Repertory Grid Method (RGM) (Varela and Ares 2012)	Consumers rate the intensity of a fixed or previously elicited set of sensory attributes by using intensity unstructured scales.	ANOVA, Tukey, LSD 's multiple comparison tests, PCA.
Sorting / Free sorting task (FST)/ Free Multiple sorting (FMS) (Minim and Silva 2016)	Samples are presented simultaneously and assessors are asked to evaluate and to sort them in mutually exclusive groups based on product-perceived similarities. After grouping procedures, subjects can be asked to provide a few terms to characterize each group.	CA, MDS, MCA, multi-block analysis
Projective Mapping (PM)NappingPartial Napping (PN) Global Napping (GN)Sorted napping (SN) (Minim and Silva 2016; Valentin et al. 2012)	Spatial arrangement procedure, in which subjects are asked to evaluate all the samples and then to position them on an A3 white sheet of paper, according to the similarities or dissimilarities between these products. Some variations according to attributes and grouping result in partial and sorted napping, respectively.	MFA, PCA
Flash Profiling (FP) (Dairou and Sieffermann 2002; Varela and Ares 2012; Delarue et al. 2014)	Mixed procedure of Free Choice Profiling (FCP) and RDA. Consumers are requested to individually elicit attributes and then order samples according to their intensity, with no level of previous training.	GPA, PCA, HCA
Ultra-flash profiling (UFP) (Minim and Silva 2016)	The same procedure of global napping, after which subjects are asked to write the descriptors direct on the sheets to characterize a sample or a group of samples.	CA, MFA by using descriptive data as supplementary variable.
Free Choice Profiling (FCP) (Minim and Silva 2016; Jack and Piggott 1991)	Individual descriptors elicitation. Use of intensity scales to individually quantify each descriptor of each sample. Monadic presentation.	GPA
Free listing (FL) (Hough and Ferraris 2010)	Subjects are asked to write terms related to samples	MDS and Cluster Analysis
Check-all-that- apply (CATA) (Ares et al. 2014)	Subjects receive samples with a list of descriptors previously ellicited and are asked to check the terms that are suitable to describe each sample	Cochran's Q test, CA; MFA, MCA, multi-block Hellinger Analysis. Friedman's test.
Rate-all-that-apply (RATA) (Ares et al. 2014)	Subjects receive samples with a list of terms and they asked selected as ''applies'', to rate intensity (using a 3-pt scale with anchors 'low', 'medium' or 'high') or rate applicability (using a 5-pt scale anchored at ''slightly applicable'' and ''very applicable'').	Cochran's Q test, Friedman's test, CA.
Polarized Projective Mapping (PPM) (Ares, et al. 2013; Pagès 2005)	Assessors are invited to locate a set of samples on a sheet of paper where three reference samples have already been placed. Then, they must locate samples according to their similarities and differences in comparison with the references. Considering that samples that are placed close to each other are similar and those that are far from each other are different. After positioning the samples, assessors are invited to use terms to describe them.	MFA, Confidences ellipses.

[1]Semi-trained or highly trained evaluators.

ANOVA: Analysis of Variance; **CA**: Correspondent analysis; **GPA**: Generalized Procrustes Analysis; **HCA**: Hierarchical Cluster Analysis; **MCA**: Multiple Correspondence Analysis; **MDS**: Multidimensional Scaling; **MFA**: Multiple Factor Analysis; **PCA**: Principal Components Analysis.

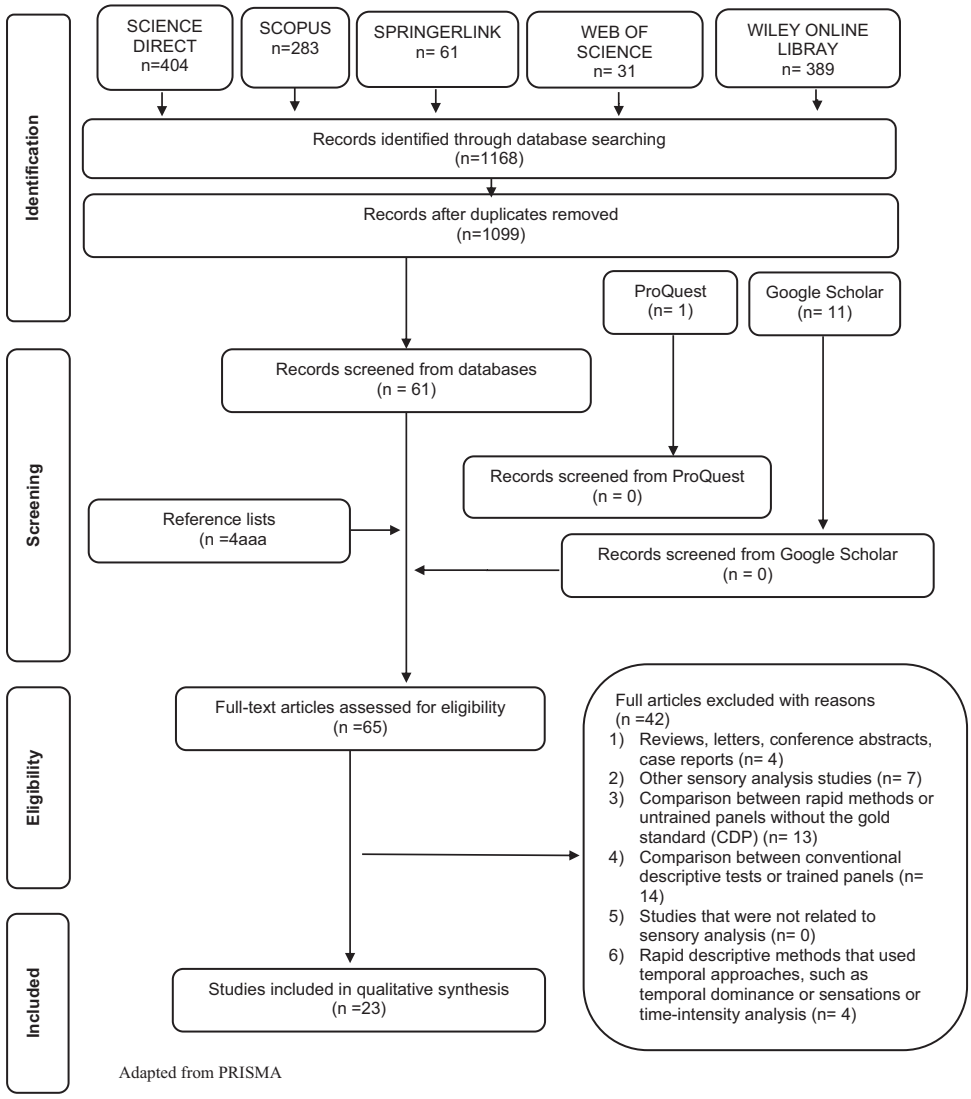

Figure 1. Flow diagram of literature search and selection criteria.[1]

semi-rapid methods, Free Choice Profile, had only one study comparing its performance with the CDP.

