Chapter

Sensory Evaluation of EVOO: Do Different Test Locations Have a Relevant Impact on Data Quality?

Annette Bongartz, Martin Popp and Richard Retsch

Abstract

In natural sciences, in general, the most important challenge is to ensure the reliability and validity of collected data and results. The identification of relevant influencing factors and the definition of adequate methodological approaches helps to minimize "noise." Within the sensory evaluation of extra virgin olive oil, many potentially influencing factors are known. Tasting procedures, therefore, are standardized. However, not all criteria have the same impact on data quality. The study at hand focuses on the possible influence of different test locations, comparing the situation in sensory labs (*in situ*) with so-called home testing stations (remote) for two well-experienced olive oil panels. Panel performance of both panels meets all regulatory requirements. Looking at the results from the overall statistical data analysis, slight differences between results coming from the two panels can be seen (nevertheless, not exceeding the requirements), but almost no differences are found between results coming out of different test situations. Knowing that the influence of testing through different panels is small but nevertheless bigger than a potential impact of testing in different test locations (sensory lab versus home testing stations), shows us the great potential for future use of remote test designs likewise to lab designs to obtain valid data.

Keywords: sensory evaluation, test location, remote testing, extra virgin olive oil (EVOO), data validity

1. Introduction

In natural sciences, the standardization of operation procedures, aiming at methodologies and the reliability as well as the validity of resulting data, is most important.

Thinking of test settings in sensory science, especially to control and monitor panelist and panel performance during the panel work in defined test settings is necessary. In the context of sensory evaluation of olive oil specifically, several official requirements do exist—on the one hand the EEC regulation 2568/91 [1], as amended, as well as several underlying documents and guidelines from the International Olive Council (IOC) [2–4] and moreover the general EN ISO/IEC 17025 regulations for any kind of testing laboratories [5].

Data quality must be the overriding objective in natural sciences and therefore is indispensable. Assuring a high data quality during data collection and assessment requires a clear focus on "data reliability" (high precision → same/similar results) and "data validity" (high accuracy → correct results). Well known is that one can gain high precision in measuring something, but at the same time can miss the target—meaning that results are precise, but not correct. So, overall high data quality can only be achieved, if data are on the one hand precise (reliable) and at the same time as well accurate (valid).

- Reliability: consistency and precision of measurement (in sensory analysis—repeatability of results—over time, of single panelists, of whole panels, indifferent test locations, ...)

- Validity: accuracy and trueness of measurement (in sensory analysis—homogeneity of results—between panels, in different test locations, ...)

Factors that might have a negative impact on data quality in sensory science are manifold. They can be related to the execution of the general procedure (test methodology), to the handling of test samples (blinding, distribution, temperature), to training and monitoring aspects of panelists and panels, to statistical analysis and data management as well as to the test infrastructure.

All above-named regulations and guidelines have in common to standardize and control procedures and finally to minimize "noise" in resulting data. This is helpful and valuable, but nevertheless, not all possible and focused so-called "influencing factors" cause a similar or even a relevant impact on data quality—some of them, presupposing a specific framework of instructions and settings, even have none.

2. Background and objective

The study at hand focuses on the aspect of different test locations for objective sensory analysis and their possible impact on the quality of resulting data.

Normally tests in sensory analysis take place as central location tests (CLT) in standardized testing rooms, such as sensory laboratories. Most of the regulations and guidelines propose these "*in situ*" (on-site) approaches. Panelists in these cases come to the lab, use the provided infrastructure (test cabins, hard and software for data collection, maybe additional technical equipment, such as heaters) for training and testing and at the same time, all other surrounding conditions (e.g., light, climate (humidity, temperature), else) are defined and controlled automatically. This is normally the rule and easily to be organized when using so-called "internal panels," consisting of panelists that work nearby, in the same institution, company, etc. Panelists can be scheduled quite easily and are able to take part in either regularly planned or as well spontaneously organized synchronic test situations.

But how about panelists in so-called "external panels," who must travel to be able to participate in regular on-site and synchronic trainings and test situations? Such dates can be organized and scheduled only medium or long-term presupposed. Rather flexible and spontaneous testing under such conditions is almost not possible. This consideration shows us—independently from pandemic situations—the need for additional appropriate test settings, that on the one hand can secure high standards of data quality/validity in sensory analysis and are on the other hand flexible as well

as time and cost-efficient. Doing sensory trainings and tests with panelists "remote" (off-site), meaning that panelists work at home respectively at defined and standardized "home-testing-stations," cause less costs for traveling (time per panelist, transportation). Data can be collected and exchanged online (web-based) either synchronously or asynchronously and as well panel meetings can take place remotely. Moreover, the latest data even show a better availability of panelists, due to less necessary time effort for traveling and thereby as well a higher motivation for contribution in panel work, compared to more elaborate (*in situ*) settings [6].

The aim of the study at hand is to prove the overall performance of participating sensory panels and the quality of the collected data. The focus lies on the evaluation of the suitability of "*in situ*" test situations (sensory laboratory) versus "remote" test situations (home test stations) for the sensory evaluation of olive oil. Additionally, the criteria of homogeneity between the considered panels as well as homogeneity, consistency, and repeatability of each single panel are particularly interesting in this context.

3. Materials and methods

3.1 Panels and panelists

The study at hand compares results from two sensory olive oil panels, namely the German Olive Oil Panel/DOP and the Swiss Olive Oil Panel/SOP.

Both panels are objective expert panels whose members have many years of experience in the sensory evaluation of olive oil. Specific infrastructure makes it possible to either carry out sensory tests synchronously and "*in situ*" in sensory laboratory (or comparable) situations, equipped with test booth, heaters, etc. On the other hand, both panels have the possibility to test asynchronously at so-called home testing stations. All panelists from both panels are regularly trained and undergo permanent monitoring of their panelist and panel performance. Additionally, both panels take part in regular inter-laboratory comparisons (proficiency tests). This is—in addition to other requirements—an important basis for their consistent, reliable, and successful work. Engaged in research projects as well as services, both panels contribute to the continuous development of quality on the olive oil market.

3.1.1 Swiss olive oil panel (SOP)

The SOP consists overall of 38 panelists and was founded in 2002. Since 2006 the SOP is accredited in accordance with EN ISO/IEC 17025 [5] and has been recognized by the International Olive Council (IOC) between 2009 and 2021. In the study at hand, the same nine panelists contributed "*in situ*" and synchronously to a test in a sensory laboratory situation as well as "remote" and asynchronously at their home testing stations.

3.1.2 German olive oil panel (DOP)

The DOP consists overall of 25 panelists and was founded in 1999. Since 2012 the DOP is accredited in accordance with EN ISO/IEC 17025 [5] and has been recognized by the International Olive Council (IOC) between 2012 and 2021. In the study at hand, the same 11 panelists contributed "*in situ*" and synchronously to a test in a sensory laboratory situation and "remote" and asynchronously at their home testing stations.

3.2 Test situations

There are at least two options to conduct sensory tests—on the one hand "*in situ*" the more common version of central location tests (CLT) in standardized testing rooms, for example, sensory laboratories, where panelists work synchronously and on the other hand "remote" the—at least in sensory analysis—less common home tests (HT) at so-called home testing stations, meaning in panelists homes, where panelists work in personal, but nevertheless standardized workspaces. Thinking about the advantages and disadvantages of the named test situations, we find pros and cons for both sides.

*3.2.1 Sensory laboratories (*in situ*)*

The sensory laboratory of SOP is located on the ZHAW campus in Wädenswil. It consists of 12 separate test booths. Each booth is equipped with a computer (incl. data acquisition software FIZZ by Biosystemes) and a heating device by Ettore Pasquali (mod. 145). Each heating device is recorded in a device list of the QMS and is checked regularly. Detailed information concerning the procedure of testing in the sensory laboratory (and concerning additional equipment, such as test glasses, spittoons, and palate-cleansing agents) is described in the QMS (internal document: Standard Operation Procedure: LMT-SEN-A5-302_translated EN → Sensory Evaluation at ZHAW (Sensory Laboratory)).

The sensory laboratory of DOP is in Nuremberg, Maxfeldstrasse 50. It consists of 12 separate mobile test booths. Each booth is equipped with a computer (incl. data acquisition software SENSORY by IMEDIA) and a heating device by Ettore Pasquali (mod. 145). Each heating device is recorded in a device list of the QMS and is checked regularly. Detailed information can be found in the QMS (internal document: Standard Operation Procedure SOP 07-02-02).

On the "Pro" side (**Table 1**) there is of course a high level of standardization. Samples are prepared in an absolute neutral way by the panel leader (or a technician), the testing takes place synchronously (at the same time) and electronic equipment is used for data collection and analysis.

On the "Contra" side (**Table 1**) we see low flexibility in terms of scheduling tests, especially because panelists must be available at the same time. If panelists are not collaborators and work on-site, they must travel. This is time- (and cost-) consuming.

Pros	Cons
ISO 17025 accreditation leads to a high level of standardization (controlled test situation, panelist/panel performance, etc.)	Synchronous testing causes low flexibility in terms of timing/scheduling tests → all testers have to be available at the same time
Sample preparation (pouring 15 ml) in coded glasses is done by the panel leader	Necessity for testers to get on-site requires traveling-time and is rather time-consuming for them
Synchronous testing (8–12 panelists)	Other potential influencing factors (malfunction interruption, noise, etc.) …
Use of electronic equipment to collect, compile and analyze data	

Table 1.
Pros/cons of a test situation in a sensory laboratory.

Moreover, even in a sensory lab, there is potential for additional interruptions (noise, malfunction, or else) with negative effects.

