soder John Control Con

Original scientific paper

Use of attribute agreement analysis (AAA) in the validation of sensory evaluation methods: Case study for the visual determination of parasites in fish

Zoran Petrović^{1*}, Jelena Ćirić¹, Jelena Babić Milijašević¹, Milan Milijašević¹, Mirjana Lukić¹, Jelena Jovanović¹ and Aleksandra Nikolić¹

¹ Institute of Meat Hygiene and Technology, 11000 Belgrade, Serbia

ARTICLE INFO

Keywords: Attribute agreement analysis (AAA) Sensory methods Validation Fish Parasites

ABSTRACT

Validation of sensory evaluation employs a process similar to any other method validation procedure in analytical chemistry. However, the parameters often measured in sensory testing are different, because the human sensory apparatus and brain are the instruments being calibrated. A total of 10 fish samples were examined by an expert 5-member evaluation panel for the visual presence of parasites in frozen fish. From each frozen hake sample, a group of six slices of fish muscle was formed by longitudinal sectioning. The total of 60 sliced samples were divided into 10 plates with six samples each. The first plate contained six slices, each of which contained parasites. The attribute agreement analysis showed strong agreement in the overall ratings. It was found that at least 75% of all tests achieved the highest level of agreement. The results of the tests were presented in the form of tables and graphs summarizing the subjective test results using Fleiss' Kappa and Cohen's Kappa statistics.

1. Introduction

Attribute agreement analysis is a statistical method used to determine if trained expert sensory panels are using a particular scale consistently and in the same way (*MoreSteam*, 2024). This method is used in the validation of sensory tests and is widely adopted in the food industry. In order to use attribute agreement analysis, it is essential that the objects, reference standards, and the rating scales themselves are precisely defined. The use of such sensory expert panels is highly regulated, and their fitness for purpose and the methods they use must be scientifically validated (*Djekic et al.*, 2021; *Sipos et al.*, 2021).

Sensory evaluation methods, such as the one validated in this case study, usually employ the use of trained expert panels who are experienced in the subject area. Such panels are often used in the food and drink industries to assess product attributes, and

in new product development and quality control, as well as being required for the labeling and marketing of products. Use of such expert sensory panels is highly regulated, and their fitness for purpose and the methods they employ have to be scientifically validated (*Barbieri et al.*, 2020; *Da Costa et al.*, 2020; *Mihafu et al.*, 2020; *Mohammadi-Moghaddam and Firoozzare*, 2021; *Gupta et al.*, 2022).

When a consistent and valid method is developed and used, not only is the capability of the product to meet consumer expectations and preferences increased, but the improvements and changes made to the product can be scientifically shown to result in an improved sensory output (*Quintão et al.*, 2020; *Stone et al.*, 2020; *Vivek et al.*, 2020). This is important to both the producer and the consumer of the product, as it demonstrates the validity and trustworthiness of the sensory results. Furthermore, the use of validated methods will

^{*}Corresponding author: Zoran Petrović, zoran.petrovic@inmes.rs

strengthen the possibility of making and defending scientific claims about a product (*Pavli et al.*,2020; *Rose and Johnson*, 2020; *Sürücü and Maslakci*, 2020). The use of qualitative methods of sensory evaluation, such as for parasites in fish, requires certainty that the method is reproducible and repeatable and, more importantly, that the assessors have a high agreement rate (*Freitas et al.*, 2020; *Jurica et al.*, 2021).

Attribute agreement analysis has a lot of benefits, key among which is its ability to remove subjectivity in the validation process. First of all, validation using this method is not dependent on the knowledge of a particular expert in a certain field. Also, attribute agreement analysis removes individualism. This is because what may be of high severity to one assessor may not necessarily be so to another. By matching the severity ratings and looking at the percentage agreement, it is possible to tell if the assessors are in agreement in as far as qualitative grading is concerned (Xiong et al., 2020). This method is used in the validation of sensory tests and is widely adopted in the food industry (Carpenter et al., 2000). In order to use attribute agreement analysis, it is essential that the objects, reference standards, and the rating scales themselves are precisely defined (*Hubbard*, 2012).

The focus in this paper was on how attribute agreement analysis can be used to validate a sensory examination method. The goal of validating the sensory method was to demonstrate the appropriateness of adopting such a method in order to fulfil accreditation requirements (*ISO 17025*, 2017). For the parasites in fish, we wanted to show that the method is consistent, precise, and less subjective or more sensitive (or specific) than alternative methods (*Zhang et al.*, 2022).