Regarding the criteria used to evaluate similarity between RDM and CDP, 48% of the 23 studies applied RV coefficients between samples configurations, 43% supported their conclusions on visual inspection of the perceptual maps, 26% compared the percentages of explanations of the first dimensions of the maps, 13% evaluated panel repeatability and degree of discrimination among samples, 9% used confidence ellipses to group and discriminate samples and 26% used comparison of samples clustering (HCA- visual inspection of the dendograms).

3.1. Projective mapping, napping and ultra flash profiling

Napping (partial napping, global napping) or Projective Mapping (partial and global) is a technique based on global differences among samples, in which assessors are asked to group samples according to their similarities and differences, in a positioning procedure on a A3 paper sheet. Similarities and dissimilarities are demonstrated through distances among samples in the two-dimensional space. Additionally, assessor can be asked to write some descriptive, or even affective, comments about the samples or groups of samples, which sets the Ultra Flash Profile (UFP) method (Dehlholm et al. 2012). Eligible studies included evaluation of dark chocolates (Kennedy and Heymann 2009), hot beverages (Moussaoui and Varela 2010), fish nuggets (Albert et al. 2011), liver pâté (Dehlholm et al. 2012), potato pureé (Jiménez, Canet, and Alvarez 2013), wine (Torri et al. 2013), pictures of fruit and vegetable mixes (Mielby et al. 2014), honeybush infusions (Moelich et al. 2017) and apple cultivars (Pickup, Bremer, and Peng 2017).

Similarity between descriptive profiles was concluded in five out of nine studies (Kennedy and Heymann 2009; Albert et al. 2011; Dehlholm et al. 2012 and Mielby et al. 2014, Moelich et al. 2017). When Global Napping (GN)/UFP was carried out in the study of Dehlholm et al. (2012), a partial relation was found between RDM and CDP, like in the study of Jiménez et al. (2013). Moussaoui and Varela (2010), Torri et al. (2013) and Pickup, Bremer, Peng (2017) have found low similarity between PM and CDP because PM performed poorly in terms of product discrimination and repeatability (Moussaoui and Varela 2010), presented different sample configuration (Torri et al. 2013) and the method generated a characterization of the

samples in a less analytical way (Pickup, Bremer, and Peng 2017).

Kennedy and Heymann (2009) suggest that a combination of PM and CDP data helps in the interpretation of sensory maps. Their conclusions were based on strong correlations (higher than 0.8) of product spaces of RDM and CDP. The correlation between first dimensions of the consensus map and the descriptive attributes can be used to examine the criteria used by the assessors in assessing overall product similarity. However, this evaluation must be done carefully once data generated from PM do not stablish a relationship among the attributes. It is important to mention that RV coefficient was calculated for the matrices of samples coordinates and not attributes, once they are different between the two methodologies. Therefore, the conclusion for similarity of the descriptive profiles is quite limited since correlation is not based on descriptors data.

Albert et al. (2011) concluded that napping can be used as a RDM to describe fish nuggets. By comparing the methods, authors showed that CDP terms were less numerous, but more specific for the food category. Relative positioning of the samples in the sensory spaces was similar. The RV coefficient for CDP vs PM was strong (higher than 0.8). PM and CDP grouped the samples in three clusters, however using different criteria, although PM was closer to CDP than FP. Another issue is that authors did not mention if RV was significant or not, which is a frequent limitation to studies that apply the RV coefficient to compare methods.

Dehlholm et al. (2012) found similarities between PN/UFP and CDP but only a partial similarity for GN/UFP method. The authors observed that PN might limit the holistic perception by restricting the ability to apply cross-modal discriminators, making descriptive analysis closer to CDP when compared to GN. They also showed that GN/UFP generated larger 95% confidence ellipses, as well as a RV of samples configurations of 0.48, not configurationally like CDP for one of the panels. Another issue was the different semantics applied, although configurations were sometimes similar.

Mielby et al. (2014) found similar maps for CDP and PM but the RDM performed better discrimination of the samples than CDP, for visual stimuli. The authors observed variations in replicates for RDM and stated that it is necessary more than one evaluation to provide stable and valid results.

Moelich et al. (2017) found high RV coefficient values ($\geq$ 0.86) when comparing CDP with PM variations, and concluded that PM is a reliable method for the sensory profile of the infusions of *Cyclopia spp*. For global PM the RV coefficient was the largest (0.90). The study by Jiménez et al. (2013) also used PM and found a partial relationship of RDM with CDP because the perception of "firm/firmness" of potato purees was not well correlated to the trained panel assessment.

Torri et al. (2013) concluded that the level of experience and knowledge of assessors influenced the performance of PM replacing CDP. The method was informative about similarities/dissimilarities among samples, but not for naïve consumers for whom differentiation of samples was more related to liking. The comparison between perceptual map from PM by trained panel (GPA) and CDP (PCA) showed that the second PCA component was closely related to the first GPA dimension. However, samples distributed along the second GPA dimension differed from those along the first PCA component. Projective mapping with hot beverages and naïve consumers performed poorly in terms of product discrimination and repeatability, according to Moussaoui and Varela (2010). Pickup, Bremer and Peng (2017) found a weak correlation when comparing CDP with Napping. Napping was able to facilitate the characterization of the apple samples tested and the differences found were expected due to the different configuration mechanisms of the samples.