3.2.2 Home testing stations (remote)

Within the SOP, each panelist has set up a personal testing station in a room in his or her home. Each testing station is equipped with a heating device (Rosenstein & Söhne), a thermometer (Testo Mini penetration thermometer), and blue glasses including cover glasses. Each heating device and the thermometer is recorded in a device list of the QMS and is checked regularly. Detailed information concerning the procedure of testing in a home testing station is described in the QMS (internal document: Standard Operation Procedure: LMT-SEN-A5-303_translated EN → Sensory Evaluation at Home Testing Stations).

Within the DOP each panelist has set up a personal testing station in a room in his or her home, which meets the test conditions of the IOC in terms of light, temperature, noise, and odors (COI/T.20/DOC. No 6/Rev. 1). The mobile booth is made up of folding elements in such a way that the panelist is isolated from negative ambient conditions. Each test station is equipped with a heating device (Ettore Pasquali, mod. 145), a thermometer (Testo Mini penetration thermometer), and blue glasses, including a cover glass. Each heating device and the thermometer is recorded in a device list of the QMS and is checked regularly. Detailed information can be found in the QMS (internal document: Standard Operation Procedure SOP 07-02-02).

On the "Pro" side (**Table 2**) there is—like for a lab situation—as well a high level of standardization because as well home testing stations provide ISO accreditation. Asynchronous testing increases flexibility for scheduling tests, meaning that testers do not have to be available at the same time, only in a defined period. Less traveling time and costs are positive and of course—like in the lab situation—as well at home testing stations, electronic equipment for data collection and analysis is used.

On the "Contra" side (**Table 2**), we see that a sample dispatch is needed, which must be well organized regarding packaging and preparation of samples. For temperature protection during transport, Styrofoam boxes are used, and the oil is coded

Pros	**Cons**
ISO 17025 accreditation leads to a high level of standardization → controlled test situation, panelist/panel performance, etc.	Sample dispatch is needed → 30 ml (per olive oil) in dark glass bottles and use of styroporous boxes which secure sample temperature
Asynchronous testing enables high flexibility in terms of timing/scheduling tests → testers do not have to be available at the same, they have to respect deadlines, but otherwise can plan rather independently	Sample preparation (pouring 15 ml) in coded glass is done by each panelist → half of the whole 30 ml quantum
	Asynchronous testing (8–12 panelists) → no relevance because of training status
Less requirement of traveling-time because of remote testing (at home)	Other potential influencing factors (malfunction interruption, noise, etc.) ...
Use of electronic equipment to collect, compile and analyze data	

Table 2.
Pros/cons of a test situation at a home testing station.

and bottled to dark 30 ml bottles. Sample preparation, directly before testing, is done by the panelist. He or she must pour out exactly half of the bottle (15 from 30 ml). Asynchronous testing is of course different compared to asynchronous lab testing situation, but since panelists are well trained on using the methodology and it is always (as well in the lab) a single panelist evaluation before compiling data, this does not lead to any problems. Finally, and again like in the lab situation, there is of course potential for additional interruptions (noise, malfunction, or else).

3.3 Test methodology (EU 2568/91, as amended)

The applied sensory methodology is based on the official panel test according to the regulation EEC regulation 2568/91 [1] and related IOC documents.

3.4 Validation concept

To be able to record high data quality in the context of sensory evaluation of olive oil, the reliability and validity of raw data must be ensured. Therefore, a study concept, based on the recommendations of the IOC (COI/T.28/Doc. No.1/Rev. 5 2019), was considered. Among other criteria, analyzing the panelist and panel performance, the concept focuses especially on the aspect of the test situation (sensory laboratory versus home testing stations) (**Table 3**).

3.5 Test samples

Both participating panels did evaluate the same selection of 10 olive oils (same lot number) "*in situ*" (in the sensory laboratory) and "remote" (at home testing stations), see **Table 4**. The order of testing was the same in all test situations but in each test situation individual three-digit codes were used to avoid influencing the testers.

	Test situation	Panel	Panel performance
Focus	Sensory laboratory vs. Home testing stations	DOP vs. SOP	• Homogeneity • Consistency • Repeatability
Attributes	• Fruitiness • Bitterness • Pungency • Defect	• Fruitiness • Bitterness • Pungency • Defect	• Fruitiness • Bitterness • Pungency • *Defect
Methodological approach	• Graphical visualization • Mixed-model ANOVA • Analyzed for each panel separately	• Graphical visualization • Mixed-model ANOVA	• Trueness (<2.0) 1. *z-Score → homogeneity* 2. *Deviation (DNp)* from other panels per session → homogeneity • Precision (<2.0) 1. *Normalized error (En) — panel mean → repeatability* 2. *Precision number (PNp) — panel mean → consistency* • Analyzed for each panel separately

*not analyzed in present study

Table 3.
Validation criteria.

No.	Sample code*	Product information
P1	104	100% Italian olive oil; different varieties, extra virgin
P2	507	European Blend; different varieties, extra virgin
P3	620	100% Italian olive oil, 100% Nocellara, extra virgin
P4	733	100% Italian olive oil; different varieties, extra virgin
P5	249	100% French olive oil; different varieties, extra virgin
P6	362	Olive Oil from IOC Org 2—2020, extra virgin
P7	878	Olive Oil from IOC Org 1—2020, defective
P8	168	Olive Oil from IOC Org 1—2020, extra virgin
P9	055	Olive Oil from IOC Org 2—2020, defective
P10	652	Olive Oil from IOC Org 2—2020, defective

**For each test situation, individual three-digit codes were used to avoid influencing the testers.*

Table 4.
Test samples.

3.6 Data collection

Both participating panels did evaluate the test samples in the same period, but independently. The evaluation criteria on the used profile sheets from both panels (electronically/paper) were identical, corresponding with the EEC regulation 2568/91 [1].

3.6.1 Swiss olive oil panel/SOP

To collect data in the sensory laboratory of ZHAW, panelists from SOP are provided with PC's in the test booth, equipped with the sensory software "Fizz" (Biosystemes Fizz for Windows 2.46 A), which allows direct electronic recording of individual panelist data on an electronic profile sheet.

For the collection of data at home testing stations, panelists from SOP use a profile sheet (paper) and transfer individual panelist data, online via the internet, to the panel leader, using the software "LimeSurvey."

3.6.2 German olive oil panel/DOP

To collect data in the sensory laboratory situation, panelists from DOP are provided with PC's in the test booth, equipped with the software "SENSORY" (IMEDIA), which allows direct electronic recording of individual panelist data on an electronic profile sheet.

For the collection of data at home testing stations, panelists from DOP use a profile sheet (paper) and transfer individual panelist data, online via the internet, to the panel leader, using the software "SENSORY" (IMEDIA).

3.7 Data evaluation (assessment)

All results of the different tasting sessions with all panelists and all panels were combined to a common data set, using the software program "Excel" (Microsoft

Office Excel 365). The following data evaluation was done with help of the Add-in Software "XLStat" (version 2020).

4. Results and discussion

In the study at hand altogether 10 olive oils (seven of them extra virgin and three defective ones) were independently tested by the Swiss Olive Oil Panel/SOP (nine panelists) and the German Olive Oil Panel/DOP (11 panelists) in different test situations—namely in sensory laboratories (*in situ*) and at so-called home testing stations (remote). Test results were analyzed regarding the validity of the data and thereby focusing on the following three aspects:

1. Agreement between test situation (*in situ* versus remote)

2. Agreement (homogeneity) between panels (SOP versus DOP)

3. Individual performance of both panels (SOP, DOP)

All data/panel results were valid according to IOC specifications (e.g., Cvr < 20%).

4.1 Agreement between different test situations (*in situ* versus remote)

First, and for both panels separately, the agreement between data collected in different test situations—sensory laboratory (*in situ*) and home testing stations (remote testing)—was checked. For this purpose, raw data were analyzed using mixed-model ANOVA. Results show that there are no statistically significant differences between

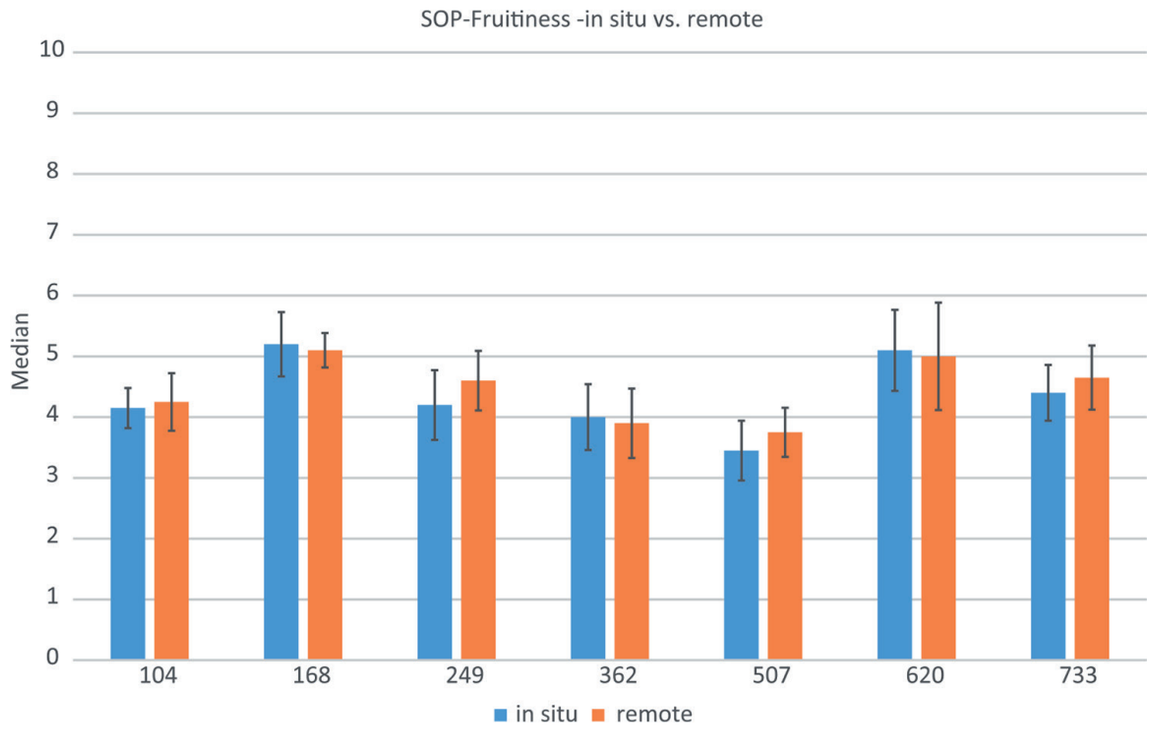

Figure 1.
Median of fruitiness (SOP)—in situ versus remote (n = 9).

the different test situations within each panel. This means that SOP and DOP can repeat their results from the laboratory situation at home testing stations.