In addition, the validation should show that the detection method actually leads to the correct conclusion regarding the presence or absence of the parasite in the fish. For example, the method should be able to show that the tested fish sample, which was declared parasite-free, actually does not contain any parasites. If, on the other hand, the sample is classified as parasite-positive, the validation should show that the method provides a correct result.

2. Materials and Methods

Prior to validating a method using attribute agreement analysis, the first step of training the assessors using an appropriate presentation, known as a consensus building session, was conducted. This session aimed to set the standards of the method and reduce the variability of the data.

2.1. Sample preparation

The validation sample consisted of 60 slices made from two groups of frozen hake one of which tested positive and the other negative for parasites.

The frozen fish were thawed at room temperature for 6 h until they became a suitable texture for cutting fillets in the semifrozen state (-3 °C). From each fish sample, a group of six slices of fish muscle was formed by longitudinal sectioning. The total of 60 sliced samples were divided onto 10 plates with six samples each. The first plate contained six slices, each of which contained visible parasites (Figure 1).

The other plates with serial numbers 2–10 each contained six slices without parasites (Figure 2).

The visual inspection was carried out by a sensory panel consisting of five trained appraisers, doctors of veterinary medicine. The visual inspection results were recorded by each appraiser individually

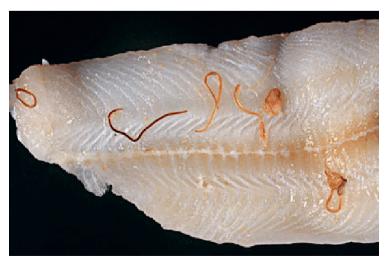


Figure 1. Fish samples with visible parasites situated on plate.

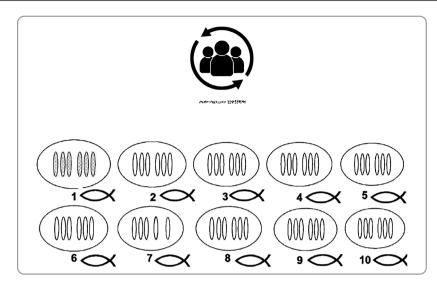


Figure 2. Scheme of validation experiment

and after the experiment were processed using appropriate statistical tools.

2.2. Temperature control

Temperature of filets was measured using digital thermometer, model, TESTO 926-1 (Germany), equipped with a wi-fi puncture probe.

2.3. Environmental and room conditions

When performing the method validation, the recommendations of the ISO 17025:2017 standard, point 6.3, were observed, and the ambient conditions were monitored and recorded. The room temperature was 20±1 °C and the relative humidity of the room (rH) was 64%. The room was illuminated

with the prescribed ambient lighting of 220 lux. The light intensity in the room was measured with a lux light meter (MMS Med Lab, UK).

2.4. Statistical analysis

The obtained data were statistically processed in the statistical package MINITAB INC. VER. 17, USA (*Minitab*, 2024), using the tool within the option: STAT/QUALITY TOOLS/ATTRIBUTE AGREEMENT ANALYSIS.

In our study, this MINITAB tool was used to assess the agreement of subjective ratings or classifications given by the five appraisers; the tool can be used for nominal and ordinal data. The attribute agreement analysis worksheet was used to create the worksheet for this study (Figure 3).

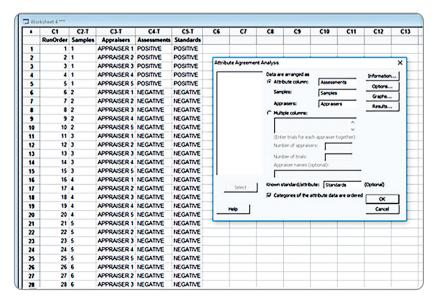


Figure 3. Layout of the statistical tool used

The data set was structured so that it was stacked in an attribute column. The results obtained from each appraiser were entered as text (positive/ negative) data. For the data in the attribute column, all responses were grouped into one column and columns were set up with grouping indicators for the appraiser and the sample number. The grouping indicators were used to define each sample. The confidence level for the interval estimate of the percentage agreement between appraisers and between each appraiser and the standard was set at 95%. We specified a column for known standards/attributes (expected outcome) in the main dialog to estimate how often each appraiser's judgments deviated from the known standard or attribute values.

