Survey on Performance Evaluation in Non-Quantitative Proficiency Testing

Published: July 2024

Eurachem

Survey on performance evaluation in non-quantitative Proficiency Testing

Published: July 2024

Editors

Marzia Mancin
Kees van Putten
Ulla Tiikkainen

Composition of the non-quantitative PT Task Force

Stephanie Albarede**, CTCB, France
Brian Brookman, LGC, UK
Thorsten Helbig**, DRRR, Germany
Caroline Laurent, BIPEA, France
Conny Lerche**, QuoData, Germany
Marzia Mancin, IZSVe, Italy
Markus Obkircher, MERCK, Switzerland
Marina Patriarca, ISS, Italy
Piotr Robouch, EC-JRC, Belgium
Erika Sarkany, QualiCont, Hungary
Ulla Tiikkainen, Labquality, Finland
Kees van Putten, Qlip, The Netherlands
**Not a member of the Eurachem PT Working Group

Acknowledgements

This report has been produced by the Eurachem Proficiency Testing Working Group (PTWG) with the composition shown (right). The editors are grateful to all these individuals and organisations, who have contributed comments, advice and assistance.

Citation

This publication should be cited* as: “M. Mancin, U. Tiikkainen, and K. van Putten (Eds.) Eurachem Report: Survey on performance evaluation in non-quantitative Proficiency Testing.

Available from www.eurachem.org.

**Subject to journal requirements*

Survey on performance evaluation in non-quantitative PT

Published July 2024

Contents

Foreword

Eurachem Reports provide summaries of work conducted by Eurachem Working Groups or others on behalf of the Eurachem Executive. Eurachem reports are issued for information only and do not constitute guidance or statements of policy.

This report summarises activities and responses received in a survey undertaken in support of the Eurachem Proficiency Testing Working Group (PTWG). The survey aimed to gather information about the performance evaluation of non-quantitative proficiency testing (PT) with a view to providing improved guidance on the topic, and support potential harmonisation of approaches.

1 Introduction

The paper by Tiikkainen et al. [1] regarding the importance of laboratory performance evaluation in qualitative and interpretative proficiency testing (PT)) and External Quality Assessment (EQA) was published in 2022, following initial work by the Eurachem PTWG. Tiikkainen et al. [1] outline the key findings of the initial survey conducted in 2014 [2] and a literature review performed in 2020 on the statistical techniques used for the performance evaluation of qualitative and interpretative PT/EQA. Both the survey results and the literature review indicated that, despite the majority of scoring systems being based on simple "yes/no" or "absence/presence" responses, various types of performance evaluation criteria are employed for qualitative and interpretative PT/EQA schemes.

2 Investigative Process

In response to the findings presented in Tiikkainen et al. [1], the Eurachem PTWG decided to conduct further investigation. A Task Force was established, comprising (i) members from the Eurachem PTWG and (ii) external experts invited to join. The Task Force was tasked with developing a new survey on non-quantitative PT (i.e., the set of qualitative and interpretative PTs) to delve deeper into the issues highlighted in the aforementioned paper.

Multiple on-line meetings were held to discuss and finalise a comprehensive list of questions. The draft list was presented to the PTWG for feedback, comments, and eventual approval.

The final version of the survey (Annex 1) included questions pertaining to the following areas:

- Type of non-quantitative PT;
- Field of application;
- Methods used to define the assigned value (according to Annex B of ISO/IEC 17043 [3]);
- Methods employed for performance evaluation (according to Tiikkainen et al. [1]);
- Information regarding accreditation status.

3 Survey implementation

The survey was developed through the joint work of the Task Force, the Eurachem PTWG and the Observatory Office of the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe, a public veterinarian institute from Italy). The IZSVe Survey instrument, a web application based on the LimeSurvey Open Source Software [4], was utilised to ensure a user-friendly interface and efficient data collection process. The online questionnaire was organised into multiple sections, each referring to a type of PT/EQA scheme, enabling the creation of a tailored digital version based on the PT/EQA schemes used by individual respondents.

The survey was formally announced and launched during the 10th Eurachem PT workshop held in Windsor, UK during September 2023, with an initial submission deadline set for December 2023 (Annex 2). However, due to the limited response, the deadline was further extended until January 31, 2024. The dataset containing the survey data was downloaded in Excel format and subsequently made available to the Eurachem PTWG for the analysis.

The initial review of the survey results was presented at the Eurachem PTWG meeting in Alicante in March 2024 (Annex 3). This report summarises the key findings and outcomes of the project. The software R version 4.2.3 [5] was used for the result visualisation along with the "ggplot2" [6] and "VennDiagram" [7] packages for creating the bar-plots and the Euler-Venn diagram, respectively.

4 Main results

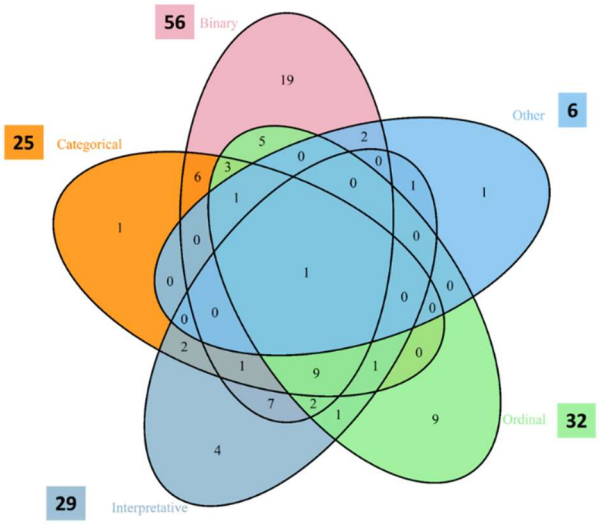
The survey included 124 responses, of which 34 were deleted (due to incomplete, nonsensical or duplicated replies). The remaining 90 acceptable records were analysed for the summary results. Among these, 76 pertained to non-quantitative PT, while 14 concerned quantitative PT. Focusing specifically on non-quantitative PT, 67 distinct PT providers participated in the survey. The majority of these PT providers were European, followed by American. Within Europe, France and Italy emerged as the most represented countries.

Type of non-quantitative PT

The answer type required from a PT provider defines the type of PT. In non-quantitative PT, the results can be binary, based on two possible outcomes (e.g. positive/negative, presence/absent); categorical, with more than two possible outcomes (e.g. blood group determination: A, B, AB); ordinal, with ordered outcomes, grades or rankings, or sensory evaluations (e.g. determination of cancer stage: I, II, III, IV, chemical reaction of type 1+, 2+, 3+); or interpretative, where no measurement is involved but a judgement within the participant's competence (e.g interpretation of an X-ray image by a clinician), as shown in Annex 1.

In the survey, the most prevalent type of non-quantitative PT was "binary", followed by "ordinal", "interpretative" and "categorical". Considering the possibility of multiple choice (with PT providers offering more than one type of non-quantitative PT), the most frequently encountered combinations were binary-categorical, binary-interpretative and binary-interpretative- categorical- ordinal (Figure 1).

Figure 1: Euler-Venn diagram showing the combination of non-quantitative PT offered.