3.2. Flash profiling

Flash profiling, a technique formerly proposed by Dairou and Sieffermann (2002), was applied to evaluate jams (Dairou and Sieffermann 2002), strawberry blended fruit yoghurts, apricot fresh cheeses (Delarue et al. 2004), hot beverages (Moussaoui e Varela 2010), fish nuggets (Albert et al. 2011), liver pâté (Dehlholm et al. 2012) and potato puree (Jiménez et al. 2013). It is based on a mixed procedure of Free Choice Profiling and Ranking Descriptive Profiling, once consumers are requested to individually elicit attributes and then order samples according to their intensity, with no level of previous training. As the data matrix is incomplete, due to different sets of attributes per assessor, Generalized Procrustes Analysis (GPA) is applied, resulting in a descriptive map, similar to FCP. The number of assessors in FP ranges from 8 to 30 (Dairou and Sieffermann 2002). The ideal number for FP is not well stated in the literature. As it derives from FCP and RDP, the number of panelists were similarly low in the initial studies. However, aligned to other rapid methods with untrained assessors requiring a larger number, FP had the panel increased to 18 and 30 consumers in the latter studies (Varela and Ares 2014).

Similarity between FP and CDP panels was concluded in four out of six studies. For jams with untrained assessors and previous experience in sensory analysis (Dairou and Sieffermann 2002), FP and CDP showed global similarity in the terms created and the panels' performances were consistent, although CDP was more repeatable and discriminating. In addition, similar information was produced, regarding how samples were grouped or discriminated, percentages of retained information on the map axes were comparable and products' clusters were alike except for one sample.

Analogously for PM, Albert et al. (2011) concluded the similarity between FP and CDP in the evaluation of fish nuggets, using the same statistical criteria. According to the authors, the interpretation of sensory descriptors in FP is not always easy due to the large number of attributes and the lack of definitions and evaluation proceedings.

For the study of Dehlholm et al. (2012) with liver pâté, FP was similar to CDP. Like PN, FP showed significant RV and smaller 95% confidence ellipses. For the authors, an interpretational drawback could be the excessively large vocabulary elicited in FP.

Although the similarity of CDP and PM was only partially observed for sensory analysis of potato purée by Jiménez et al. (2013), good similarity was stated between CDP and FP for the same samples. The authors found a high RV coefficient, 0.92, considered strong despite not mentioning a significance test. Samples clustering was slightly different between CDP and the studied RDM. Furthermore, texture attributes were highlighted

by trained assessors and neglected by consumers who, in turn, valued appearance and aroma descriptors, neglected by trained assessors.

From the study of Delarue and Sieffermann (2004) the conclusion found was that the type of food influences performance of FP against CDP. All criteria used to evaluate validity of FP was quite satisfactory for blended fruit yoghurts, but not for apricot fresh cheeses. It raises the question if RDM should be validated food by food, method by method. Partial similarity was also concluded by Moussaoui and Varela (2010) for hot beverages with untrained assessors, in spite of a strong RV (0.91, significance test not mentioned).

3.3. Sorting, free sorting task, free multiple sorting, free sorting

Sorting, Free sorting task (FST)/ Free Multiple sorting (FMS)/ Free Sorting (FS) was tested against CDP by Cartier et al. (2006) with commercial breakfast cereals, Moussaoui and Varela (2010) with hot beverages, Dehlholm et al. (2012) with liver pâté, Cardinal et al. (2014) with salmon, and by Mielby et al. (2014) with pictures of fruit and vegetable mixes. In this method, samples are simultaneously presented, and consumers are requested to evaluate and group them according to their similarities. After that, consumers are asked to provide descriptors for each group to characterize samples within it. Sorting techniques were shown to be similar to CDP in three out of five studies (Dehlholm et al. 2012; Cardinal et al. 2014; Mielby et al. 2014). In one study, sorting showed partial correspondence with CDP (Cartier et al. 2006). Another study concluded that sorting did not generate a similar profile with CDP (Moussaoui and Varela 2010).

Dehlholm et al. (2012) concluded for similarity of FMS if compared with CDP. The authors used two panels of experts in the study: panel A from University environment, mainly students, and panel B from a research institute that supports meat industry. For panel A, a high RV coefficient was found for FMS compared with CDP. On the other hand, for panel B, a lower RV coefficient was presented. Moreover, confidence ellipses for each sample showed larger ellipses for FMS, indicating noisier data. In the study with salmon by-products and untrained assessors, Cardinal et al. (2014) found a strong RV coefficient between sample configurations from FS and CDP. However, more samples had their confidence ellipses separated, showing more differentiation in CDP. Descriptors elicited in both methods were very analogous, but a specific odor for one sample was shown in CDP.

As occurred for PM, Mielby et al. (2014) found similarity and higher discrimination of samples for sorting when compared with CDP, but sorting provided clearer separation of fruit samples from the other ones. Distatis performed better than MDS to sorting data, by grouping tightly similar samples. On the other hand, it is possible that the holistic product evaluation associated to PM and sorting leads to change in assessors' evaluation criteria thus providing differences between replicates for the RDM.

Cartier et al. (2006) found partial similarity between sorting and CDP in a study with breakfast cereals. A moderate RV coefficient (0.31), a slightly higher percentage of variance explained by CDP, besides small differences in sample grouping and in number and description of attributes supported the conclusion.

Moussaoui and Varela (2010) concluded that sorting, as PM, performed poorly in terms of hot beverages discrimination and repeatability and thus only a rough description of the products can be obtained from these methods.

3.4. Check all that apply

Milk dessert (Bruzzone et al. 2012; Ares et al. 2015), orange flavored powered drinks, raspberry coulis, white wine (Ares et al. 2015), and dry fermented sausages (dos Santos et al. 2015) were used in studies of Check all that apply (CATA) method validation. This method is based in a list of descriptors previously developed in which consumers are asked to check all attributes that they considered appropriate to describe products.

Similarity against CDP was observed in a study performed by Bruzzone et al. (2012). Ares et al. (2015) found that CATA could partially replace CDP, whereas Dos Santos et al. (2015) did not find similarity between CATA and CDP for dry fermented sausages analyzed by consumers. According to Bruzzone et al. (2012), similitude of the methods in a study with milk desserts and untrained assessors were supported by the proximity of the methods' results in the first two dimensions of consensus representation using MFA, indicating very similar information.

Correspondence of descriptive profiles was partial in the study of Ares et al. (2015), with several products (milk desserts, orange flavor powered drinks, raspberry coulis and white wine). The percentage of CDP descriptors with conclusions indicating differences among samples differed between CDP and CATA ranged from 0 to 25%. On the other hand, the RV coefficients of samples configuration of all product studies were significant and strong, although many differences between sample configurations from CDP and CATA questions were observed in visual inspection of the maps. Moreover, the discriminative ability of CATA questions with consumers was lower than that of CDP with trained assessors for some attributes. Dos Santos et al. (2015) did not find similarity between CDP and CATA in the study with fermented sausages, supported by the different clusters of products formed by Hierarquical Cluster Analysis, from the different methods.