3. Results and Discussion

The attribute agreement analysis output included graphical and numeric output in the forms shown in the text below. The statistical programme displayed three assessment agreement tables: Each appraiser vs standard; between appraisers, and; all appraisers vs standard.

3.1 Attribute agreement analysis for assessment reports

Samples: 10 Appraisers: 5 Replicates: 1 Total runs: 50 Date of study: 10 2023

Reported by: Head of sensory panel

Name of product: Fish

Misc: Method validation — Visual

determination of parasites in

fish

3.2 Each appraiser vs standard

Tables 1 to 4 show output tables from the MINITAB statistical programme, using operation 3.2. The parameter analyzed in a statistical test lies between the endpoints of the confidence limit interval. In this case, as shown in Table 1, 74.11% to 100% of the appraisers correctly identified the positive samples. In Table 2, and in accordance with the findings in Table 1, there were no deviations in the overall evaluation of the fish samples or of those fish samples with parasites that were previously declared as standard.

Appraiser	# Inspected	# Matched	Percent	95% CI
APPRAISER 1	10	10	100.00	(74.11, 100.00)
APPRAISER 2	10	10	100.00	(74.11, 100.00)
APPRAISER 3	10	10	100.00	(74.11, 100.00)
APPRAISER 4	10	10	100.00	(74.11, 100.00)
APPRAISER 5	10	10	100.00	(74.11, 100.00)

Table 1. Assessment agreement

Table 2. Assessment disagreement

Appraiser	# NEGATIVE/POSITIVE	Percent	# POSITIVE/NEGATIVE	Percent	# Mixed Percent
APPRAISER 1	0 0.00	0	0.00	0	0.00
APPRAISER 2	0 0.00	0	0.00	0	0.00
APPRAISER 3	0 0.00	0	0.00	0	0.00
APPRAISER 4	0 0.00	0	0.00	0	0.00
APPRAISER 5	0 0.00	0	0.00	0	0.00

[#] NEGATIVE/POSITIVE: Assessments across trials = NEGATIVE / standard = POSITIVE.

[#] Matched: Appraiser's assessment across trials agrees with the known standard.

[#] NEGATIVE/POSITIVE: Assessments across trials = POSITIVE / standard = NEGATIVE.

[#] Mixed: Assessments across trials are not identical.

Fleiss' kappa and Cohen's kappa scores are included in the statistic programme's output tables (Table 3 and Table 4, respectively). The higher the kappa score, the greater the agreement between appraisers, and the better the validation of the test (MINITAB, 2024). The statistical software calculates Cohen's kappa when two appraisers rate a single trial, or when one appraiser rates two trials. The Fleiss kappa coefficient theoretically ranges from -1 to +1. Values close to 1 indicate a strong agreement between the overall rating and the individual appraisers. Tables 3 and 4 show kappa coefficient values of 1 for each appraiser, indicating their full agreement with the overall rating. In our study, the results of Cohen's kappa test (Table 4) were identical to the Fleiss Kappa indices. A test of

significance and its p-value are displayed to indicate the significance of each result (P=0.00008).

3.3 Between appraisers

Figures 4 to 6 show the statistical programme's output data, using the appropriate statistical operation (Figure 3). In Figure 4, we plotted the results between the appraisers, who in this case each gave an identical answer by looking at all groups of fish slices distributed on 10 plates. The first plate contained six positive samples from the same sample population (in this study, this plate was also declared to be the standard), which gave us confidence intervals with the population parameter (positive parasite findings).

Table 3. Fleiss' kappa statistics

Appraiser	Response	Kappa SE Kappa	Z	P(vs > 0)
APPRAISER 1	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 2	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 3	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 4	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 5	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008

Table 4. Cohen's kappa statistics

Appraiser	Response	Kappa SE Kappa	Z	$P(vs \ge 0)$
APPRAISER 1	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 2	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 3	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 4	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008
APPRAISER 5	POSITIVE	1 0.316228	3.16228	0.0008
	NEGATIVE	1 0.316228	3.16228	0.0008

Between Appraisers

Assessment Agreement

```
# Inspected # Matched Percent 95% CI
10 10 100.00 (74.11, 100.00)
```

Matched: All appraisers' assessments agree with each other.

Fleiss' Kappa Statistics

Response	Kappa	SE	Kappa	Z	P(vs > 0)
POSITIVE	1		0.1	10	0.0000
NEGATIVE	1		0.1	10	0.0000

Figure 4. Results of comparison between appraisers

Matched: All appraisers' assessments agree with each other.