4.1.1 Swiss olive oil panel (SOP)

Looking at the attribute fruitiness, **Figure 1** shows the comparison of the two medians per sample of all seven extra virgin olive oils. The maximum difference found was 0.4 for sample 249 (**Figure 2**). So, one can say, that—for the attribute fruitiness—there is no significant difference between results coming from *in situ* testing compared to home testing stations (remote). Results from mixed-model ANOVA are shown in **Table 5**.

For bitterness, you can easily see in **Figure 3** that there were found as well similar medians for all analyzed oils. The maximum difference was 0.3 for sample 104. Similar to fruitiness, as well as bitterness, there is no significant difference between results coming from the *in situ* testing compared to results from home testing stations (remote). Results from mixed-model ANOVA can be seen in **Table 6**.

For pungency, not surprisingly the comparison of medians in all seven oils in **Figure 4** shows only slight differences. The maximum difference found is 0.5 for sample 249. Again, there is no significant difference between results coming from the lab (*in situ*) compared to the home testing stations (remote). Results from mixed-model ANOVA can be seen in **Table 7**.

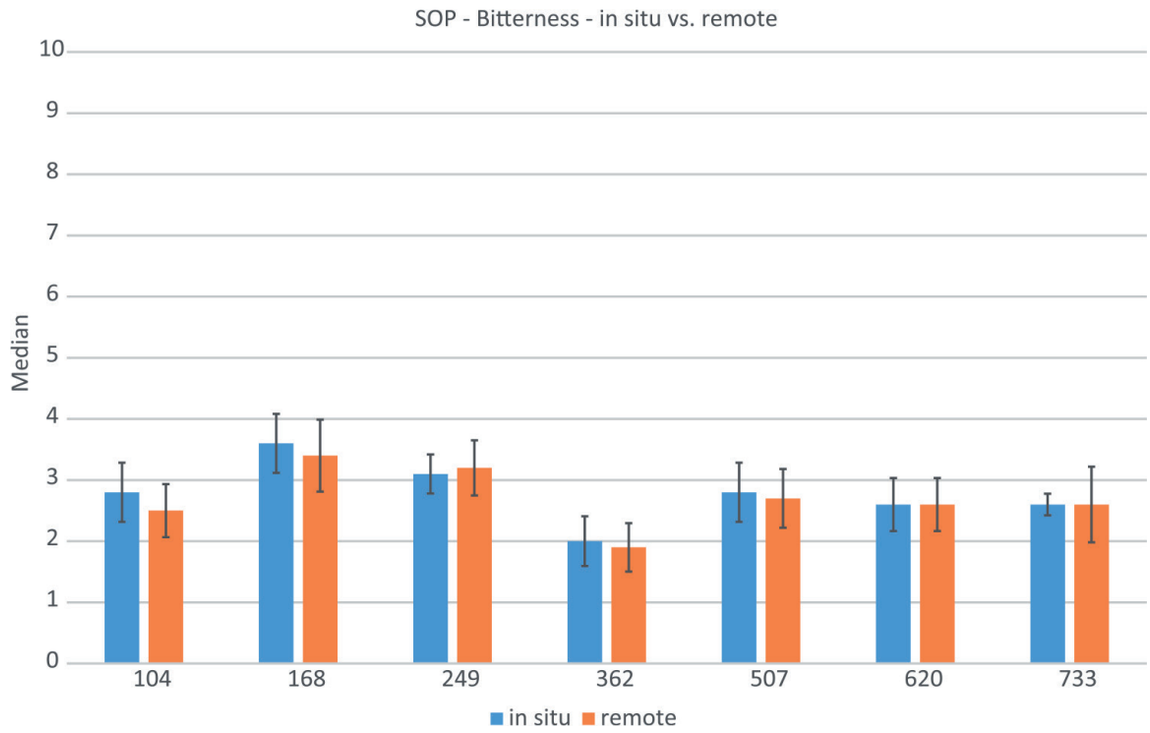

Figure 2.
Median of bitterness (SOP)—in situ versus remote (n = 9).

Contrast	Difference	Standardized difference	Critical value	P-value	Significant
In situ versus remote	0.142	0.991	2.120	0.336	No

Table 5.
Mixed model ANOVA for fruitiness (SOP)—in situ versus remote (n = 9).

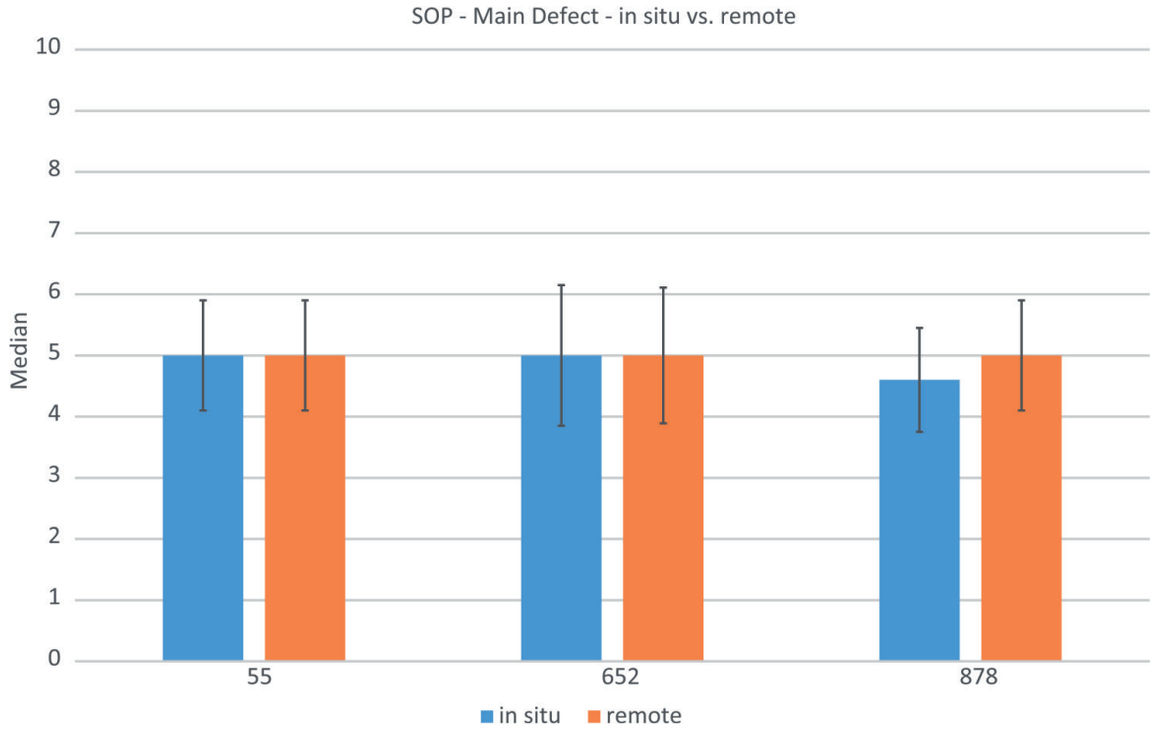

Figure 3.
Median of main defect (SOP)—in situ versus remote (n = 9).

Contrast	Difference	Standardized difference	Critical value	P-value	Significant
In situ versus remote	0.065	0.532	2.120	0.602	No

Table 6.
Mixed-model ANOVA for bitterness (SOP)—in situ versus remote (n = 9).

Finally, **Figure 5** focuses on defects and visualizes the comparison of medians of the main defect of the three defective olive oils. The maximum difference between medians was 0.4 for sample 878. So as well for defects, we can see no significant difference between results coming from the *in situ* testing compared to the home testing stations (remote). Results from mixed-model ANOVA are shown in **Table 8**.

4.1.2 German olive oil panel (DOP)

Looking at the attribute fruitiness, **Figure 6** visualizes the comparison of the medians of all seven extra virgin olive oils analyzed. Like for SOP, as well for DOP the maximum difference between medians for very small, in this case, 0.4 for sample 362. This proves that there is no significant difference between results coming from the *in situ* testing compared to the home testing stations (remote), as well for DOP. Results from mixed-model ANOVA are shown in **Table 9**.

For bitterness, **Figure 5** visualizes the comparison of the median of all extra virgin olive oils analyzed. The maximum difference between medians is 0.2 for samples 104 and 507. There is no significant difference between results coming from *in situ* testing compared to the home testing stations (remote). Results from mixed-model ANOVA are shown in **Table 10**.