3.5. Ranking descriptive analysis

Ranking Descripitive Analysis (RDA) was applied by Rodrigue et al. (2000) for corn, by Richter et al. (2010) for chocolate pudding and by Silva et al. (2013) for chocolate formulations. The procedure is similar to CDP except for the fact that assessors are not trained in intensity scales after descriptors elicitation and there is a single contact with reference samples for a basic calibration (Minim and Silva 2016). Assessors compare samples by ranking them according to each attribute. Regarding the three studies of validation of this RDA, similarity to CDP was found in all of them.

Rodrigue et al. (2000) evaluated types of corn with untrained assessors and RDA. Based on statistical criteria, they observed similar discrimination among samples, but slightly higher for

some products and attributes for the trained panel, besides consensus among assessors for most attributes and maps visually alike and very few non-concordant attributes.

Richter et al. (2010) evaluated chocolate puddings by RDA in comparison with CDP and showed the validity of the method, based on the lower RDA panel dispersion and low residual variance in GPA analysis. Similarities were observed in configuration and in the most discriminating attributes. It is important to note that for the two first dimensions, the percentage of variance explained for CDP was low (46%) and thus the other evaluated methods appeared to be better than they really are. Furthermore, for a higher percentage of explained variance, RV coefficient could be used to perform a more precise comparison between the maps.

Silva et al. (2013) showed that RDA is similar and equivalent to CDP once maps showed similar spatial configurations, despite a lower discrimination of samples in RDA, detected by "confounding regions". By generating ordinal measures, it is not possible to state if these groups present large or small differences in sensory attributes in the RDA results. Nonetheless, this method showed the higher correlations between sensory and instrumental measurements, when compared to the other methods in the study.

3.6. Optimized descriptive profile

Optimized Descriptive Profile (ODP) is a rapid method with semi trained assessors that stands out by providing quantitative measures of sensory attributes. The technique proposed by Silva et al. (2012) was applied to evaluate light cheesecurds and chocolates (Silva et al. 2013). Training with reference samples was replaced by presentation of reference materials during final evaluation of the products, helping the allocation of attribute intensity on the unstructured scale. ODP was shown to be similar to CDP in both studies. According to Silva et al. (2012), the similarity between ODP and CDP study with cheesecurds was supported by the lower effect product x assessor achieved in ODP, compared with CDP, in ANOVA analysis; by similar spatial configuration of PCA maps; by the significant correlation ($p < 0.1$) and nonsignificant t-test ($p > 0.1$) of sensory description obtained by ODP and CDP for all attributes; and by the existence of the same correlations suggested by the sensory data from CDP x instrumental data and ODP x instrumental data.

Regarding ODP with chocolates (Silva et al. 2013), similar descriptive maps were achieved in comparison with CDP, besides similar power of F test, with higher variance due to samples and lower variance due to interaction for ODP. Similar estimation of residual variance due to the random error and higher sample mean square was also found with ODP. Both methods presented similar discrimination power. From F-test, interaction between the formulations and assessors was significant ($p < 0.10$) for all attributes in ODP (except for one) and in CDP. Both methods presented differentiated utilization of the unstructured scale portions, based on chi-square test between the distributions ($p < 0.10$) for all sensory attributes. Finally, for correlation between sensory and instrumental data for texture, ODP presented higher and significant correlation coefficients than CDP. The use of RV coefficient in this study could have provided additional support for their conclusions.

3.7. Intensity scales

Intensity Scales (IS) is a method based on the evaluation of individual attributes. Assessors are asked to quantify a certain set of attributes by using intensity scales, as commonly performed by trained assessors in CDP. Descriptors are provided by the sensory analyst and no training is carried out (Varela and Ares 2014). IS was tested in comparison with CDP in the studies of Bruzzone et al. (2012) and Ares Bruzzone and Giménez (2011), both with milk desserts.

Validation of IS was achieved in the study of Bruzzone et al. (2012). The similarity of trained assessors and consumers' intensity scores discriminating samples, the good correlations between textural attributes and the similar PCA configuration for IS and CDP methods supported the conclusion, but the RV coefficient was not applied. The authors also observed that the use of larger number of consumers compared with the trained assessors compensated the larger variability and lack of training of the RDM panel.

Ares et al. (2011) working with the same type of product reached a different conclusion. Although global panel performance for IS method was confirmed through ANOVA model, the poor individual assessors' performance of consumers from IS panel was indicated by low agreement with the rest of the panel, by the spread distribution of assessors in the two dimensions of PCA map, besides the low discriminative capacity and reproducibility.

3.8. Repertory grid, free listing and free choice profile

Repertory Grid method (RG), Free Listing (FL) and Free Choice profile (FCP) are techniques found in only one study each when considering their validation against CDP.

The Repertory Grid Method (RG) is an elicitation technique that associates meanings with products as bipolar constructs/attributes (Russel and Cox 2004). In this method, assessors can be given three samples and they are asked to say in what ways two of them are similar and how they are different from the third. The technique requires from 50 to 100 assessors (Varela and Ares 2014) and is generally conducted in two sessions. In the first session, attributes are elicited with the described procedure and attribute definitions can be requested. Then, in a next session, assessors give intensity scores to samples for each of the elicited attributes. The procedure is very similar to Free Choice Profiling (FCP) once each assessor builds his/her set of attributes and a configuration of N objects in K dimensions is obtained (Moussaoui and Varela 2010; William and Langron 1984). The sensory profiles of hot beverages were not totally similar between RG and CDP in the study of Mossaoui and Varela (2010). RG formed similar clusters of samples, presented a strong RV coefficient (0.88) in comparison with CDP and generated a rich vocabulary. Nevertheless, according to the authors, the description of complex attributes as mouthfeel or after feel was not well detailed or not accordant in the RG panel.

Free Listing (FL) is a regular technique used in anthropology (Hough and Ferraris 2010) in which assessors are asked to write terms related to samples. The number of assessors is not defined and may range between 20 and 378 (Hough and Ferraris 2010). In the single study of validation (Dos Santos

et al. 2015) of FL, results indicated that the method is not similar to CDP, as concluded for CATA in the same study with fermented sausages. The different clusters formed by the methods supported the conclusion.