3.3.1 Assessment Agreement

The software is able to determine any deviation of a particular appraiser from the standard. However, this was not calculated in our study, as the current data are from only a single trial for each appraiser.

3.4 All appraisers vs standard

3.4.1 Assessment Agreement

The software output also displays a graph of the confidence intervals (CI), comparing each appraiser against the standard, as shown in Figure 6.

This figure shows the CIs, comparing each appraiser (1–5) against the standard (fish samples on plate No. 1 (Figure 1) and that were declared to be positive (so did contain parasites).

It should also be noted that each matching percentage is associated with a confidence interval (Figure 6). The results of the statistical operation, presented in Table 1, show complete agreement of the results of the evaluation of all samples by the appraisers, including the samples declared as the standard (the first plate with all six positive samples) at the confidence level of 95% within the corresponding confidence interval (CI).

All Appraisers vs Standard

Assessment Agreement

```
# Inspected # Matched Percent 95% CI
10 10 100.00 (74.11, 100.00)
```

Matched: All appraisers' assessments agree with the known standard.

Fleiss' Kappa Statistics

```
Response Kappa SE Kappa Z P(vs > 0)
POSITIVE 1 0.141421 7.07107 0.0000
NEGATIVE 1 0.141421 7.07107 0.0000
```

Cohen's Kappa Statistics

```
Response Kappa SE Kappa Z P(vs > 0)
POSITIVE 1 0.141421 7.07107 0.0000
NEGATIVE 1 0.141421 7.07107 0.0000
```

Figure 5. Results of comparison between all appraisers vs standard

Measurements in sensory analysis (by humans) are subjective assessments by people rather than direct physical measurements. In these situations, the quality characteristics are difficult to define and evaluate (Nute, 2010). To obtain meaningful classifications, more than one appraiser should classify the response measure. If the appraisers agree, there is a possibility that the ratings are correct. If the appraisers disagree, the usefulness of the rating is limited. In this method, each of the appraisers rates or grades a series of samples. The sensory measurement depends on not only the human factors of the sensory appraisers, such as their experience and acuteness of sense, but also on the laboratory environment, the experimental condition, and the presentation of the test samples. In practice, the traditional validation methods are usually time-consuming and costly, and validation of sensory methods often becomes just an administrative burden to the quality assurance personnel.

4. Conclusions

It is clear that the validation of sensory evaluation methods is important, as it separates the field of sensory science from an industry that relies heavily on guesswork and subjectivity. A statistical kappa analysis was performed to determine the degree of agreement between each subjective assessment given by the appraisers and the actual parasitic status (positive or negative) of the fish samples. At the end of the analysis, the subjective assessment that had the highest agreement with the objective measurement was determined. It was found that, statistically, at least 75% of all appraisals achieved the highest level of agreement with the objective status of the fish samples, which allowed the attribute's discriminatory ability in describing sensory differences to be verified. Therefore, this case study validated the sensory evaluation method, based on attribute agreement.

One of the main advantages of attribute agreement analysis is that it provides a quick and accurate method for evaluating both the sensory method and the attributes. In addition, the graphical representation provided by the MINITAB statistical programme provides information about the quality of the appraisers and whether the scale is correctly calibrated or not.

By assessing the strength of agreement among appraisers using well-categorized and standardized measurement criteria, attribute agreement analysis helps us to establish the degree of the quality of the sensory evaluation. This makes the evaluation

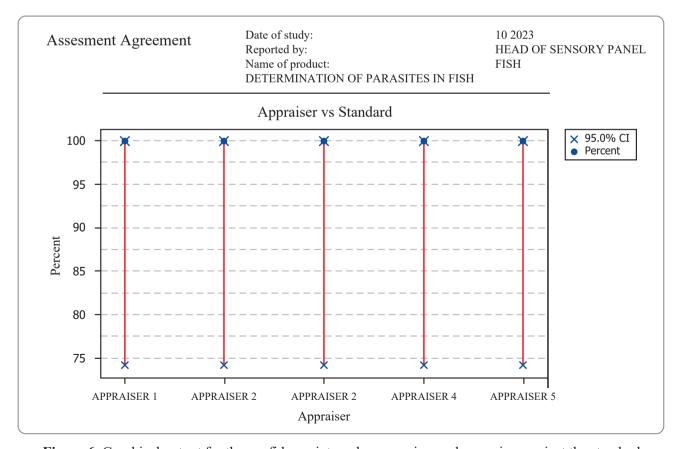


Figure 6. Graphical output for the confidence intervals comparing each appraiser against the standard

results more reliable and acceptable. In our case, the method should be able to show that no visible parasites are actually present in cases where the fish sample tested is declared parasite-free. If, on the other hand, a fish sample is classified as parasite-positive, the validated method should also produce a correct result stating that.