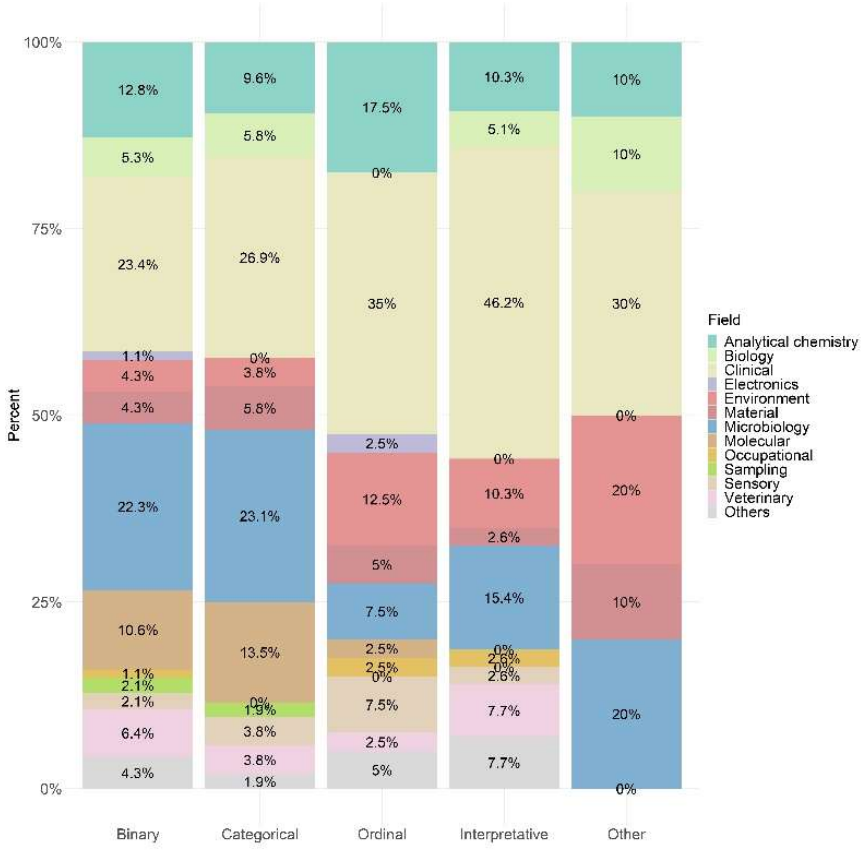


Field of application

The following fields were considered in the survey (Annex 1): analytical chemistry (agriculture, cosmetics, food & feed, veterinary, water), biology or pharmacology, clinical (chemistry/immunochemistry, cytology, haematology, pathology), electrotechnics/electronics, environment (air, soil, water), material or mechanical testing, microbiology (agriculture, clinical, consumer, cosmetics, food & feed, products, surface, veterinary, water), molecular biology (clinical, food & feed, veterinary), occupational safety, sampling, sensory testing, and veterinary (chemistry/immunochemistry, cytology, haematology, pathology).

The Clinical field had the highest number of non-quantitative PT offers, followed by microbiology, analytical chemistry, molecular and environmental. Analysing the distribution of fields by PT type, the clinical field dominated as the most represented field for each non-quantitative PT type, followed by microbiology for binary, categorical and interpretative PT and analytical chemistry for ordinal qualitative PT. (Figure 2).

Figure 2: Bar-plot showing the fields distribution by non-quantitative PT.



Methods used to define the assigned value

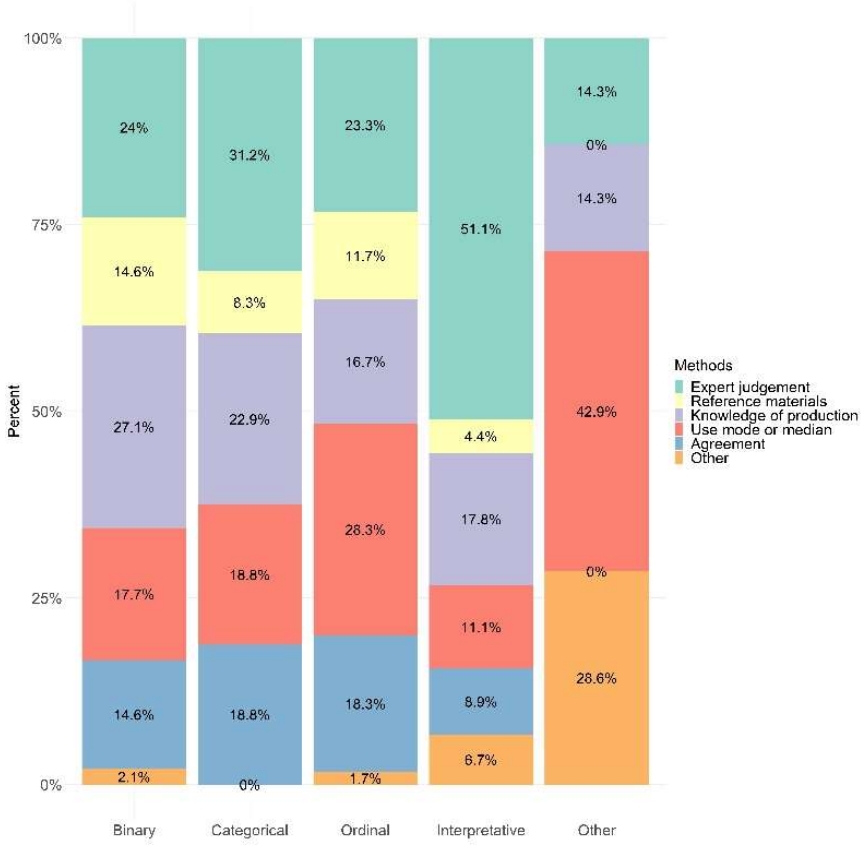
The list of methods used to define the assigned value was derived from ISO/IEC 17043. The survey showed that the assigned value for PT type varied. Expert judgement was employed for interpretative and categorical qualitative PT, knowledge of the origin or production for binary PT, and mode or median for ordinal qualitative PT. Expert judgement served as the second method for deriving the assigned value for binary and ordinal PT, while knowledge of the origin or production was the second method for categorical and interpretative PT (Figure 3).

Methods employed for performance evaluation.

The following methods for the performance evaluation have been identified by Tiikkainen *et al.* [1]:

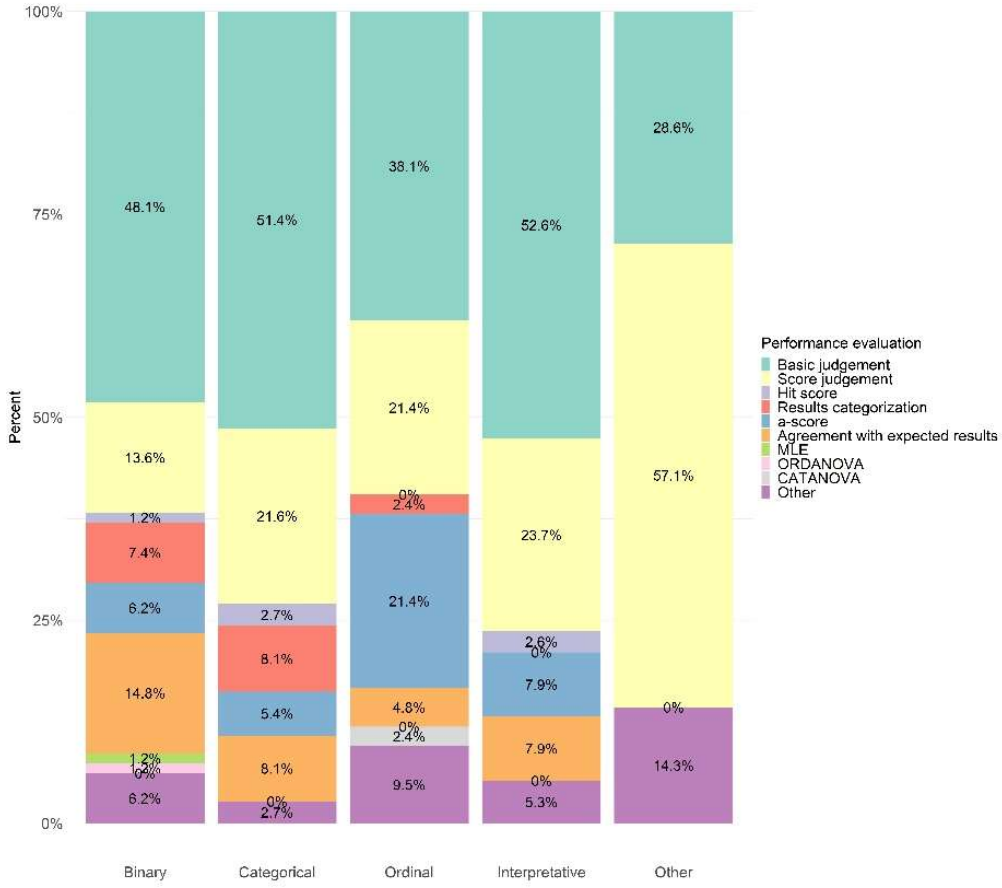
- Basic judgement [8,9];
- Score judgement [9];
- Hit score [10];
- Results categorization [11,12];
- Quantitative score, such as a-score [13];
- Agreement with expected results [14];
- Score by using the Maximum likelihood method [10,15,16];
- ORDANOVA: Equivalence of results from different laboratories, one factor [17]; and
- CATANOVA Equivalence of results from different laboratories and technicians: two factors [18].