One of the first RDM developed and applied in sensory literature (Parolari 1994; Barcenas et al. 2003; Pérez Aparicio, Angeles Toledano Medina and Rosales 2007; Pérez Aparicio, Toledano Medina, and Lafuente Rosales, 2007; Guàrdia et al. 2010), Free Choice Profile (FCP) is a method based on the assumption that assessors present similar perception of the sensory characteristics but differ in the way how they describe these features (William and Langron 1984). Assessors develop their own vocabulary to describe the sensory perceptions and use this vocabulary to rate the characteristics of a set of samples (Jack and Piggott 1991). The number of assessors for FCP is similar to CDP (10 to 12). Richter et al. (2010) compared FCP with CDP with chocolate puddings as samples. FCP panel presented good repeatability for all samples but the percentage of information on principal components was low (38%). The analysis of the general configuration and residual variance of assessors of each panel demonstrated that there was consensus in FCP and CDP. Comparing the samples distribution in FCP and CDP similarities were observed in configuration (primarily in dimension 1) and for the most discriminating attributes. However, aroma and texture descriptors were less consensual in FCP panel.

Varela and Ares (2012) review emphasized that some RDM cannot be considered as substitutes for conventional descriptive analysis with trained assessors. In the present systematic review, lack of similarity was found by Dos Santos et al. (2015), which analyzed dry fermented sausages using CATA and FL; by Moussaoui and Varela (2010), which analyzed hot beverages by Napping and Sorting; and by Torri et al. (2013) that evaluated wines using Napping.

3.9. Rate-all-that-aplly (RATA)

Rate-all-that-apply (RATA) questions are a variation of check-all-that-apply (CATA) where consumers indicate whether the terms in a list apply to describe a product and, if the term is selected, the consumer also indicates the intensity using a scale of 3 to 5 points (Ares et al. 2014). Oppermann et al. (2017) found similarities between RATA and CDP. The results found in their study suggest that RATA intensity scores with untrained raters show discriminative ability when compared with trained CDP raters. In this way the RATA can be used to measure the perceived intensity of the sensory terms using non-trained evaluators when samples with small differences are evaluated. These conclusions agree with the comparison of MFA analysis maps, which found similar maps, and maps related to RATA intensity are more similar with CDP than RATA frequency maps. The RV values also support this conclusion.

3.10. Polarized projective mapping (PPM)

The PPM is a combination of Projective Mapping (PM) and Polarized Sensory Positioning (PSP), in which assessors must situate the samples on a sheet of paper in which 3 reference samples or poles have been previously settled. The three poles must be previously tried and then each sample; to locate each of them in the sheet according to their similarities and differences in comparison with the references. Assessors could describe the samples to understand why they differ or not and to obtain consumers' vocabulary regarding the product (Ares et al. 2013).

Horita et al. (2017) compared PPM with CDP and found that trained CDP and untrained PPM evaluators were able to discriminate the samples in a similar way, and it was possible to reach this conclusion by comparing the maps generated by the MFA and by the value of the coefficient Rv (0.649), which was significant.

3.11. Risk of bias (RB) assessment

RB assessment of individual studies is part of assessing the strength of a body of evidence (Viswanathan et al. 2012). We formulated a tool based on classical and well stablished literature criteria for bias control in sensory studies and recommended statistical analyses to check if each study has properly performed the sensory tests thus indicating if it is shortly or highly exposed to RB.

Appendix C (Supplementary electronic material) presents a table with detailed evaluation of each RB criteria for each study, resulting in the final RB assessment of Table 2. Low and moderate RB evaluations were reached by 87% and 13% of the studies, respectively. It shows that the conclusions of this systematic review are well supported, according to the evaluation of bias control and statistical procedures conducted and reported in the 23 eligible studies.

The use of "randomized 3-digit coded samples" is a basic requirement to assure the control of the error of expectation (Stone and Sidel 2004; Lawless and Heymann 2010; Meilgaard et al. 2006). This is a usual tool to blind the samples so that assessors do not have extra information about them during sensory tests. Five studies did not mention the use of this procedure in methods and were classified as "unclear", whereas the other 78% of the studies reported the procedure.

Seventy-four percent of the studies used "appropriate number of samples per session". Excessive number of samples per session is associate to the bias of sensory fatigue or adaptation (Stone and Sidel 2012; Meilgaard et al. 2006; Lawless and Heymann 2010), impairing sensory results. The ideal number of samples is not well stated in the literature and depends on the nature of the product, the number of questions and the experience of the panelists (Stone and Sidel 2004; Lawless and Heymann 2010; Meilgaard et al. 2006). Generally, until ten samples per session is considered an appropriate number in CDP, being therefore the criterion for this review.

The "balanced order of presentation of coded samples for both tests" is another important control of bias procedure related to the minimization of psychological and physiological biases, such as expectancy error, and the effects of contrast, group, standard, central tendency or positioning, time-order, habituation and anticipation or adaptation (Stone and Sidel 2004; Meilgaard et al. 2006; Lawless and Heymann 2010). This requisite was met for 78% of the studies.

All the studies also reported the "recruitment and selection of assessors", a very important step to obtain reliable results. For instance, people who do not consume a product are not

suitable to evaluate it in a sensory test. Furthermore, for QDA or CDP, there are certain requirements such as availability, ability to perceive differences within that class of products, ability to verbalize and work as a group, ability to use scales of intensity (Stone and Sidel 2004), that must be raised by means of people screening protocols before participation.

The "appropriate number of assessors for CDP" was reached by all studies. Stone and Sidel (2012) recommends a QDA panel with ten to twelve assessors. Heymann et al. (2012) concluded that at least 8 assessors, and preferably 10, are sufficient for a CDP.

The "Statistical validation of trained panel", an important tool to assure that assessors present sample discrimination, repeatability and panel coherence, is requisite for a gold standard descriptive sensory method. Although much importance is given to the number and duration of sessions for training and descriptors elicitation, for two studies (Kennedy et al. 2009; Albert et al. 2011), the occurrence of validation is unclear or not mentioned and therefore was not positively classified in this criterion. Some studies reported the use of ISO 8586:2012 (ISO 2012) and then we considered that correct statistical procedures were applied to select assessors after training sessions (Ares et al. 2015; 2011; Bruzone et al. 2012). In eight studies (Rodrigue et al. 2000; Delarue et al. 2004; Cartier et al. 2006; Dehlholm et al. 2012; Torri et al. 2013; Mielby et al. 2014; Ares et al. 2015, Horita et al. 2017), it was unclear whether this criterion was met, and in 57% of the articles this criterion has been accomplished.