Finally, this tool can be used to check whether the measurement error is at an acceptable level before performing a data analysis. The attribute agreement analysis quantifies three types of variations: variations within the repeated measurements of a single appraiser (repeatability); variations between the measurements of different appraisers, (reproducibility) including corresponding confidence intervals for both of the measuring characteristics, and; variations between an appraiser's measurements and a reference or standard.

Upotreba Attribute Agreement Analysis – AAA u validaciji senzornih metoda ispitivanja: Studija slučaja za vizuelno određivanje parazita u ribama

Zoran Petrović, Jelena Ćirić, Jelena Babić Milijašević, Milan Milijašević, Mirjana Lukić, Jelena Jovanović i Aleksandra Nikolić

INFORMACIJE O RADU

Ključne reči: Analiza saglasnosti atributa (AAA) Senzorne metode Validacija Ribe Paraziti

APSTRAKT

Validacija senzornih metoda ispitivanja prema zahtevima treba da koristi proces sličan svakom drugom postupku validacije metoda u npr. hemijskim ispitivanjima. Međutim, parametri koji se često mere u senzornom testiranju su različiti, jer su ljudski senzorni aparat i um instrumenti koji se "kalibrišu". Ukupno 10 uzoraka ribe je ispitano od strane petočlanog panela za procenu vizuelnog prisustva parazita u smrznutom osliću. 1 uzorak ribe je izabran iz prve grupe koja je imala parazite, dok je 9 uzoraka riba izabrano iz druge grupe bez parazita. Od svakog smrznutog uzorka oslića, uzdužnim sečenjem formirana je grupa od 6 tankih odsečaka mišića ribe. Test uzorci za ispitivanje su obuhvatili 60 isečenih uzoraka koji su raspoređeni u 10 tanjira sa po 6 narezanih uzoraka. Prvi tanjir je sadržao 6 tankih fileta, od kojih je svaki sadržao parazite. Rezultati primenjenog statističkog pristupa pokazali su jako slaganje u ukupnim ocenama. Statističkom obradom u ovoj studiji utvrđeno je da je najmanje od 75% do 100 % (interval poverenja) svih pojedinačnih testova članova panela postiglo najviši nivo saglasnosti. Rezultati statističkih testova su predstavljeni u obliku tabela i grafikona koji sumiraju subjektivne rezultate testa korišćenjem Fleiss Kappa i Cohen Kappa statističke.

Disclosure statement: No potential conflict of interest was reported by the authors.

Funding: This study was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia, (Grant No. 451-03-66/2024-03/200050 dated 05.02.2024).

References

Barbieri, S., Brkić Bubola, K., Bendini, A., Bučar-Miklavčič, M., Lacoste, F., Tibet, U., ... & Gallina Toschi, T. (2020). Alignment and proficiency of virgin olive oil sensory panels: The OLEUM approach. Foods, 9 (3), 355.

Carpenter, R. P. Lyon, D. H., & Hasdell, T. A. (2000). Guidelines for Sensory Analysis in Food Product Development and Quality Control. In: Putting sensory analysis into practice. Eds. Carpenter, R. P. Lyon, D. H.,

Hasdel, T. A., 129–141, Springer Science & Business Media, US.

Da Costa, G. M., De Paula, M. M., Costa, G. N., Esmerino, E. A., Silva, R., de Freitas, M. Q., ... & Pimentel, T. C. (2020). Preferred attribute elicitation methodology compared to conventional descriptive analysis: A study using probiotic yogurt sweetened with xylitol and added with prebiotic components. *Journal of Sensory Studies*, 35 (6), e12602.