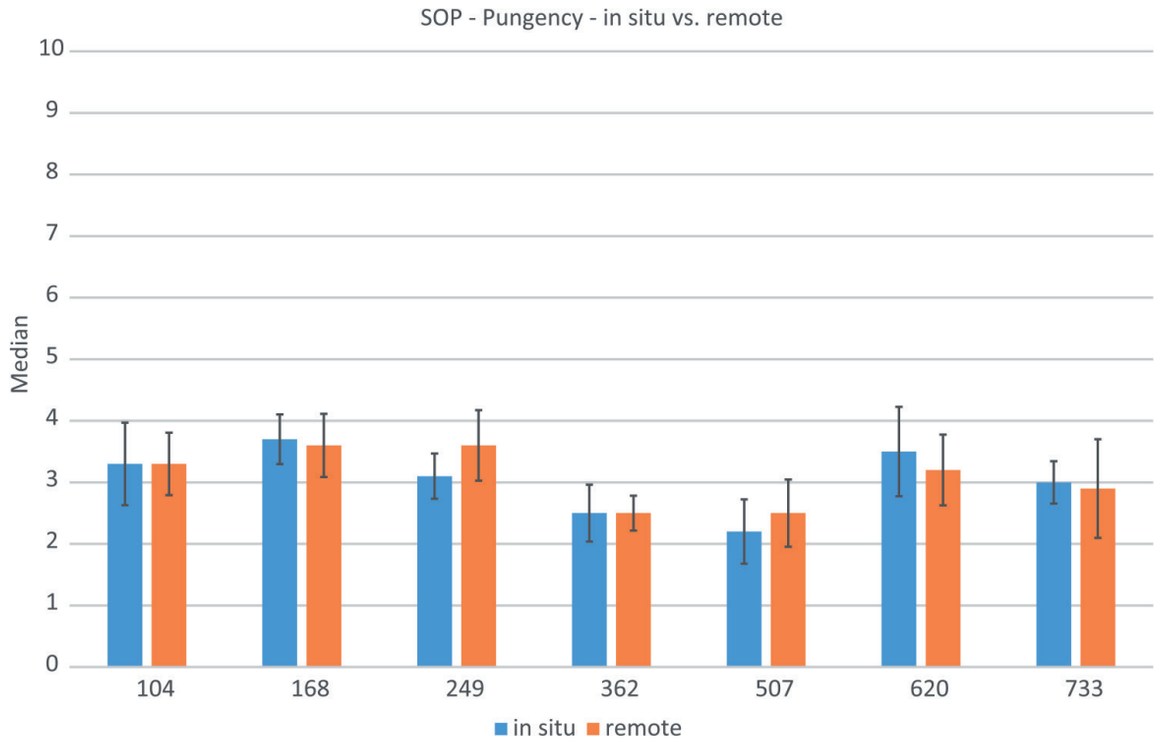

Figure 4.
Median of pungency (SOP)—in situ versus remote (n = 9).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.017	0.125	2.120	0.902	No

Table 7.
Mixed-model ANOVA for pungency (SOP)—in situ versus remote (n = 9).

For pungency, **Figure 7** visualizes the comparison of the median of all extra virgin olive oils analyzed. The maximum difference between medians is 0.3 for samples 104 and 362. There is no significant difference between results coming from the *in situ* testing compared to the home testing stations (remote). Results from mixed-model ANOVA are shown in **Table 11**.

Finally, **Figure 8** focuses on defects and visualizes the comparison of the median of the main defect of the 3 defective olive oils. The maximum difference between medians is 0.2 for samples 055 and 652. There is no significant difference between results coming from the *in situ* testing compared to the home testing stations (remote). Results from mixed-model ANOVA are shown in **Table 12**.

4.2 Agreement between different panels

Second, mixed-model ANOVA was used to analyze whether there exist differences between the data collected from the two IOC-recognized panels separately. It was found that there are statistically significant differences between the two panels. The largest difference for the mean value is 0.5 on a 10 cm scale. This means, that panels

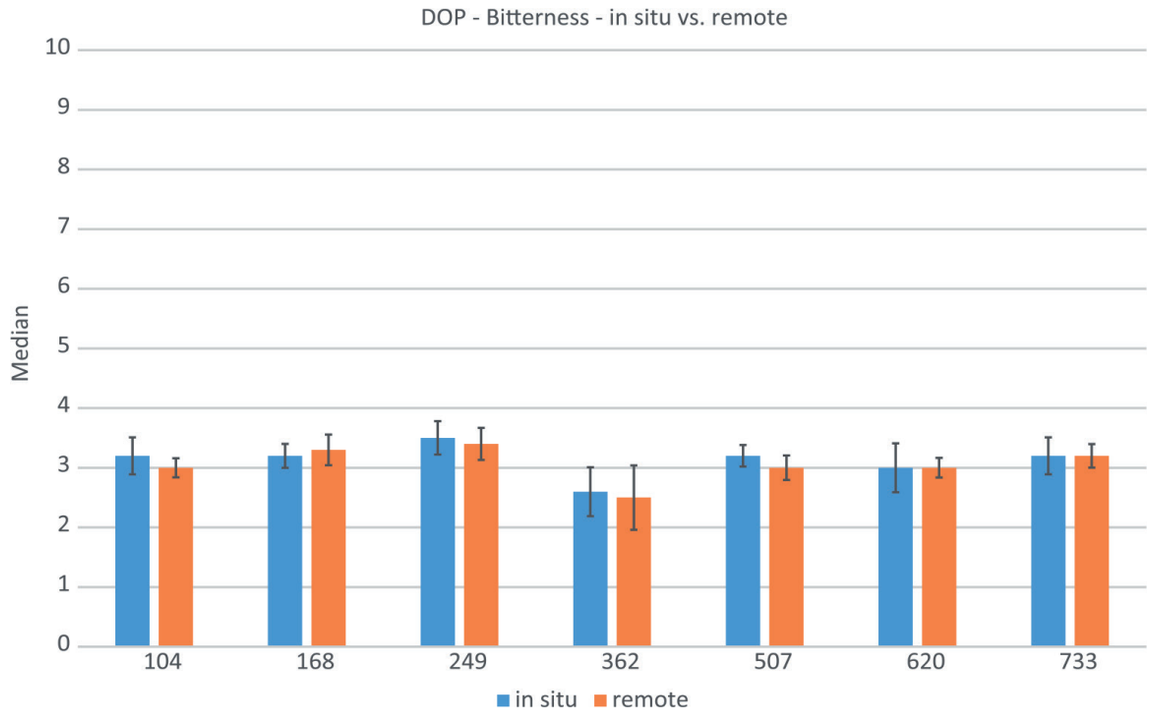

Figure 5.
Median of bitterness (DOP)—in situ versus remote (n = 11).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.082	0.289	2.086	0.775	No

Table 8.
Mixed-model ANOVA for main defects (SOP)—in situ versus remote (n = 9).

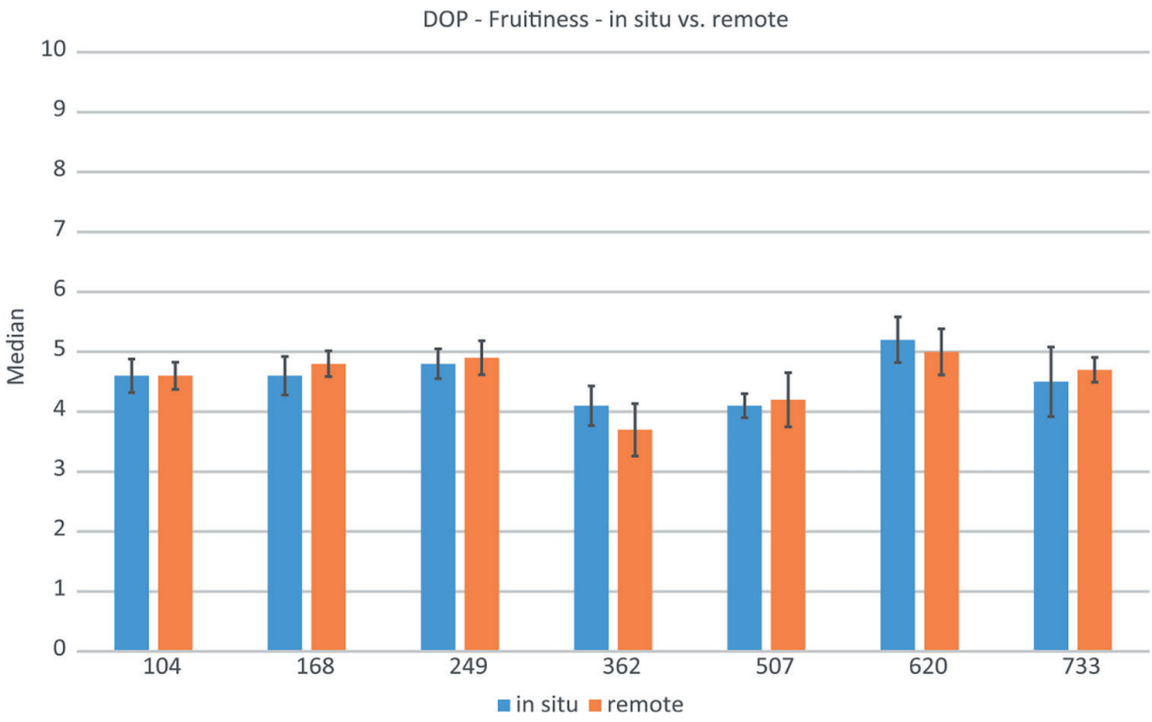

Figure 6.
Median of fruitiness (DOP)—in situ versus remote (n = 11).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.004	0.044	2.086	0.966	No

Table 9.
Mixed-model ANOVA for fruitiness (DOP)—in situ versus remote (n = 11).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.095	1.500	2.086	0.149	No

Table 10.
Mixed-model ANOVA for bitterness (DOP)—in situ versus remote (n = 11).