All studies used "appropriate number of samples to the applied data analysis and vice versa", which is important to the reliability of the statistical results. Multivariate data may be non-quantitative (categorical or ordinal) or quantitative (range or ratio scale) and statistical analyzes must take into account the type of data collected. Sixty-five percent of the studies applied appropriate statistical method in relation to the type of data. A controversial common practice, however, lies on the use of GPA for ranking data, like in Flash Profile and Ranking Descriptive Analysis. Among the studies that used FP and RDA, all, except for Dehlholm et al. (2012), used GPA to analyze data. Although some authors justify the use of GPA to ranking data (Rodrigue et al. 2000), GPA is applicable to quantitative data (Lawless 2013), and the use of MFA by Dehlholm et al. (2012) is the most appropriate method, given that it is suitable for different types of data (including ordinal), and not only quantitative (such as GPA).

The "percentage of explanation of CDP maps", which was accepted as higher than 70% (Favero et al. 2009) was achieved by 83% of the studies. Cartier et al. (2006) e Richter et al. (2010) presented, respectively, 66% e 46% in the first two dimensions of PCA map. This percentage was unclear in the study of Delholm et al. (2012).

On the other hand, only 48% of the studies applied "RV coefficient in relation to Conventional Descriptive Profile" to evaluate correlation between the perceptual map coordinates from the rapid method and from the gold standard.

RV-coefficient, a multivariate generalization of the correlation coefficient, is calculated to provide a global index of the proximity of the two factorial configurations (Robert and Escoffier 1976). According to Cartier et al. (2006); Nestrud and Lawless (2008), an excellent agreement between configurations corresponds to a coefficient value close to 1. However, as for the correlation coefficient, a statistical test is required to assert significance of RV coefficient (Josse, Pagès and Husson 2008). The RV coefficient was used in eleven out of twenty-three studies, and only six of them reported to perform tests of significant for RV (Dehlholm et al. 2012; Cartier et al. 2006; Ares et al. 2015; Oppermann et al. 2017; Moelich et al. 2017; Horita et al. 2017). In all studies, the description of the RV coefficient is relative to the coordinate samples, except for the study of Kennedy and Heymann (2009) in which it is not clear if samples or attributes are used in RV analysis. A limitation of using only the samples' coordinates is that high and significant RV coefficients do not necessarily mean similar sensory profiles, but similar samples configurations that could be grouped by different criteria or diverse perceived characteristics. As the elicited attributes are usually different between CDP and RDM, calculation of the RV on attributes is unfeasible and therefore it is not appropriate to conclude regarding similar sensory profiles.

4. Conclusion

Systematic Review was a feasible tool to evaluate the validation of RDM against CDP. Twenty-three studies currently support the state of the art of comparison of rapid methods against the conventional profile The similarities of the studies by method were: Flash Profile, 67%; Sorting and derivatives, 60%; Projective Mapping and derivatives, including ultra-flash profile, 56%; IS, 50%; and CATA, 33%, Ranking Descriptive Analysis, Optimized Descriptive Profile, Polarized Projective Mapping and Rate-all-that-apply 100%. Validation was supported by many statistical strategies, but mainly by the visual inspection of the perceptual maps (47%), RV coefficients between samples configurations (67%) and comparison of the percentages of explanations of the first two maps' dimensions (27%). Risk of bias was low for 87% of the studies and moderate for 13%, showing general good quality to support their conclusions. Although CDP is a robust method to describe food, RDM are important alternatives as they can also be applied for other purposes, such as obtaining consumers language. Regarding their correspondence to CDP to describe food, conclusions on validity must be supported by more studies in many of the RDM evaluated, and with more food categories.

Acknowledgments

We acknowledge Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the scholarship support.

ORCID

Lauro Melo http://orcid.org/0000-0001-5410-8752

References

Albert, A., P. Varela, A. Salvador, G. Hough, and S. Fiszman 2011. Overcoming the issues in the sensory description of hot served

food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality and Preference* 22:463–73. doi:10.1016/j.foodqual.2011.02.010.

Ares, G., L. Antúnez, F. Bruzzone, L. Vidal, A. Giménez, B. Pineau, M. K. Beresford, D. Jin, A. G. Paisley, S. L. Chheang, C. M. Roigard, and S. R. Jaeger. 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. doi:10.1016/j.foodqual.2015.05.007.

Ares, G., F. Bruzzone, and A. Giménez 2011. Is a consumer panel able to reliably evaluate the texture of dairy desserts using unstructured intensity scales? Evaluation of global and individual performance. *Journal of Sensory Studies* 26:363–70. doi:10.1111/j.1745-459X.2011.00352.x.

Ares, G., F. Bruzzone, L. Vidal, R. S. Cadena, A. Giménez, B. Pineau, D. C. Hunter, A. G. Paisley, and S. R. Jaeger. 2014. Evaluation of a rating-based variant of check-all-that-apply questions: Rate-allthat- apply (RATA). *Food Quality and Preference* 36:87–95. doi:10.1016/j.foodqual.2014.03.006.

Ares, G., L. de Saldamando, L. Vidal, L. Antúnez, A. Giménez, and P. Varela. 2013. "Polarized Projective mapping: Comparison with polarized sensory positioning approaches." *Food Quality and Preference* 28 (2):510–518. doi:10.1016/j.foodqual.2013.01.003.

Ares, G., A. Tárrega, L. Izquierdo, and S. R. Jaeger 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* 31:135–41. doi:10.1016/j.foodqual.2013.08.012.

Bárcenas, P., F. J. Pérez Elortondo, and M. Albisu. 2003. "Comparison of free choice profiling, direct similarity measurements and hedonic data for ewes' milk cheeses sensory evaluation." *International Dairy Journal* 13 (1):67–77. doi:10.1016/S0958-6946(02)00139-5.

Bruzzone, F., G. Ares, and A. Giménez 2012. Consumers' texture perception of milk desserts II- Comparison with trained assessors' data. *Journal of Texture Studies* 43:214–26. doi:10.1111/j.1745-4603.2011.00332.x.