- Djekic, I., Lorenzo, J. M., Munekata, P. E., Gagaoua, M., & Tomasevic, I. (2021). Review on characteristics of trained sensory panels in food science. Journal of Texture Studies, 52 (4), 501-509.
- Freitas, J., Vaz-Pires, P., & Câmara, J. S. (2020). From aguaculture production to consumption: Freshness, safety, traceability and authentication, the four pillars of quality. Aquaculture, 518,734857.
- Gupta, M. K., Torrico, D. D., Ong, L., Gras, S. L., Dunshea, F. R., & Cottrell, J. J. (2022). Plant and dairy-based vogurts: a comparison of consumer sensory acceptability linked to textural analysis. Foods, 11 (3), 463.
- Hubbard, M. R. (2012). Statistical Quality Control for the Food Industry, Springer Science & Business Media, US.
- Jurica, K., Brčić Karačonji, I., Lasić, D., Bursać Kovačević, D., & Putnik, P. (2021). Unauthorized food manipulation as a criminal offense: Food authenticity, legal frameworks, analytical tools and cases. Foods, 10 (11), 2570.
- Mihafu, F. D., Issa, J. Y., & Kamiyango, M. W. (2020). Implication of sensory evaluation and quality assessment in food product development: A review. Current Research in Nutrition and Food Science Journal, 8 (3), 690-702.
- MINITAB, (2024). https://support.minitab.com/en-us/minitab/21/ help-and-how-to/quality-and-process-improvement/measurement-system-analysis/supporting-topics/attribute-agreement-analysis/what-is-an-attribute-agreement-analysis-also-called-attribute-gage-r-r-study/ assessed March of 16th.
- Mohammadi-Moghaddam, T. & Firoozzare, A. (2021). Investigating the effect of sensory properties of black plum peel marmalade on consumers acceptance by Discriminant Analysis, Food Chemistry: X, 11, 100126.
- MoreSteam, Attribute Agreement Analyss, https://www.moresteam.com/help/engineroom/attribute-agreement-analysis, assessed March of 7th, 2024.
- Nute, G. R. (2010). Sensory Evaluation of Meat Products. In Handbook of Meat Processing, Eds. F. Toldrá. 457-468, Blackwell Publishing.

- International Organization for Standardization. (2017). General requirements for the competence of testing and calibration laboratories (ISO/IEC 17025:2017). https:// www.iso.org/standard/.
- Pavli, F. G., Argyri, A. A., Chorianopoulos, N. G., Nychas, G. J. E., & Tassou, C. C. (2020). Effect of Lactobacillus plantarum L125 strain with probiotic potential on physicochemical, microbiological and sensorial characteristics of dry-fermented sausages. LWT, 118, 108810.
- Quintão, C., Andrade, P., & Almeida, F. (2020). How to Improve the Validity and Reliability of a Case Study Approach? Journal of Interdisciplinary Studies in Education, 9 (2), 264-275.
- Rose, J., & Johnson, C. W. (2020). Contextualizing reliability and validity in qualitative research: toward more rigorous and trustworthy qualitative social science in leisure research. Journal of Leisure Research, 51 (4) 432-451.
- Sipos, L., Nyitrai, Á., Hitka, G., Friedrich, L. F., & Kókai, Z. (2021). Sensory Panel Performance Evaluation—Comprehensive Review of Practical Approaches. Applied Sciences, 11 (24), 11977.
- Stone, H., Bleibaum, R. N., & Thomas, H. A. (2020). Sensory evaluation practices. In organizing and operating sensory sciece capabilites, 23, Publisher Elsevier Science,
- Sürücü, L., & Maslakci, A. (2020). Validity and reliability in quantitative research. Business & Management Studies: An International Journal, 8 (3), 2694–2726.
- Vivek, K., Subbarao, K. V., Routray, W., Kamini, N. R., & Dash, K. K. (2020). Application of fuzzy logic in sensory evaluation of food products: A comprehensive study. Food and Bioprocess Technology, 13, 1–29.
- Xiong, Y., Li, W., & Liu, T. (2020). Risk early warning of food quality safety in meat processing industry. International Journal of Environmental Research and Public Health, 17 (18), 6579.
- Zhang, Z., Sun, Y., Sang, S., Jia, L., & Ou, C. (2022). Emerging approach for fish freshness evaluation: principle, application and challenges. Foods, 11 (13),1897.

Authors ORCID info (ID)



Zoran Petrović https://orcid.org/0000-0003-2016-5681

Jelena Ćirić https://orcid.org/0000-0002-8118-7676

Jelena Babić Milijašević https://orcid.org/0009-0003-8923-7046

Milan Milijašević https://orcid.org/0000-0003-4269-236X

Mirjana Lukić https://orcid.org/0009-0001-7749-864X

Jelena Jovanović https://orcid.org/0000-0003-0301-729X

Aleksandra Nikolić https://orcid.org/0009-0009-0012-3418