Figure 3: Bar-plot showing the distribution of methods to derive the assigned value by non-quantitative PT.



The survey results indicated that the predominant methods to provide the performance evaluation by PT type were the basic judgment for all qualitative PT types. Score judgement was the second most common choice for categorical and interpretative PT. For ordinal PT, score judgement and a-score [13] were the second most common choice, whereas for binary, it is the agreement with the results (Figure 4).

Figure 4: Bar-plot showing the distribution of the performance evaluation methods by non-quantitative PT.



Information regarding accreditation status.

Analysing information regarding the number of PTs offered by each provider and the number of those accredited, it was observed that overall, more than 50 % of the PTs were accredited. Particularly, in case of binary qualitative PT, the percentage increased to 70 %. Conversely, the lowest percentage of accredited PTs corresponded to interpretative PT.

5 Conclusion

In conclusion, the survey conducted by the Eurachem PTWG yielded valuable insights into non-quantitative PT. The comprehensive approach adopted in designing and executing the survey provided a robust foundation for further analysis and action. Moving forward, the next deliverables include (i) a letter in Accreditation and Quality Assurance presenting the outcome of this survey, and (ii) a draft guidance document explaining statistical evaluation of qualitative PTs, similar to ISO 13528:2022 [19] for quantitative PT, which could provide input into the next revision of ISO 13528.

Acknowledgment

Appreciation is extended to the task force members, along with the Observatory Office of the Istituto Zooprofilattico Sperimentale delle Venezie (IZSVe), specifically to Marco Zago for his role as survey developer, Mirko Ruzza for web support, Stefania Crovato and Giulia Mascarello for their assistance in survey organization. Additionally, gratitude is expressed to the Eurachem PTWG for their valuable comments, and to all participants of the survey.

References

[1] Tiikkainen, Ulla, et al (2022). *Is harmonisation of performance assessment in non - quantitative proficiency testing possible/necessary?* Accreditation and Quality Assurance 27, 1-8. https://doi.org/10.1007/s00769-021-01492-6.

[2] Eurachem (2014). Google form: Eurachem survey on performance evaluation of qualitative and interpretative PT/EQA. https://bit.ly/3zzzLLg. Accessed on 08/July/2024.

[3] ISO/IEC 17043:2023. *Conformity assessment -General requirements for the competence of proficiency testing providers.* Issued by the International Standardization Organization, Geneva, Switzerland.

[4] LimeSurvey GmbH, Hamburg, Germany. http://www.limesurvey.org. Accessed on 08/July/2024.

[5] R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org. Accessed on 08/July/2024.

[6] Wickham H. (2016). *ggplot2 - Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://doi.org/10.1007/978-3-319-24277-4.

[7] Chen H (2022). VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.7.3. https://CRAN.R-project.org/package=VennDiagram. Accessed on 08/July/2024.

[8] Snell JJS (1985). *United Kingdom National external quality assessment scheme for microbiology*. Eur J Clin Microbiol 4:464–467. https://doi.org/10.1007/BF02014425.

[9] James VLA (2012). *Evaluating participant performance in qualitative PT/EQA schemes*. Accred Qual Assur 17:375–378.https://doi.org/10.1007/s00769-012-0895-1.

[10] Schilling P, Powilleit M, Uhlig S (2006). *Macrozoobenthos interlaboratory comparison on taxonomical identification and counting of marine invertebrates in artificial sediment samples including testing various statistical methods of data evaluation*. Accred Qual Assur 11:422–429. https://doi.org/10.1007/s00769-006-0139-3.

[11] Clarke J, Wilson JF (2005). *Proficiency testing (external quality assessment) of drug detection in oral fluid*. Forensic Sci Int 150:161–164. https://doi.org/10.1016/j.forsciint.2004.11.025.

[12] Chabirand A, Anthoine G, Pierson O, Hostachy B (2014). *The organization of proficiency testing in plant pathology (qualitative methods of analysis) according to the ISO/IEC 17043: example of the French national reference laboratory*. Accred Qual Assur 19:111–125. https://doi.org/10.1007/s00769-014-1034-y.

[13] Beavis G, Wilson J, Sykes M (2019). *Quantitative scores for binary qualitative proficiency testing*. Accred Qual Assur 24:263–269. https://doi.org/10.1007/s00769-019-01386-8.

[14] Mancin M, Barco L, Saccardin C, Ricci A (2015) *Proposed statistical analysis to evaluate qualitative proficiency testing of Salmonella serotyping*. Accred Qual Assur 20:305–310. https://doi.org/10.1007/s00769-015-1129-0.

[15] Uhlig S, Bläul C, Frost K, Sgorzaly S, Colson B, Simon K (2015) *Qualitative PT data analysis with easy-to-interpret scores*. Accred Qual Assur 20:347–353. https://doi.org/10.1007/s00769-015-1174-8.

[16] Bashkansky E, Turetsky V (2016) *Proficiency testing: binary data analysis.* Accred Qual Assur 21:265–270. https://doi.org/10.1007/s00769-016-1208-x.

[17] Bashkansky E, Gadrich T, Kuselman I (2012) *Interlaboratory comparison of test results of an ordinal or nominal binary property: analysis of variation*. Accred Qual Assur 17:239–243. https://doi.org/10.1007/s00769-011-0856-0.

[18] Gadrich T, Kuselman I, Andrić I (2020) *Macroscopic examination of welds: Interlaboratory comparison of nominal data*. SN Appl Sci 2:2168. https://doi.org/10.1007/s42452-020-03907-4.

[19] ISO 13528:2022, "*Statistical methods for use in proficiency testing by interlaboratory comparison*". Issued by the International Standardization Organization, Geneva, Switzerland.