Cardinal, M., R. Baron, C. Kouakou, C. Prost, and P. Courcoux 2014. Comparative value of a sorting procedure and quantitative descriptive analysis to investigate the influence of processing parameters: Case study of hydrolysate production from salmon by-products. *Journal Of Sensory Studies* 29 (2):159–70. doi:10.1111/joss.12091.

Cartier, R., A. Rytz, A. Lecomte, E. Poblete, J. Krystlik, E. Belin, and N. Martin. 2006. Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference* 17 (7–8):562–71. doi:10.1016/j.foodqual.2006.03.020.

Dairou, V., and J.-M. Sieffermann 2002. A comparison of 14 jams characterized by conventional profile and a quick original method, flash profile. *Journal of Food Science* 67:826–34. doi:10.1111/j.1365-2621.2002.tb10685.x.

Deeks, J. J. 2001. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal* 323 (7305):157–62. doi:10.1136/bmj.323.7305.157.

Dehlholm, C., P. B. Brockhoff, L. Meinert, M. D. Aaslyng, W. L. P. Bredie 2012. Rapid descriptive sensory methods – Comparison of free multiple sorting, partial napping, napping, flash profiling and conventional profiling. *Food Quality and Preference* 26 (2):267–77. doi:10.1016/j.foodqual.2012.02.012.

Delarue, J., and J. M. Sieffermann 2004. Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference* 15:383–92. doi:10.1016/S0950-3293(03)00085-5.

Delarue, Julien, Ben Lawlor, and Michel Rogeaux 2014. Rapid Sensory Profiling Techniques. 1st edn. Applications in New Product Development and Consumer Research. Delarue, J. "1 - The use of rapid sensory methods in R & D and research: an introduction." In Rapid Sensory Profiling Techniques, 3–25. Cambridge: Woodhead Publishing.

Devillé, W. L., F. Buntinx, L. M. Bouter, V. M. Montori, H. C. de Vet, D. A. van der Windt, and P. D. Bezemer 2002. Conducting systematic reviews of diagnostic studies: Didactic guidelines. *BMC Medical Research Methodology* 2 (1):1–13. doi:10.1186/1471-2288-2-9.

Dos Santos, B. A., P. C. Bastianello Campagnol, A. G. da Cruz, M. T. E. L. Galvão, R. A. Monteiro, R. Wagner, and M. A. R. Pollonio (2015). Check all that apply and free listing to describe the sensory characteristics of low sodium dry fermented sausages: Comparison with trained panel. *Food Research International* 76:725–34. doi:10.1016/j.foodres.2015.06.035.

European Food Safety Authority. 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal* 2010 8:1637. doi:10.2903/j.efsa.2010.1637.

Favero, L. P., P. Belfiore, F. L. Silva, and B. L. Chan 2009. *Análise de dados. Modelagem multivariada para tomada de decisões*. Rio de Janeiro: Elsevier.

Guàrdia, M. D., A. P. S. Aguiar, A. Claret, J. Arnau, and L. Guerrero. 2010. "Sensory characterization of dry-cured ham using free-choice profiling." *Food Quality and Preference* 21 (1):148–55. doi:10.1016/j.foodqual.2009.08.014.

Hemingway, P., and N. Brereton 2009 What is a systematic review?, What is…? series [online]. URL: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Systreview.pdf [accessed 02.02.16] doi:10.1016/j.foodqual.2009.08.014.

Heymann, H., B. Machado, L. Torri, and A. L. Robinson 2012. How many judges should one use for sensory descriptive analysis? *Journal of Sensory Studies*. 27 (2):111–22. doi:10.1111/j.1745-459X.2012.00373.x.

Higgins, J. P. T., and S. Green 2011 (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Higgins, J. P. T., D. G. Altman, and J. A. C. Sterne 2011. (editors). Chapter 8: Assessing risk of bias in included studies. In: J. P. T., Higgins, S., Green (editors). Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available from www.handbook.cochrane.org.

Horita, C. N., E. A. Esmerino, V. A. S. Vidal, J. S. Farah, G. V. Amaral, H. M. A. Bolini, A. G. Cruz, and M. A. R. Pollonio 2017. "Sensory profiling of low sodium frankfurter containing garlic products: Adequacy of polarized projective mapping compared with trained panel." *Meat Science* 131:90–98. doi:10.1016/j.meatsci.2017.05.002.

Hough, G., and D. Ferraris 2010. Free listing: A method to gain initial insight of a food category. *Food Quality and Preference* 21:295–301. doi:10.1016/j.foodqual.2009.04.001.

ISO 2012. *Sensory analysis—General guidance for the selection, training, and monitoring of selected assessors and expert sensory assessors*. ISO Standard 8586:2012. Geneva, Switzerland: International Organization for Standardization.

Jack, F. R., and J. R. Piggott 1991. Free choice profiling in consumer research. *Food Quality and Preference* 3:129–34 doi:10.1016/0950-3293(91)90048-J.

Jiménez, M. J., W. Canet, and M. D. Alvarez 2013. Sensory description of potato puree enriched with individual functional ingredients and their blends. *Journal of Texture Studie* 44 (4):301–16. doi:10.1111/jtxs.12024.

Josse, J., J. Pagès, and F. Husson 2008. Testing the significance of the RV coefficient. *Computational Statistics & Data Analysis* 53 (1):82–91. doi:10.1016/j.csda.2008.06.012.

Kennedy, J., and H. Heymann 2009. Projective mapping and descriptive analysis of milk and dark chocolates. *Journal of Sensory Studies* 24:220–33. doi:10.1111/j.1745-459X.2008.00204.x.

Lagast, S., X. Gellynck, J. J. Schouteten, V. De Herdt, and H. De Steur. 2017. Consumers' emotions elicited by food: A systematic review of explicit and implicit methods. *Trends in Food Science & Technology* 69:172–189. doi:10.1016/j.tifs.2017.09.006.

Lawless, H. T., and H. Heymann 2010. *Sensory evaluation of food: Principles and practice*. (2nd ed.). New York: Springer.

Lawless, H. T. 2013. Quantitative sensory analysis. Oxford: Wiley-Blackwell, (Chapter 10).

Leeflang, M. M., J. J. Deeks, Y. Takwoingi, and P. Macaskill. 2013. Cochrane diagnostic test accuracy reviews. *Syst Rev.* 7 (2):82. doi:10.1186/2046-4053-2-82.