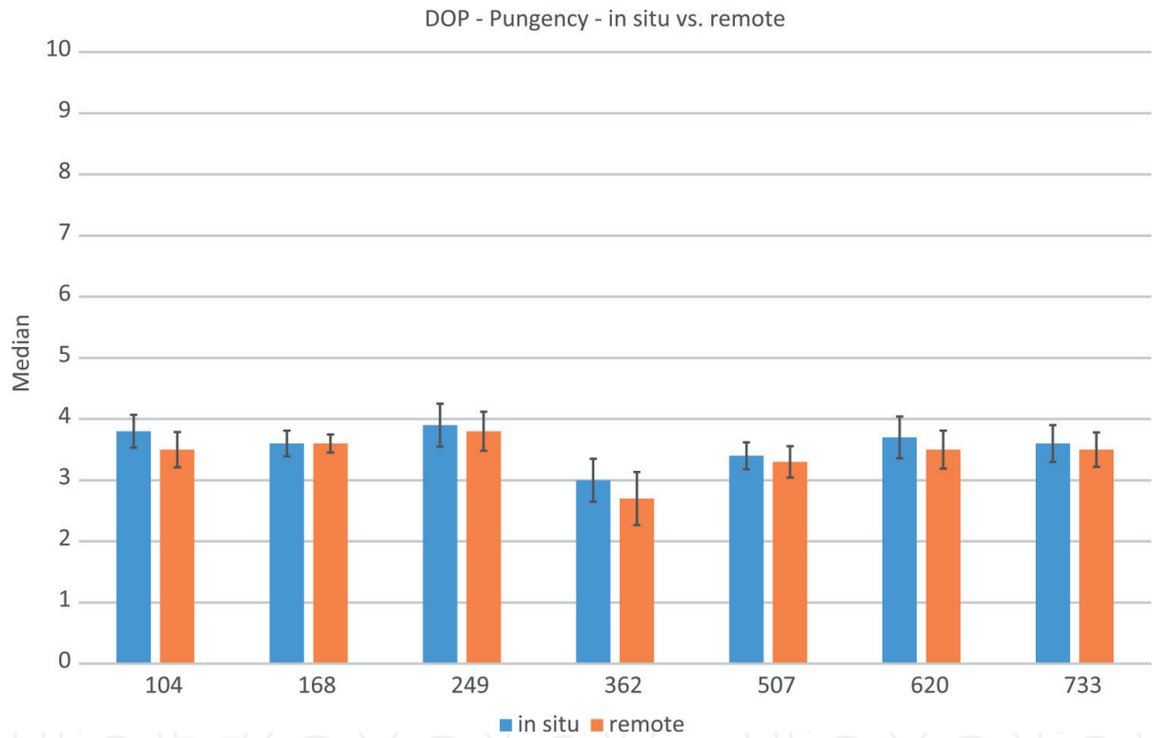

Figure 7.
Median of pungency (DOP)—in situ versus remote (n = 11).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.134	1.958	2.086	0.064	No

Table 11.
Mixed-model ANOVA for pungency (DOP)—in situ versus remote (n = 11).

show variance, but nevertheless, results are within the expected and accepted variation proposed by the IOC.

For the attribute fruitiness "*in situ*" as well as fruitiness "remote" **Figures** 9 and 10 show that the two panels do differ only slightly (0.2) and the results of the two panels show significant differences (**Table** 13).

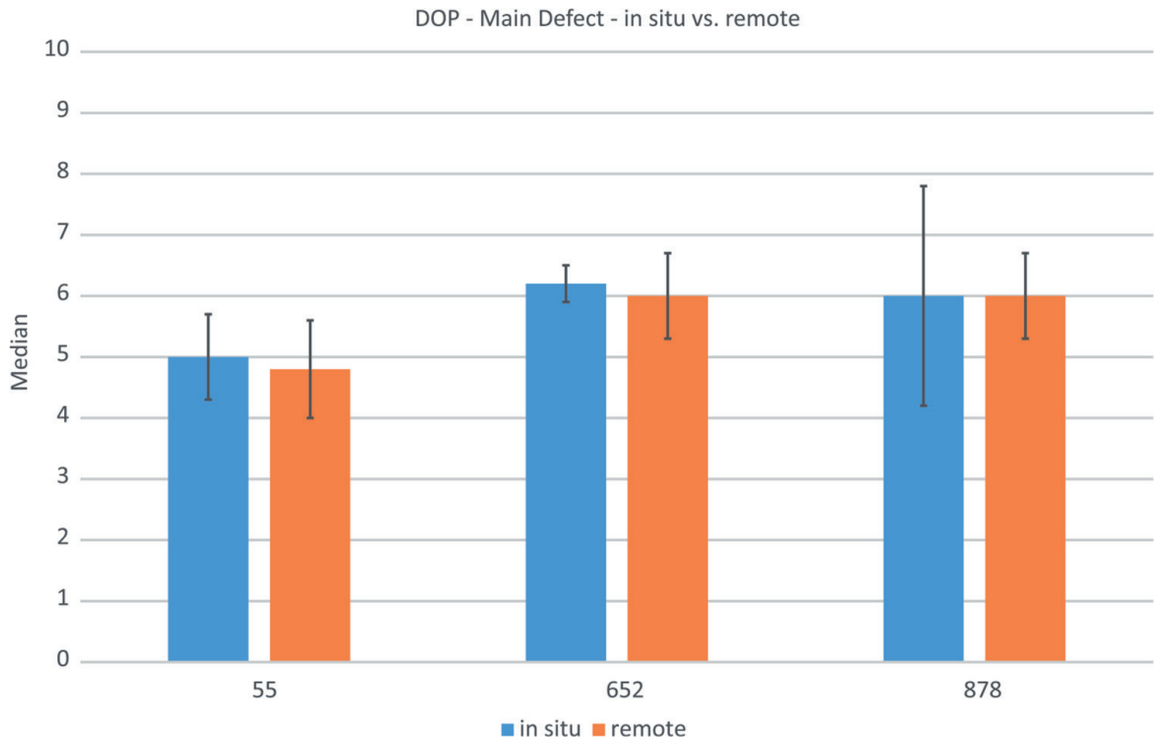

Figure 8.
Median of main defect (DOP)—in situ versus remote (n = 11).

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
In situ versus remote	0.078	0.142	2.120	0.889	No

Table 12.
Mixed-model ANOVA for main defects (DOP)—in situ versus remote (n = 11).

This is as well the case for the other positive attributes—bitterness and pungency. Regardless that the comparisons of panels show significant differences, it can be stated, that the maximum deviation in the mean value was in all attributes only 0.5.

A similar situation is found for the main defects "*in situ*" and "remote." It is shown in **Figures 11** and **12** that the two panels do differ slightly (1.5) and results from both panels show significant differences (**Table 14**).

Based on the shown data, it can be concluded, that even if the "difference between panels" (DOP/SOP) is significant, the variance in all cases is well below the IOC accepted differences between recognized panels.

4.3 Panel performance of single panels

Third, and based on the document COI/T.28/Doc. No.1/Rev. 52,019, the panel performance for both panels were analyzed according to the following selected criteria: *Z*-Score, Deviation Number as well as Normalized Error and Precision Number for both participating panels, SOP and DOP. Results show that both panels meet the IOC requirements in all criteria mentioned. The following data from SOP are shown exemplarily.

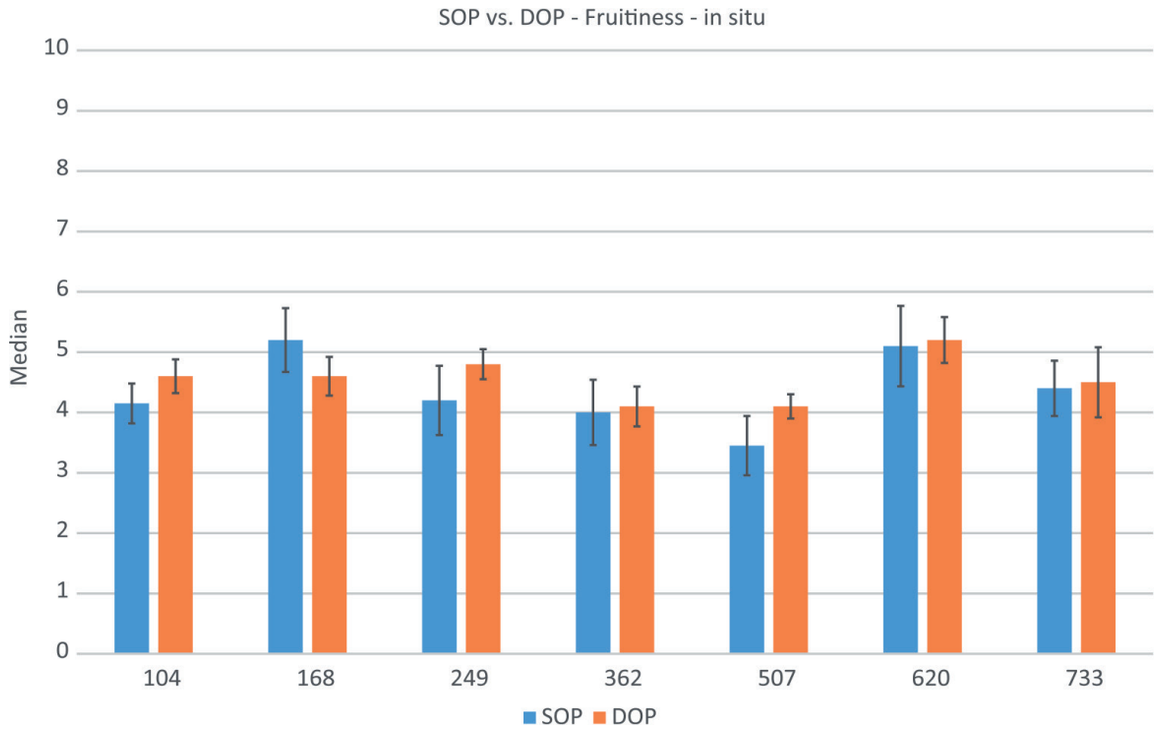

Figure 9.
Median of fruitiness (SOP versus DOP)—in situ.