Annex 1 - Survey on non-quantitative PT

Survey on non-quantitative PT/EQA - 2023

Introduction about the survey:

Dear Colleague,

The evaluation of participants' performance is a key issue of any proficiency testing (PT) or external quality assessment (EQA), and it is crucial that participants are assessed fairly and consistently. However, whereas ISO/IEC 17043 and ISO 13528 deal extensively with this issue for the PT/EQAs concerned with quantitative measurements, guidance related to qualitative and interpretative PT/EQA is limited. Eurachem* is currently organising a survey to collect information on the current practice applied by PT/EQA providers for the evaluation of participants' performance in qualitative and interpretative PT/EQA, with the aim to review current practices, identify elements for harmonization and provide guidance on this issue. As a PT/EQA provider, you are invited to share with us your strategies for performance evaluation by completing this questionnaire, so that your experience can be considered and become part of a larger picture of shared knowledge. The questionnaire could be completed at different times by clicking "Resume later" on the top-right of the page and providing the required information. A summary of the survey results will be made available as soon as possible to all those who take part. We would appreciate if you could provide us by email a copy of typical reports you send to participants for each type of PT (or other documents sent to participants to interpret their assessment).

(*) Eurachem (http://www.eurachem.org/) is a network of organisations in Europe aiming to promote measurement traceability and good quality practices in analytical sciences.

For any further information, please contact the PT WG Secretariat via the website [redacted]

Thank you for submitting your contribution by ** **December 15th, 2023** **

/ The Eurachem PT working group

Privacy information about "IZSVe Survey"
Pursuant to sect. 13 and 14 of (EU) Regulation 2016/679 (GDPR)
Data controller

Istituto Zooprofilattico Sperimentale delle Venezie, with registered office in Viale dell'Università 10 – 35020 Legnaro (PD), Italy, VAT 00206200289, in the person of its Director General legal representative, tel. [redacted], e-mail [redacted]

☐ **To continue please first accept our survey privacy policy.**

Name of PT provider*: .....

URL (web site) *: .....

Country *: .....

The following definitions are used in this survey:
Qualitative PT/EQA are schemes where the objective is to identify or describe one or more characteristics of the proficiency test item. The results of qualitative tests can be either **categorical** (or "**nominal**") values, e.g. identity of micro-organisms, identification of the presence of a specific analyte (such as a drug) or **ordinal** values, including, for example, responses such as grades or rankings, sensory evaluations, or the strength of a chemical reaction (e.g. 1+, 2+, 3+).

Interpretative PT/EQA are schemes where sets of data or other information are supplied and the information is processed to provide an interpretation (or other outcome). In interpretative tests, the "proficiency test item" is a test result (e.g. a descriptive morphology statement), a set of data (e.g. to determine a calibration line) or other set of information (e.g. a case study), concerning an interpretative feature of the participant's competence.

According to the previous definitions, do you provide qualitative/interpretative PT/EQA?*

- ☐ Yes
- ☐ No

(if No, the questionnaire stops. Message: Thank you for your feedback!!)

According to definitions below:

Binary results: two possible outcomes, e.g. positive/negative, presence/absent;

Categorical results: more than two possible outcomes, e.g. blood group determination: A, B, AB, O;

Ordinal results: ordered outcomes, grades or rankings, sensory evaluations, e.g. determination of cancer stage: I, II, III, IV, chemical reaction (e.g. 1+, 2+, 3+).

Interpretative results: no measurement is involved but a judgement within the participant's competence: interpretation of an X-ray image by a clinician.

Select the PT type of your PT/EQA (multiple choice. If your PT/EQA scheme has more than one types of answers (e.g. binary and interpretive), select all the relevant options)*:

- ☐ Qualitative: Binary
- ☐ Qualitative: Categorical
- ☐ Qualitative: Ordinal
- ☐ Interpretative
- ☐ Other

If other, please specify

How do you define the assigned value? (multiple choice) *

- ☐ By expert judgement
- ☐ By use of reference materials as PT items
- ☐ From knowledge of the origin of production of the PT item(s)
- ☐ Using the mode or median of participant results
- ☐ Agreement of a predetermined majority percentage of responses
- ☐ Other, specify

Which is the field of your PT/EQA? (multiple choice) *.

☐ Analytical chemistry [Agriculture, cosmetics, food & feed, veterinary, water]	☐ Environment [Air, soil, water]	☐ Occupational safety
☐ Biology, pharmacology	☐ Material or mechanical testing	☐ Sampling
☐ Clinical [Chemistry/immunochemistry, cytology, haematology, pathology]	☐ Microbiology [Agriculture, clinical, consumer, cosmetics, food & feed, products, surface, veterinary, water]	☐ Sensory testing
☐ Electrotechnics /Electronics	☐ Molecular biology [Clinical, food & feed, veterinary]	☐ Veterinary [Chemistry/immunochemistry, cytology, haematology, pathology]
☐ Other field, please specified		

How do you evaluate the performance? (multiple choice) *

	Ref
☐ Basic judgement (e.g. correct/wrong)	[1; 2]
☐ Score judgement (e.g.: 1 = Poor, 2 = Unsatisfactory, 3 = Satisfactory, 4 = Good, 5 = Very Good)	[2]
☐ Hit score	[3]
☐ Results categorization (e.g. true positive (TP), true negative (TN), sensibility (SE), specificity (SP), Accuracy (AC))	[4; 5]
☐ Quantitative score (e.g. a-score)	[6]
☐ Agreement with expected results (e.g. unweighted/weighted K of Cohen)	[7]
☐ Score by using the Maximum likelihood method (e.g. L-score)	[3; 8; 9]
☐ Equivalence of results from different laboratories: one factor (e.g. ORDANOVA)	[10]
☐ Equivalence of results from different laboratories and technicians: two factors (e.g CATANOVA)	[11]
☐ Other method, please specify and provide literature reference	

References

[1] https://doi.org/10.1007/BF02014425	[7] https://doi.org/10.1007/s00769-015-1129-0
[2] https://doi.org/10.1007/s00769-012-0895-1	[8] https://doi.org/10.1007/s00769-015-1174-8
[3] https://doi.org/10.1007/s00769-006-0139-3	[9] https://doi.org/10.1007/s00769-016-1208-x
[4] https://doi.org/10.1016/j.forsciint.2004.11.025	[10] https://doi.org/10.1007/s00769-011-0856-0
[5] https://doi.org/10.1007/s00769-014-1034-y	[11] https://doi.org/10.1007/s42452-020-03907-4
[6] https://doi.org/10.1007/s00769-019-01386-8	

How many PT schemes do you have? * .....

How many of them are accredited? * .....

Please, report any other comments/observations.

.....
.....

If you wish to share a typical PT report of yours, please send it by e-mail to **Eurachem-ptwg@gmail.com** and indicate the following subject of your email "**quali-survey-2023".** Your report will provide additional useful information that we will treat confidentially. Be assured that your report will be kept confidential, and will not be distributed.

If there is a need for clarification or interview, could we contact you? *

- ☐ No
- ☐ Yes

If yes,

Name of contact person*: .....

e-mail*: .....

_____

Thank you for your feedback!!

Annex 2 - Poster on non-quantitative PT presented at the Eurachem 10th PT Workshop in Windsor, UK

Eurachem

A Focus for Analytical Chemistry in Europe

Survey on non-quantitative PT/EQA, 2023

SCOPE

The evaluation of participants' performance is a key issue of any **proficiency testing (PT)** or **external quality assessment (EQA)**, and it is crucial that participants are assessed fairly and consistently. However, whereas ISO/IEC 17043 and ISO 13528 deal extensively with this issue for the PT/EQAs concerned with quantitative measurements, guidance related to qualitative and interpretative PT/EQA is limited. **Eurachem* is organising a survey** to collect information on the current practice applied by PT/EQA providers for the **evaluation of participants' performance in qualitative and interpretative PT/EQA**, with the aim to review current practices, identify elements for harmonization and provide guidance on this issue.