Mielby, L. H., H. Hopfer, S. Jensen, A. K. Thybo, and H. Heymann 2014. Comparison of descriptive analysis, projective mapping and sorting performed on pictures of fruit and vegetable mixes. *Food Quality and Preference* 35:86–94. doi:10.1016/j.foodqual.2014.02.006.

Minim, V. P. R., and RdedsS.Nda. Silva 2016. *Análise sensorial descritiva*. Ed: UFV.

Moelich, E. I., M. Muller, E. Joubert, T. Næs, and M. Kidd 2017. "Validation of projective mapping as potential sensory screening tool for application by the honeybush herbal tea industry." *Food Research International* 99:275–86. doi:10.1016/j.foodres.2017.05.014.

Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman. 2009. The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLoS Med* 151:264–69.

Moussaoui, K. A., and P. Varela 2010. Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality and Preference* 21:1088–99. doi:10.1016/j.foodqual.2010.09.005.

Nestrud, M. A., and H. T. Lawless 2008. Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professional and consumers. *Food Quality and Preference* 19:431–38 doi:10.1016/j.foodqual.2008.01.001.

Oppermann, A. K. L., C. de Graaf, E. Scholten, M. Stieger, and B. Piqueras-Fiszman 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. doi:10.1016/j.foodqual.2016.09.010.

Pagés, J. 2005. "Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley." *Food Quality and Preference* 16:642–649. doi:10.1016/j.foodqual.2005.01.006.

Parolari, G. 1994. Special issue: Advances in sensory food science – rose marie pangborn memorial symposium taste quality of Italian raw ham in a free-choice profile study. *Food Quality and Preference* 5 (1): 129–33. doi:10.1016/0950-3293(94)90018-3.

Pérez Aparicio, J., M. Angeles Toledano Medina, and V. L. Rosales. 2007. Descriptive sensory analysis in different classes of orange juice by a robust free-choice profile method. *Analytica Chimica Acta* 595 (1–2):238–47. doi:10.1016/j.aca.2007.02.054.

Pickup, W., P. Bremer, and M. Peng. 2017. "Comparing conventional descriptive analysis and napping®-UFP against physiochemical measurements: A case study using apples." *Journal of the Science of Food and Agriculture*. doi:10.1002/jsfa.8616.

Pineli, LdLdO., S. Rodrigues Jda, A. M. Costa, H. C. de Lima, M. D. Chiarello, and L. Melo 2014. Antioxidants and sensory properties of the infusions of wild passiflora from Brazilian savannah: potential as functional beverages. *Journal of the Science of Food and Agriculture* 95 (7):1500–1506.

Pineli, L., G. Oliveira, M. Mendonça, L. Borgo, É. Freire, S. Celestino, M. Chiarello, and R. Botelho. 2015. Tracing chemical and sensory characteristics of baru oil during storage under nitrogen. *LWT – Food Science and Technology* 62 (2):976–82. doi:10.1016/j.lwt.2015.02.015.

Richter, V. B., T. C. A. de Almeida, S. H. Prudencio, and M. B. de Toledo (2010). Proposing a ranking descriptive sensory method. *Food Quality and Preference* 21 (6):611–20. doi:10.1016/j.foodqual.2010.03.011.

Robert, P., and Y. Escoufier 1976 A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied Statistics* 25:257–65. doi:10.2307/2347233.

Rodrigue, N., M. Guillet, J. Fortin, and J. F. Martin 2000. Comparing information obtained from ranking and descriptive tests of four sweet corn products. *Food Quality and Preference* 11:47–54. doi:10.1016/S0950-3293(99)00063-4.

Russel, C. G., and D. N. Cox 2004. Understanding middle-aged consumers' perception of meat using repertory grid methodology. *Food Quality and Preference* 15:317–29. doi:10.1016/S0950-3293(03)00073-9.

Silva, RdCdS.Nd., V. P. R. Minim, JdeD. S. Carneiro, M. Nascimento, S. M. Della Lucia, and L. A. Minim 2013. Quantitative sensory description using the optimized descriptive profile: Comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality and Preference* 30 (2):169–79. doi:10.1016/j.foodqual.2013.05.011.

Silva, R., V. P. R. Minim, A. A. Simiquel, L. E. da Silva Moraes, A. I. Gomide, and L. A. Minim 2012. Optimized descriptive profile: a rapid methodology for sensory description. *Food Quality and Preference* 24:190–200. doi:10.1016/j.foodqual.2011.10.014.

Simiqueli, A. A., V. P. R. Minim, R. D. C.D S N. D. Silva, A. N. D. Silva, and L. A. Minim 2015. How many assessors are necessary for the optimized descriptive profile when associated with training? *Food Quality and Preference* 44:62–69. doi:10.1016/j.foodqual.2015.03.019.

Stone, H. 2012. *Sensory Evaluation Practices*. Elsevier Science.

Stone, H., and J. Sidel 2004. *Sensory Evaluation Practices*. San Diego,CA: Academic Press, Inc. p. 377

Stone, H., J. Sidel, S. Oliver, A. Woosley, and R. C. Singleton 1974. Sensory evaluation by quantitative descriptive analysis. *Food Technology* 28: 24–34.

Torri, L., C. Dinnella, A. Recchia, T. Naes, H. Tuorila, and E. Monteleone 2013. Projective mapping for interpreting wine aroma differences as perceived by naïve and experienced assessors. *Food Quality and Preference* 29 (1):6–15. doi:10.1016/j.foodqual.2013.01.006.

Valentin, D., S. Chollet, M. Lelievre, and H. Abdi 2012. Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International Journal of Food Science and Technology* 47:1563–78. doi:10.1111/j.1365-2621.2012.03022.x.

Varela, P., and G. Ares 2012. Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International* 48:893–908. doi:10.1016/j.foodres.2012.06.037.

Varela, P., and G. Ares 2014. *Novel techniques in sensory characterization and consumer profiling*. Boca Raton: CRC Press.

Viswanathan, M., M. T. Ansari, N. D. Berkman, S. Chang, L. Hartling, L. M. McPheeters, P. L. Santaguida, T. Shamliyan, K. Singh, A. Tsertsvadze, and J. R. Treadwell 2012. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. March 2012. *AHRQ Publication* No. 12-EHC047-EF. Available at: www.effectivehealthcare.ahrq.gov.

Williams, A. A., and S. P. Langron 1984. "The use of free-choice profiling for the evaluation of commercial ports." *Journal of the Science of Food and Agriculture* 35 (5):558–68. doi:10.1002/jsfa.2740350513.