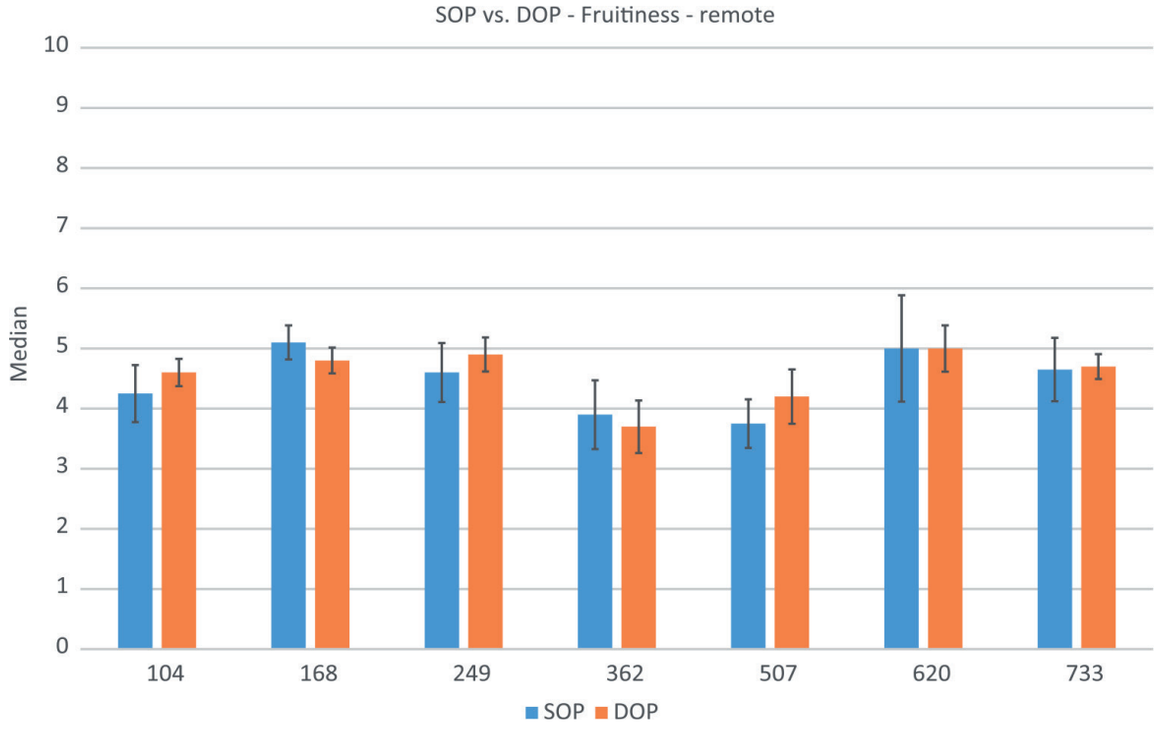

Figure 10.
Median of fruitiness (SOP versus DOP)—remote.

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
SOP versus DOP	0.181	2.702	1.978	0.008	Yes

Table 13.
Mixed-model ANOVA for fruitiness (SOP versus DOP).

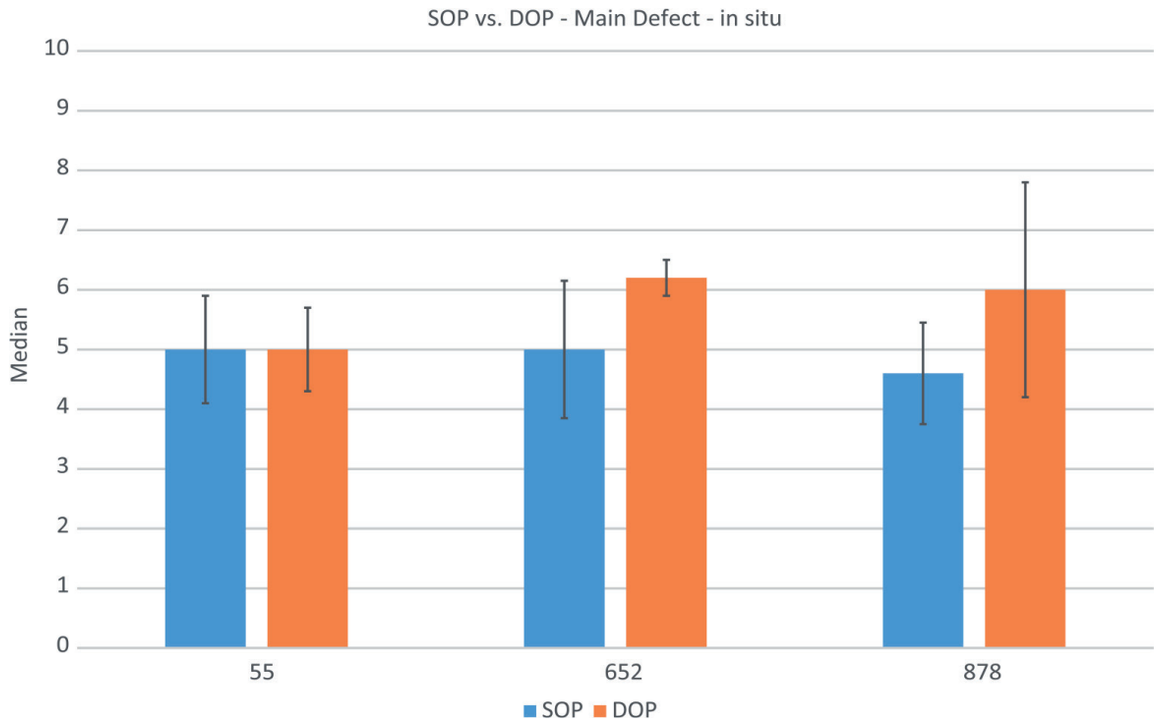

Figure 11.
Median of main defect (SOP versus DOP)—in situ.

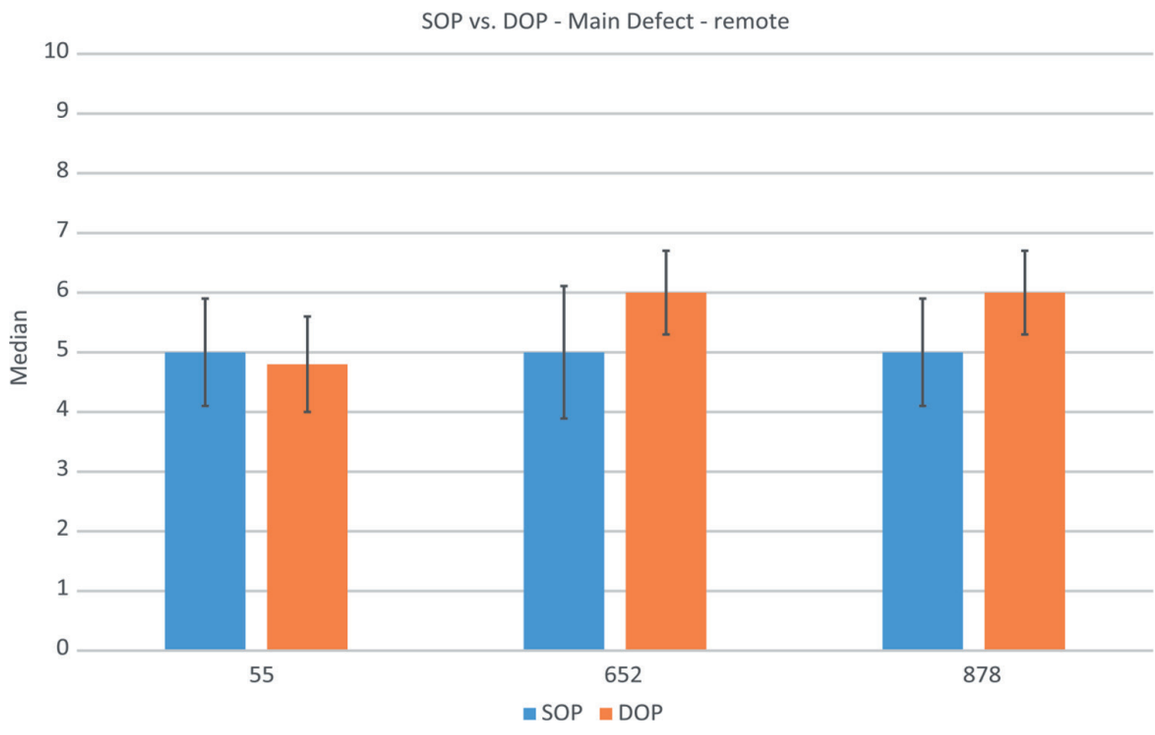

Figure 12.
Median of main defect (SOP versus DOP)—remote.

4.3.1 Trueness (homogeneity): z-score (for panels)

The calculation of the *z*-score for panels focuses on the difference between one-panel result (median) and a reference result (median) in relation to a defined SD of

Contrast	Difference	Standardized difference	Critical value	*P*-value	Significant
SOP versus DOP	1.449	3.853	2.002	<0.0001	Yes

Table 14.
Mixed-model ANOVA for main defect (SOP versus DOP).

0.7. The reference result (in this study) is defined as the mean over all four test results (DOP remote, DOP *in situ*, SOP remote, and SOP *in situ*).

Z-score = difference between one-panel result (median) and the reference result* (median) in relation to SD.

*Reference result = median of results from all four considered test situations (DOP remote, DOP *in situ*, SOP remote, and SOP *in situ*)

- SD = maximum standard deviation of the method = ±0.7
- Warning limit = ±2
- Action limit = ±3

→ Proof of trueness/homogeneous results (statistically acceptable)

We can see in **Figures 13** and **14** for the attribute fruitiness, that for *in situ* testing as well as for remote testing, the *z*-Score of SOP is in line with the requirements, that is—well below warning (±2) and action limit (±3). This is the case for bitterness and pungency as well, but not shown here. Based on these findings, it can be concluded that results from SOP are homogeneous (in both test situations) and statistically acceptable (= aspect of trueness).