As a PT/EQA provider, you are invited to share your strategies for performance evaluation by completing this questionnaire, so that your experience can be considered and become part of a larger picture of shared knowledge. A summary of the survey results will be made available as soon as possible to all those who take part. We would appreciate if you could provide us by email a copy of typical reports you send to participants for each type of PT (or other documents sent to participants to interpret their assessment).

Thank you for submitting your contribution by **December 15th, 2023**.
/ The Eurachem PT working group.

(*) Eurachem is a network of organisations in Europe aiming to promote measurement traceability and good quality practices in analytical sciences. Contact the PT WG Secretariat for information at https://eurachem.org/index.php/contacts/contact-wg/142

DEFINITIONS

- **Proficiency Test**, PT
- **External Quality Assessment**, EQA
- **Qualitative PT/EQA** are schemes where the objective is to identify or describe one or more characteristics of the proficiency test item. The results of qualitative tests can be either categorical (or "nominal") values, e.g. identity of micro-organisms, identification of the presence of a specific analyte (such as a drug) or ordinal values, including, for example, responses such as grades or rankings, sensory evaluations, or the strength of a chemical reaction (e.g. 1+, 2+, 3+ etc.).
- **Interpretative PT/EQA** are schemes where sets of data or other information are supplied and the information is processed to provide an interpretation (or other outcome). In interpretative tests, the "proficiency test item" is a test result (e.g. a descriptive morphology statement), a set of data (e.g. to determine a calibration line) or other set of information (e.g. a case study), concerning an interpretative feature of the participant's competence.

Type of PT/EQA

- **Qualitative - Binary results**: two possible outcomes, e.g. positive/negative, presence/absent;
- **Qualitative - Categorical results**: more than two possible outcomes, e.g. blood group determination: A, B, AB, 0;
- **Qualitative - Ordinal results**: ordered outcomes, grades or rankings, sensory evaluations, e.g. determination of cancer stage: I, II, III, IV, chemical reaction (e.g. 1+, 2+, 3+, ...).
- **Interpretative results**: no measurement is involved but a judgement within the participant's competence: interpretation of an X-ray image by a clinician.

Do you provide qualitative / interpretative PT/EQA?

Which is the field of your PT/EQA? *(multiple choice)*

- ❑ Analytical chemistry [Agriculture, cosmetics, food & feed, veterinary, water]
- ❑ Biology, pharmacology
- ❑ Clinical [Chemistry/immunochemistry, cytology, haematology, pathology]
- ❑ Electrotechnics /Electronics
- ❑ Environment [Air, soil, water]
- ❑ Material or mechanical testing
- ❑ Microbiology [Agriculture, clinical, consumer, cosmetics, food & feed, products, surface, veterinary, water]
- ❑ Molecular biology [Clinical, food & feed, veterinary]
- ❑ Occupational safety
- ❑ Sampling
- ❑ Sensory testing
- ❑ Veterinary [Chemistry/immunochemistry, cytology, haematology, pathology]
- ❑ Other field, *please specify*

How do you define the assigned value? *(multiple choice)*

- ❑ By expert judgement
- ❑ By use of reference materials as PT items
- ❑ From knowledge of the origin of production of the PT item(s)
- ❑ Using the mode or median of participant results
- ❑ Agreement of a predetermined majority percentage of responses
- ❑ Other type, *please specify*

How do you evaluate the performance? *(multiple choice)*

- ❑ Basic judgement (e.g. correct/wrong);
- ❑ Score judgement (from 1 = "Poor" to 5 = "Very Good");
- ❑ Hit score;
- ❑ Results categorisation (e.g. true/false positive/negative);
- ❑ Quantitative score (e.g. *a*-score);
- ❑ Agreement with expected results (Cohen statistics);
- ❑ Maximum likelihood method (e.g. L-score);
- ❑ Equivalence of results from different laboratories: one factor (e.g. ORDANOVA) or two factors (e.g. CATANOVA);
- ❑ Other method, *please specify & provide literature reference*

Istituto Zooprofilattico Sperimentale delle Venezie

Survey designed by the Istituto Zooprofilattico Sperimentale delle Venezie (www.IZSVenezie.it)

www.eurachem.org

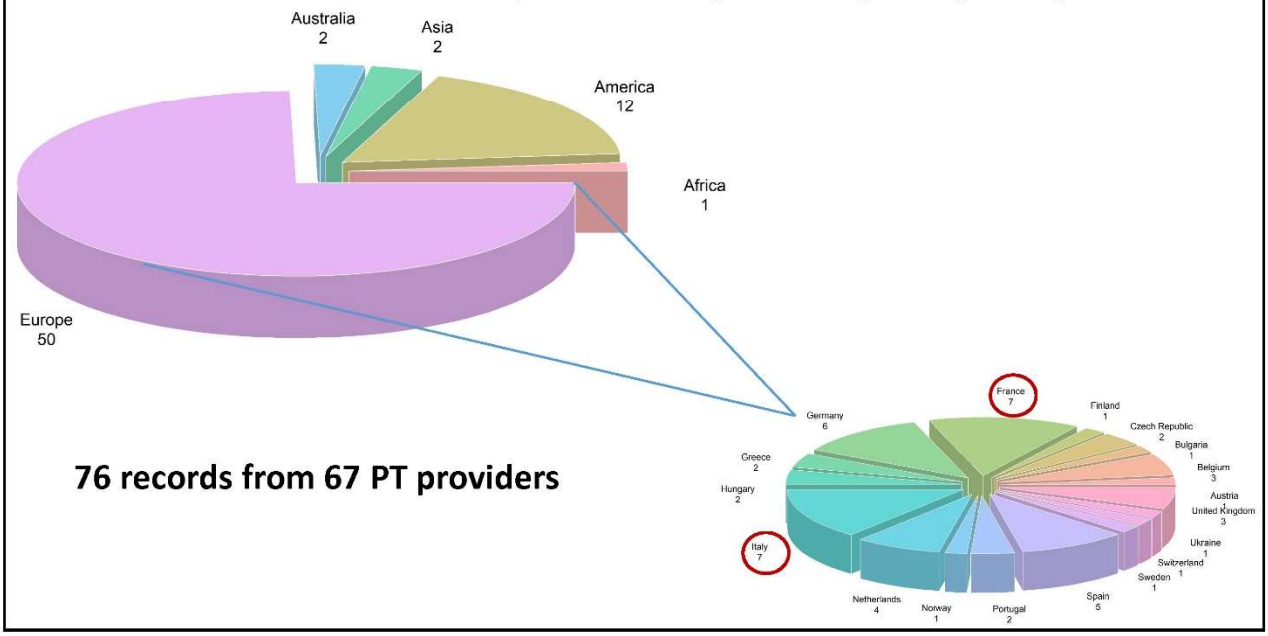
Annex 3 - Presentation of survey results to the Eurachem PT working group, Alicante

Survey results

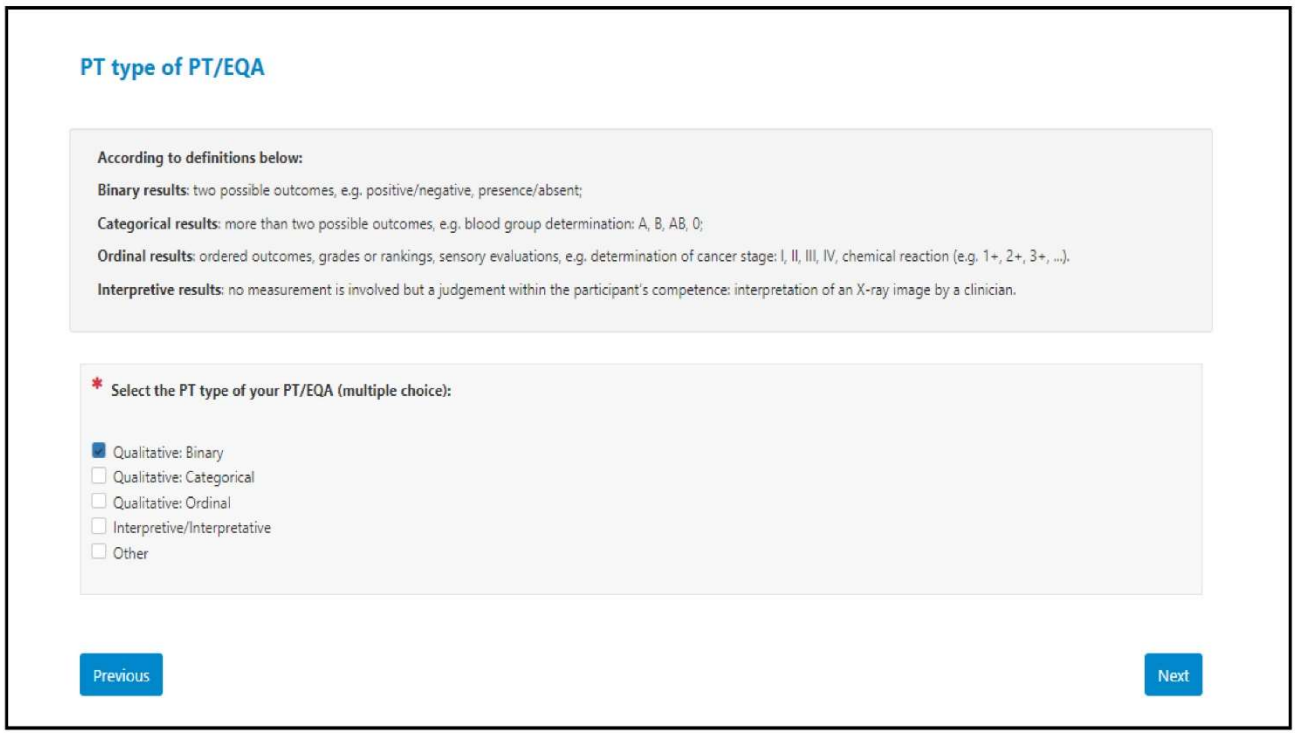
- **124 records were submitted**
- **34 records were deleted**
  - 25 incomplete answers;
  - 3 nonsensical answers;
  - 3 almost completed answers, but with unknown provider, website, country;
  - 3 duplicate records.

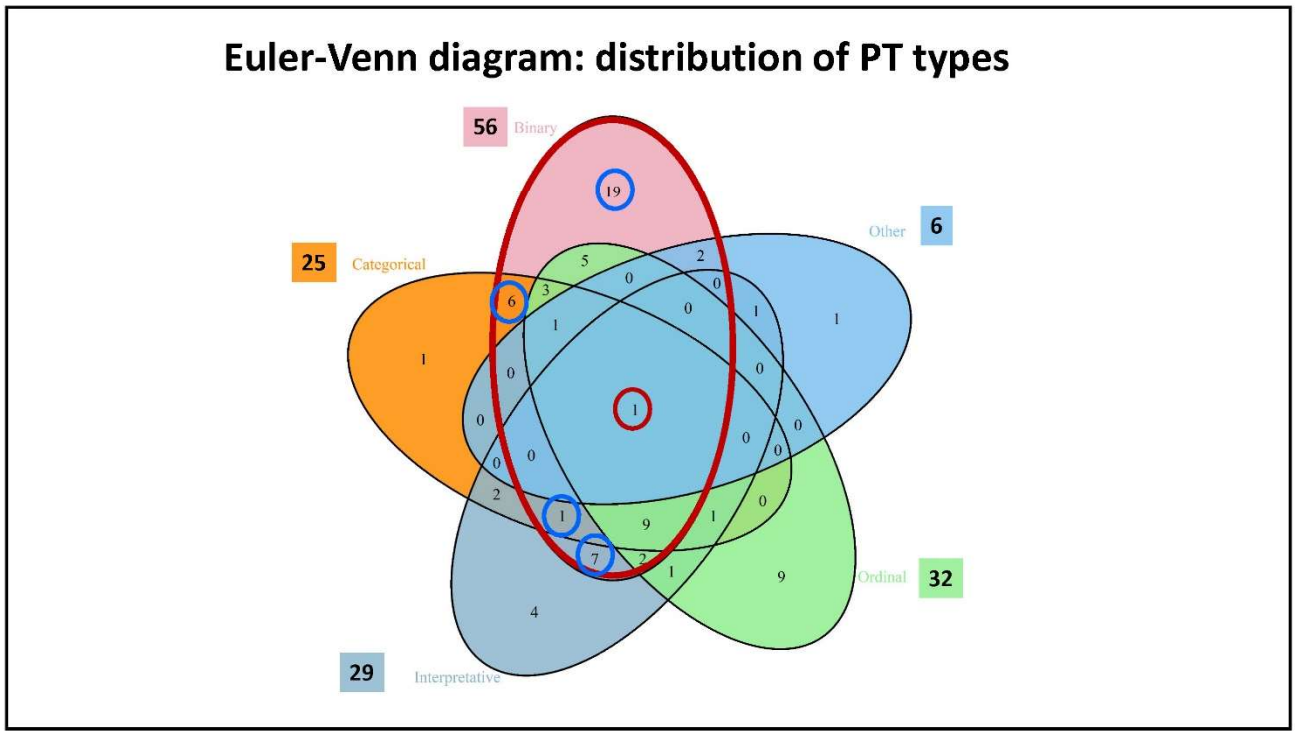
- **90 acceptable answers**
  - **76 organise non-quantitative PT**
  - 14 do NOT organise non-quantitative PT

Distribution of PT providers by country: frequency



76 records from 67 PT providers


PT type of PT/EQA
According to definitions below:
Binary results: two possible outcomes, e.g. positive/negative, presence/absent;
Categorical results: more than two possible outcomes, e.g. blood group determination: A, B, AB, 0;
Ordinal results: ordered outcomes, grades or rankings, sensory evaluations, e.g. determination of cancer stage: I, II, III, IV, chemical reaction (e.g. 1+, 2+, 3+, ...).
Interpretive results: no measurement is involved but a judgement within the participant's competence: interpretation of an X-ray image by a clinician.
* Select the PT type of your PT/EQA (multiple choice):
Qualitative: Binary
Qualitative: Categorical
Qualitative: Ordinal
Interpretive/Interpretative
Other
Previous
Next


Euler-Venn diagram: distribution of PT types
56 Binary
25 Categorical
Other 6
Ordinal 32
29 Interpretative
19
5
2
0
3
6
1
0
1
1
0
0
1
0
0
0
0
0
0
0
2
1
9
1
0
7
2
1
9
4

«Other PT types»