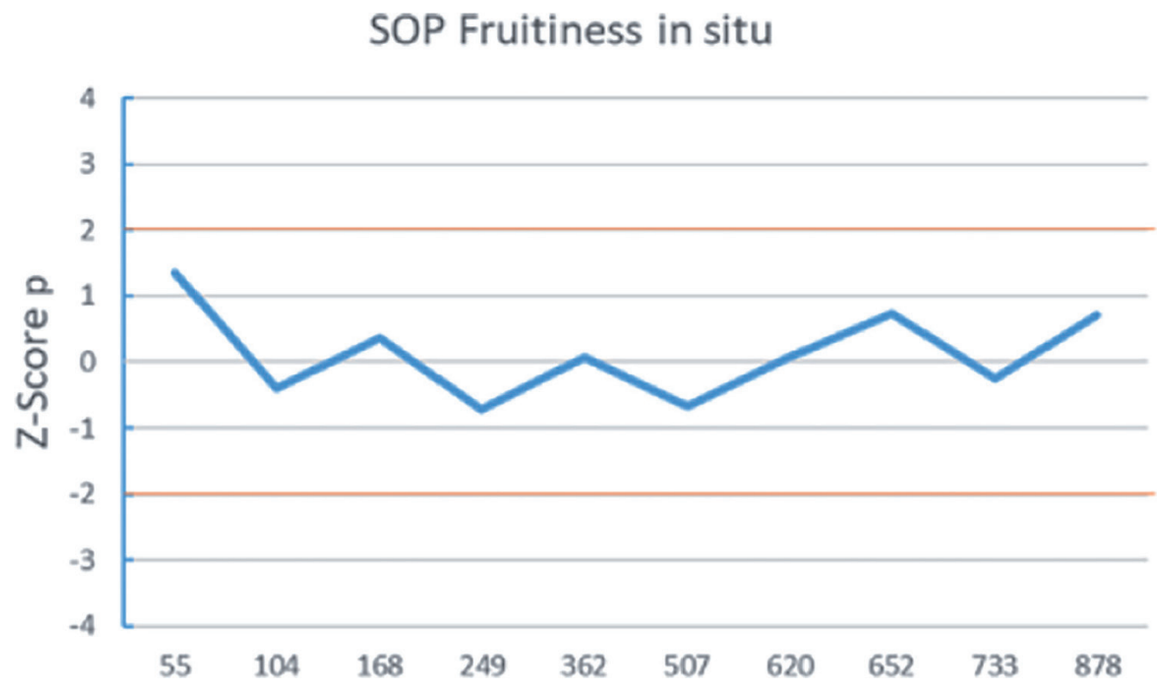

Figure 13.
z-Score (SOP) fruitiness—in situ (n = 9).

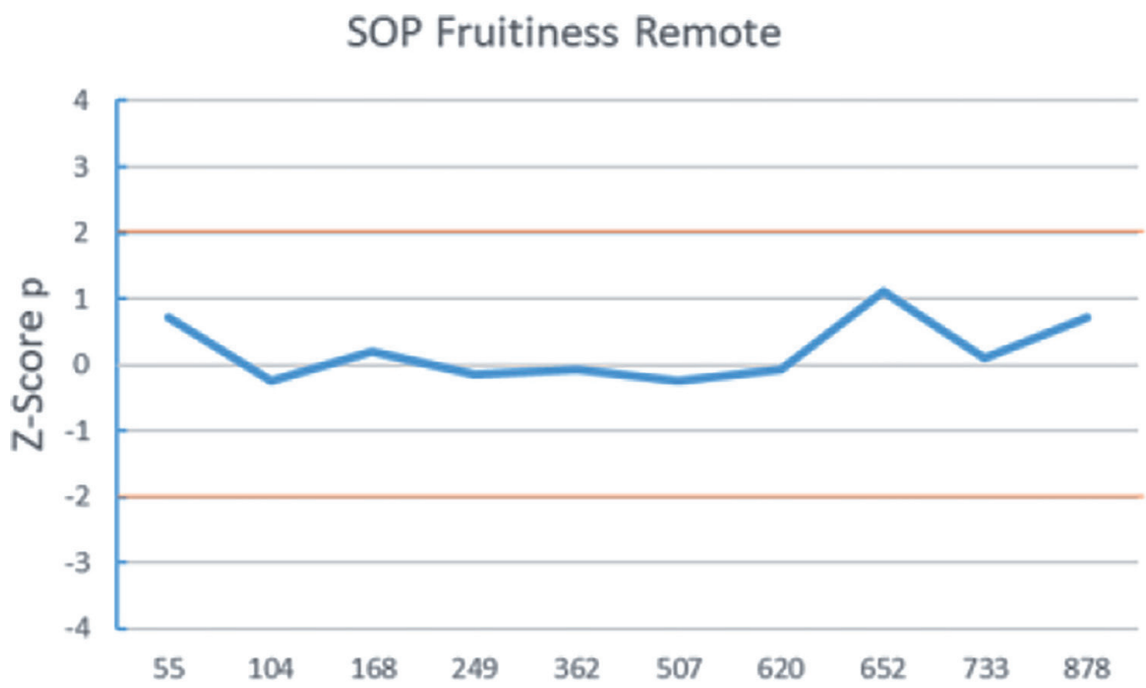

Figure 14.
z-Score (SOP) fruitiness—remote (n = 9).

4.3.2 Trueness (homogeneity): DN_p (deviation number for panels)

The calculation of the DN_p (deviation number for panels) focuses on the sum of differences (squared) between duplicate results (median) and the reference result (median) in relation to the number of reference samples (in our case 4). The reference result (in this study) is defined as the mean over all four test results.

DN_p = sum of differences (squared) between duplicate results (median) and the reference result* (median) in relation to the number of reference samples (4).

*Reference result = median of results from all considered test situations (DOP remote, DOP *in situ*, SOP remote, and SOP *in situ*)

- Duplicate = comparison between *in situ*/remote
- Number of samples building the reference mean = 4

→ Proof of trueness/homogeneous results (statistically acceptable)

Figure 15 shows for the attribute fruitiness, that the deviation number for SOP is in line with the requirements, that is: well below the warning limit of 1.0 and the action limit of 2.0). This is the case for bitterness and pungency as well, not shown here. Based on these findings, it is proven that results of SOP are homogeneous between different test situations and statistically acceptable (= aspect of trueness).

4.3.3 Precision (repeatability, consistency): E_n (normalized error)

The calculation of the E_n (normalized error) focuses on the difference between duplicate panel results (mean) in relation to a defined SD of 0.7.

E_n = difference between duplicate panel results (mean) in relation to SD.

- Duplicate = comparison between *in situ*/remote

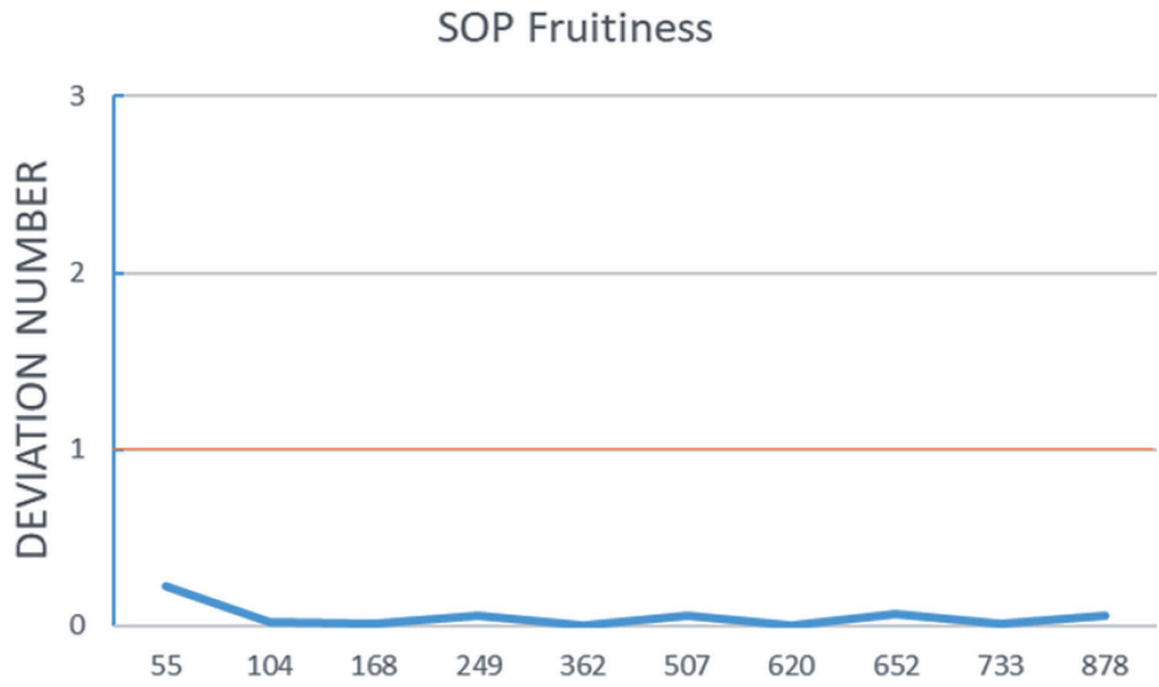

Figure 15.
DN_p (SOP) fruitiness (n = 9).

• SD = maximum standard deviation of the method (or maximum error) ± 0.7

→ Proof of precision/consistent results (statistically acceptable)

Coming to the aspect of "Precision" (repeatability, consistency), we can see in the figure for fruitiness (**Figure 16**) that results of the normalized error from SOP are in line with the requirements, that is—well below the action limit of 1.0. This is the case for bitterness and pungency as well, not shown here. Based on these findings, it can be concluded that results of SOP are repeatable between different test situations and statistically acceptable (precise, consistent).