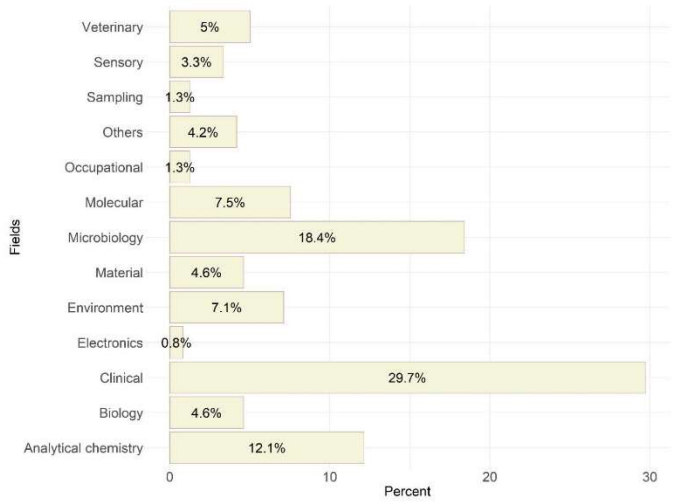
Other
Based on the set of measurements come to a clinical interpretation of the patient sample. No performance assessment on the basis of quantitative results but evaluation of the diagnostic pathway used and the final diagnosis.
Quantitative
Mechanical values like tensile strength, impact toughness, hardness, chemical composition
Multiple choice Questionnaire Sterilisation
Quantitative through measurements
Quantification of cells

Fields for each PT type

* **Qualitative: Binary - Which is the field of your PT/EQA? (multiple choice).**

- [x] Analytical chemistry [Agriculture, cosmetics, food & feed, veterinary, water]
- [ ] Biology, pharmacology
- [ ] Clinical [Chemistry/immunochemistry, cytology, haematology, pathology]
- [ ] Electrotechnics /Electronics
- [ ] Environment [Air, soil, water]
- [ ] Material or mechanical testing
- [ ] Microbiology [Agriculture, clinical, consumer, cosmetics, food & feed, products, surface, veterinary, water]
- [ ] Molecular biology [Clinical, food & feed, veterinary]
- [ ] Occupational safety
- [ ] Sampling
- [ ] Sensory testing
- [ ] Veterinary [Chemistry/immunochemistry, cytology, haematology, pathology]
- [ ] Other field, please specified

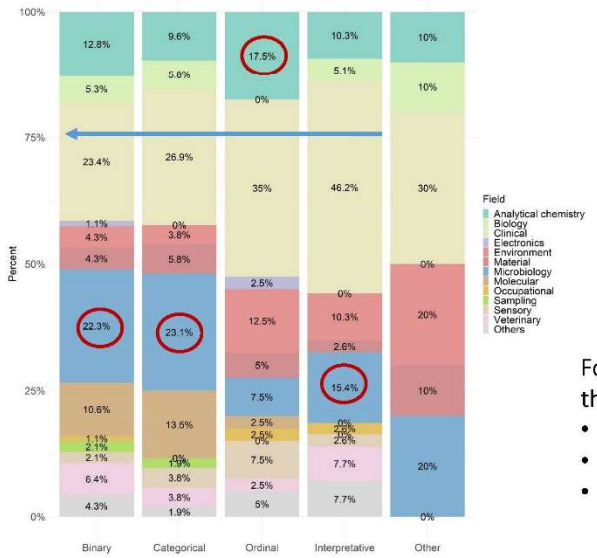
Distribution of fields



The most represented fields are:

1. Clinical;
2. Microbiology;
3. Analytical chemistry;
4. Molecular;
5. Environment.

Distribution of fields by qualitative PT types



For all types of answer,
the most represented fields are:

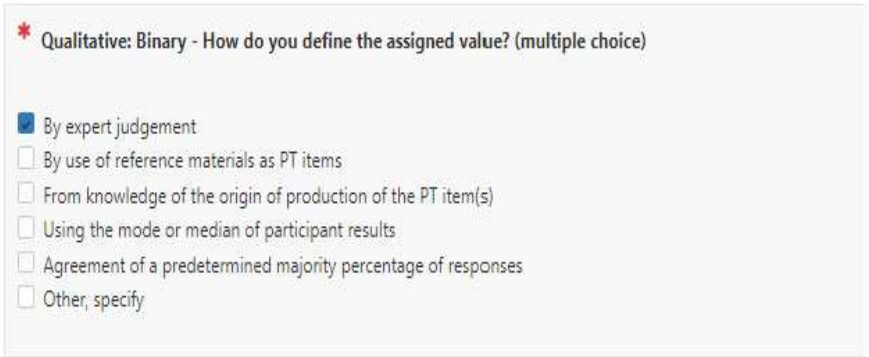
- Clinical;
- Microbiology (Binary, categorical, interpretive); and
- Analytical chemistry (ordinal).

«Other fields» for each PT type

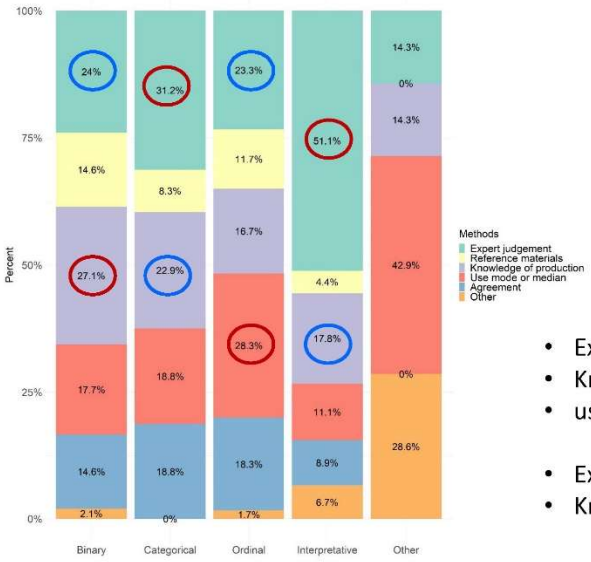
Other	
Binary	**Phytosanitary**
	Detection tests for food pathogens, e.g., presence of salmonella, Listeria, other pathogens like E, coli O157.
	Auto-immunity (Clinical)
	Physics (Calibration)
Categorical	**Auto-immunity; allergy**
Ordinal	**Phytosanitary**
	Auto-immunity, allergy
Interpretative	**Auto-immunity,** Pre analytical phase
	Fingerprints, Document examination, Digital Forensics
	Genetic and cytogenomic

Methods to obtain the assigned value by PT type

Qualitative Binary



Distribution of methods to derive the assigned value by PT type



- Expert judgement (interpretative and categorical);
- Knowledge of production (binary PT); and
- use mode or median (ordinal PT).

- Expert judgement (binary and ordinal); and
- Knowledge of production (categorical & interpretative).

«Other methods» to derive the assigned value for each PT type

Other	
Binary	by use a reference value obtained from other groups of laboratory that are able to determinate with other methods the identity and concentration of the substance.
	Using standards or guides.
Categorical	/
Ordinal	Average of participant results excluding outliers
Interpretative	**Using standards or guides.**
	No actual assigned value. The schemes (odour and taste in drinking water) are a combination of an ordinal scale (strength) and interpretive (description of taste). Some of the samples are of natural original origin and some have additives.
	Histotechnology and Immunhistochemistry Task: stain and detection of the marker molecules. In case of the technical phase-staining the tissue on the slide, there is no target value. The experts judge the quality of the staining with scores.

Methods to evaluate the PT performance by PT type

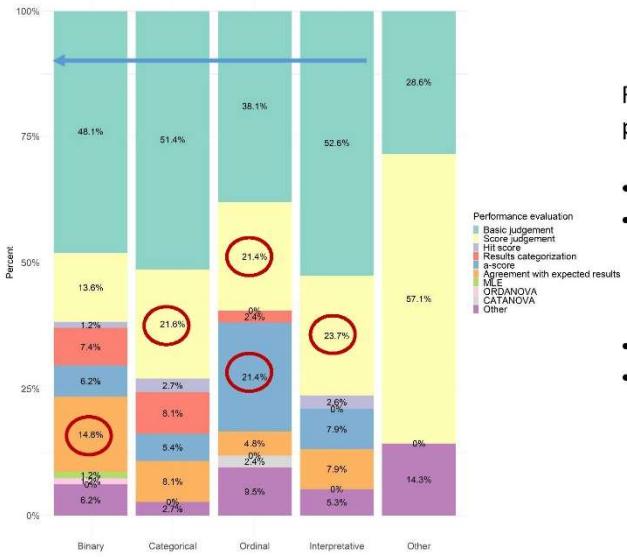
Qualitative: Binary - How do you evaluate the performance? (multiple choice)

References

[1] https://doi.org/10.1007/BF020 14425 [2] https://doi.org/10.1007/BF020 14425 [3] https://doi.org/10.1007/s00769-006-0139-3
[4] https://doi.org/10.1016/j.forsc iint. 2004.11.025 [5] https://doi.org/10.1007/s00769-014-1034-y [6] https://doi.org/10.1007/s00769-019-01386-8
[7] https://doi.org/10.1007/s00769-015-1129-0 [8] https://doi.org/10.1007/s00769-015-1174-8 [9] https://doi.org/10.1007/s00769-016-1208-x
[10] https://doi.org/10.1007/s00769-011-0856-0 [11] https://doi.org/10.1007/s42452-020-03907-4

- [x] Basic judgement (e.g. correct/wrong) [1;2]
- [x] Score judgement (e.g.: 1 = Poor, 2 = Unsatisfactory, 3 = Satisfactory, 4 = Good, 5 = Very Good) [2]
- [ ] Hit score [3]
- [ ] Results categorization (e.g. true positive (TP), true negative (TN), sensibility (SE), specificity (SP), Accuracy (AC)) [4;5]
- [ ] Quantitative score (e.g. a-score) [6]
- [ ] Agreement with expected results (e.g. unweighted/weighted K of Cohen) [7]
- [ ] Score by using the Maximum likelihood method (e.g. L-score) [3;8;9]
- [ ] Equivalence of results from different laboratories: one factor (e.g. ORDANOVA) [10]
- [ ] Equivalence of results from different laboratories and technicians: two factors (e.g CATANOVA) [11]
- [ ] Other method, please specify and provide literature reference

Distribution of the performance evaluation methods by PT type

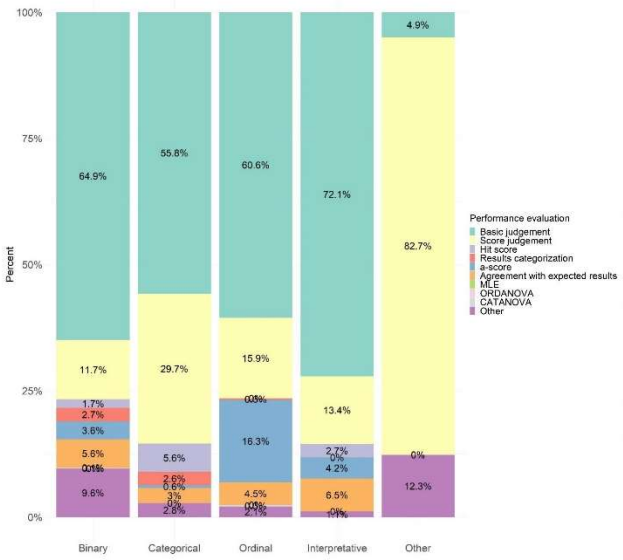


For all types of answer, the most used methods for performance evaluation are:

- Basic judgement (e.g. correct/wrong);
- Score judgement (e.g. 1 poor, 2 unsatisfactory, 3 satisfactory, 4 good, 5 very good) - for categorical, interpretative and ordinal.

- a-score is also relevant - for ordinal;
- Agreement with expected results for binary.

Distribution of the performance evaluation methods by PT type, weighted for number of PT provided



For all types of answer, the most used methods for performance evaluation are:

- Basic judgement (e.g. correct/wrong);
- Score judgement (e.g. 1 poor, 2 unsatisfactory, 3 satisfactory, 4 good, 5 very good) (Binary, categorical, interpretative);
- a-score (ordinal);
- Agreement with expected results (binary, interpretative);
- hit score (categorical); and
- score judgement (ordinal).

Distribution of the PT/accredited PT for method of performance evaluation and PT type

PT types:	Binary			Categorical			Ordinal			Interpretative			Total		
Method of perf.	N	Nacc	%Acc	N	Nacc	%Acc	N	Nacc	%Acc	N	Nacc	%Acc	N	Nacc	%Acc
Basic judgement	654	498	76.2	299	143	47.8	175	91	52.0	189	54	28.6	1317	786	59.7
Basic score	118	68	57.6	159	95	59.8	46	24	52.2	35	18	51.4	358	205	57.3
Hit score	17	0	0.00	30	0	0	0	0		7	7	100	54	7	13.0
Categorization	27	21	77.8	14	11	78.6	1	1	100	0	0		42	33	78.6
a-score	36	25	69.4	3	1	33.3	47	19	40.4	11	9	81.8	97	54	55.7
Agreement with expected results	56	35	62.5	16	2	12.5	13	1	7.7	17	3	17.7	102	41	40.2
MLE	1	0	0	0	0		0	0		0	0		1	0	0
Ordanova	1	1	100	0	0		0	0		0	0		1	1	100
Catanova	0	0		0	0		0	0		0	0		0	0	
Total	**910**	**648**	**71.2**	**521**	**252**	**48.4**	**282**	**136**	**48.2**	**259**	**91**	**35.1**	**1972**	**1127**	**57.2**

N: total number of PTs
Nacc: number of accredited PTs
%Acc: % accredited PT (= 100 * Nacc / N)

«Other performance evaluation» for each PT type

Other	
Binary	Satisfactory, unsatisfactory, questionable (in case of false positive result), congruent (The laboratory does not detect the analyte that is effectively present, as its method does not allow it. This is an information concerning the capability of the method)
	Judgement for each specimen: Acceptable, Unsatisfactory Overall performance for each survey (10 specimen/survey): Acceptable, Acceptable with Caution, Unsatisfactory
	Z-score
	direct percentage consensus
	NORMALIZED ERROR (En)
Categorical	Consensus
Ordinal	Most recent updated and consolidated version (28.06.2022) of Commission Regulation (EC) 152/2009 of 27 January 2009 laying down the methods of sampling and analysis for the official control of feed
	Z-score
	The schemes (odour and taste in drinking water) are a combination of an ordinal scale (strength) and interpretive (description of odour/taste) according to an older method of the Swedish Food Agency. For the ordinal scale we evaluate the participant's deviation from the mode. For example, for odour strength the standard used has a 5 step scale. The evaluation is -1 if the lab's response is one step lower than the mode, and +1 if one step lower than the mode.
Interpretative	The schemes (odour and taste in drinking water) are a combination of an ordinal scale (strength) and interpretive (description of odour/taste) according to an older method of the Swedish Food Agency. Since odour and taste are highly subjective, the description is currently provided in the report mainly as reference information. We provide feedback regarding the general agreement between labs and agreement with expected results.
	Case stories - we just state that "you have answered this..." and give the value or the decision made, and then report what others have reported.

Next steps on data evaluation

- Further evaluation relating the answer «Other» in the context of:
  - Type of qualitative PT;
  - Methods to derive the assigned value; and
  - Methods for the performance evaluation.