4.3.4 Precision (repeatability, consistency): PNp (the precision number for panels)

The calculation of the PN_p (the precision number for panels) focuses on the sum of differences (squared) between duplicate panel results (mean) in relation to the number of duplicate samples (in this study 10).

PN_p = sum of differences (squared) between duplicate panel results (mean) in relation to the number of duplicate samples (10)

• Duplicate = comparison between *in situ*/remote

• Number of duplicate samples = 10

→ Proof of precision/consistent results (statistically acceptable)

We can see in the figure for fruitiness (**Figure 17**) that the results of the precision number for panels from SOP are in line with the requirements, that is—well below the action limit of 2.0. This is the case for bitterness and pungency as well, but not shown here. Based on these findings, it is proven that results of SOP are consistent and precise between different test situations and statistically acceptable.

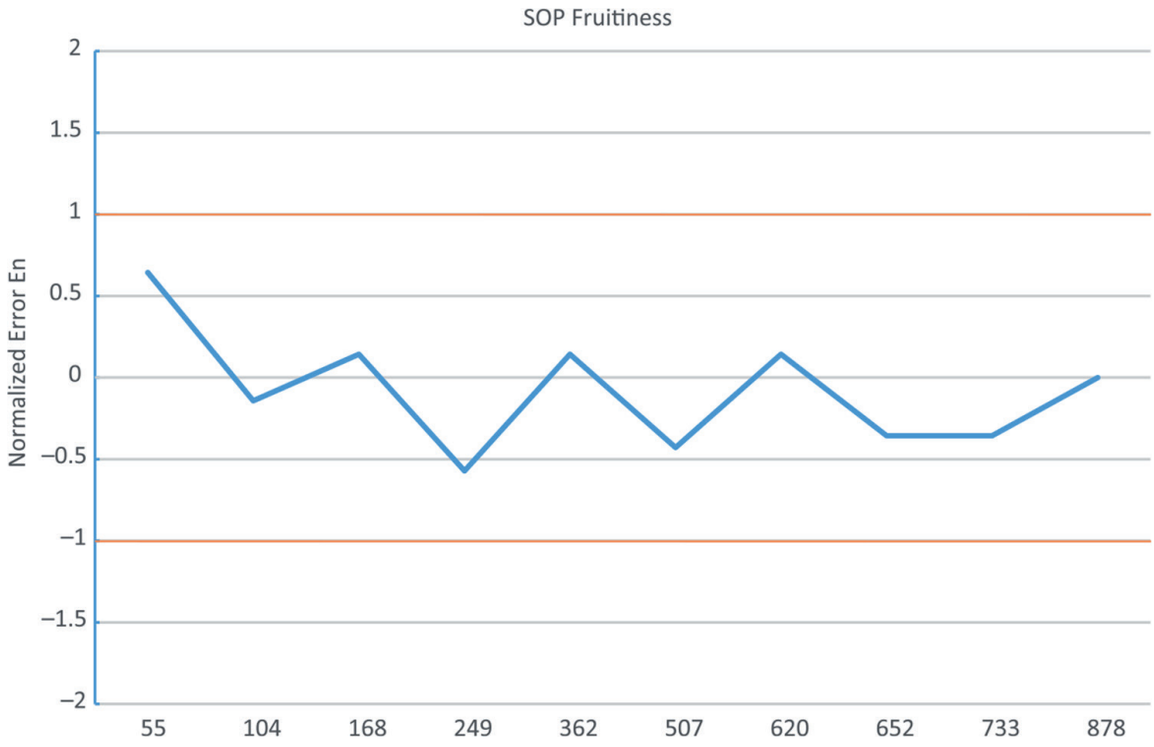

Figure 16.
E_n (SOP) fruitiness (n = 9).

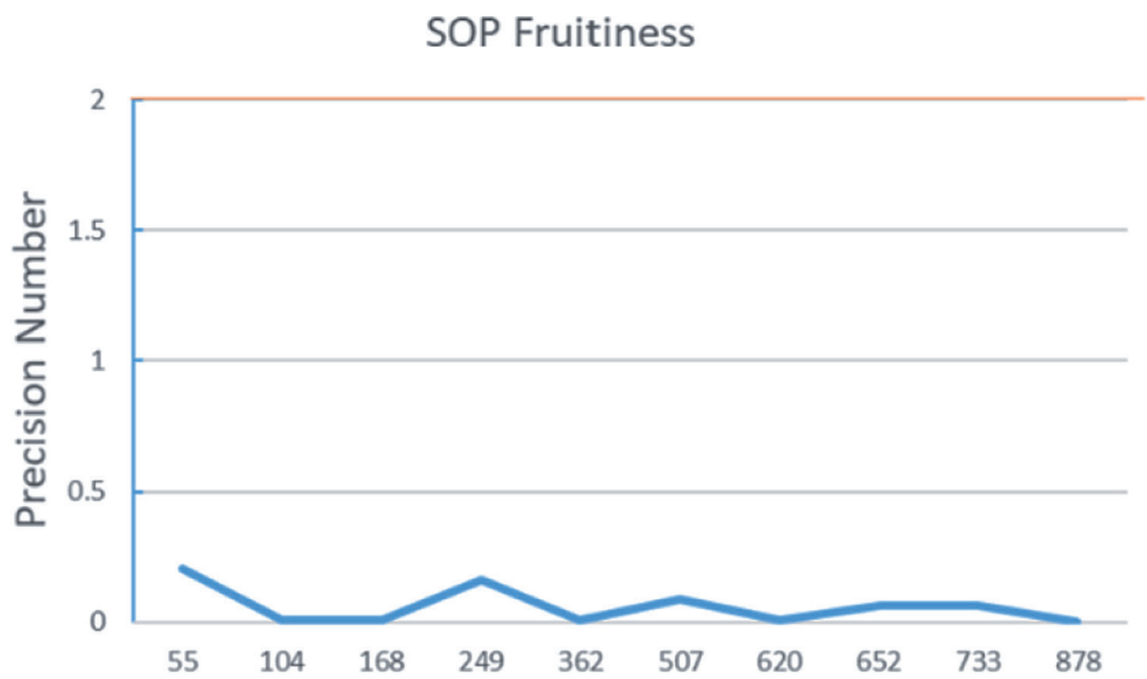

Figure 17.
PN_p (SOP) fruitiness (n = 9).

5. Conclusions and outlook

To prove data quality in terms of reliability and validity, 10 olive oils (seven of them extra virgin and three defective ones) were independently tested by the Swiss Olive Oil Panel/SOP (nine panelists) and the German Olive Oil Panel/DOP (11

panelists) in different test situations—namely in sensory laboratories (*in situ*) and at so-called home testing stations.

Analyzing the raw data, various aspects of panel performance were looked at—especially the different test situations (*in situ* versus remote), the variations between results from different panels as well as selected aspects concerning the individual panel performance per panel.

1. The individual panel performance per panel [3] shows that both panels meet the requirements from IOC.

2. Significant differences between the two considered olive oil panels were found, but the variation is within the accepted limits required by the IOC.

3. No significant differences between test situations (*in situ*/remote) were found → panels can repeat results between laboratory and home testing situation.

Overall, results show that the influence of testing through different panels (SOP versus DOP) is bigger than the impact of testing in different test situations (*in situ* versus remote). However, it is important to mention that to achieve such reliable data in both test situations, professional equipment and intense and adequate training of the panelists/panels is required. DOP and SOP fully meet these expectations. Data from both panels coming from both test situations are fully reliable and valid.

Besides the convincing findings from this study—many comparison tests over the last years took place proofing reliability and precision of tests taking place either *in situ* or remote. Remote testing—in a defined framework—therefore is proven to be a very valuable methodology and technique, independently from pandemic situations. In the meantime, the test procedure is well established in different contexts and accepted by official certification bodies (ISO 17025) as well as retailers, importers, and consumers.

In the future, it will be valuable to set up advanced follow-up studies with even more participating panels from different countries to regularly confirm findings and strengthen the trust in the data and conclusions of the study at hand.

Acknowledgements

Our thanks go especially to the panelists from the two involved panels—the German Olive Oil Panel DOP and the Swiss Olive Oil Panel SOP—as well as to their panel leaders: Richard Retsch from DOP and Martin Popp/Annette Bongartz from SOP.

Conflict of interest

The authors declare no conflict of interest.

Author details

Annette Bongartz[1*], Martin Popp[1] and Richard Retsch[2]

1 Zurich University of Applied Sciences (ZHAW)/Swiss Olive Oil Panel (SOP), Wädenswil, Switzerland

2 German Olive Oil Panel (DOP) e.V., Nuremberg, Germany

*Address all correspondence to: annette.bongartz@zhaw.ch

IntechOpen

References

[1] Commission Regulation (EEC) No 2568/91 of 11 July 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. 1991 [as amended]

[2] COI/T20/Doc. No 15/Rev. 10 2018. Sensory analysis of olive oil—Method for the organoleptic assessment of virgin olive oil. 2018

[3] COI/T28/Doc. No 1/Rev. 5 2019. Guidelines for the accomplishment of requirements of standard ISO 17025 of testing laboratories with particular reference to virgin olive oil. 2019

[4] Decision No Dec-III.1/113-VI/2020. Adopting the guidlines for the managament of virgin olive oil tasting panels in the event of a pandemic. 2020

[5] EN ISO/IEC 17025:2017. General requirements for the competence of testing and calibration laboratories. 2017

[6] Dinnella C, Pierguidi L, Spinelli S, Borgogno M, Toschi TG, Predieri S, et al. Remote testing: Sensory test during Covid-19 pandemic and beyond. Food Quality and Preference. 2022;**96**:104437. DOI: 10.1016/j.foodqual.2021.104437