Brazilian Journal of Analytical Chemistry
Pre-publication – Accepted after peer review
doi: 10.30744/brjac.2179-3425.letter.oliveira.et.al

BrJAC

LETTER

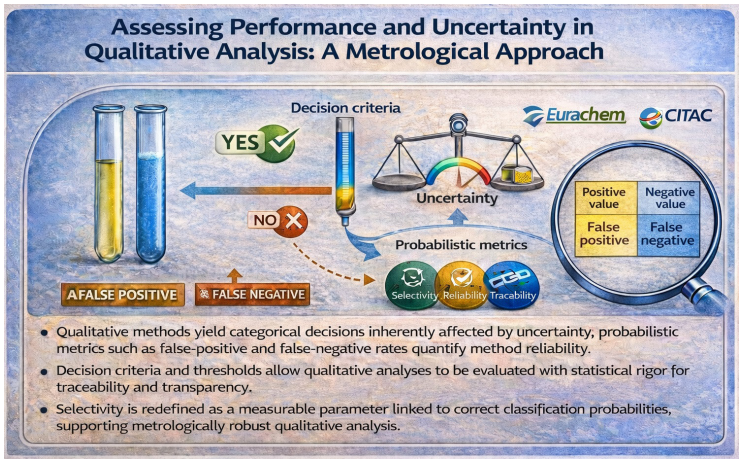
Assessment of Performance and Uncertainty in Qualitative Analytical Chemistry: A Metrological Approach based on the Eurachem/CITAC Guide

Ariadne Mayra Silva Rocha[1], Felipe Rebello Lourenço[2], Elcio Cruz de Oliveira[1,3*]

[1]Programa de Pós-Graduação em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro. Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 22451-900, Brazil

[2]Faculdade de Ciências Farmacêuticas, Universidade de São Paulo. Av. Professor Lineu Prestes, 580, Bloco 13 A, Butantã, São Paulo, SP, 05508-000, Brazil

[3]Logística, Planejamento e Controle Operacional, Medição e Gestão de Estoque de Produtos, PETROBRAS S.A. Av. Henrique Valadares, 28, Centro, Rio de Janeiro, RJ, 20231-030, Brazil



Graphical abstract created using ChatGPT (AI)

While the aim of "Quantifying Uncertainty in Analytical Measurement" is to provide a standardized framework and methodology for estimating and expressing uncertainty in quantitative analytical measurements, the Eurachem/CITAC guide for qualitative analysis expands the concept of uncertainty to categorical decisions, allowing qualitative analyses to be evaluated with the same metrological and statistical rigor.

Published in 2021, the Eurachem/CITAC Guide on Assessment of Performance and Uncertainty in Qualitative Chemical Analysis is considered as a key methodological reference for assessing the reliability of qualitative outcomes. Such qualitative tests yield categorical results, such as the presence or absence of a substance or the identification of a compound. Although such results do not yield a directly measured numerical value, they are not exempt from uncertainty. The guide proposes quantifying these uncertainties associated with the probability of false-positive or false-negative results to inform users of the analysis about the method's reliability limits.[1]

In this context, the document provides different methodological approaches that enable qualitative decisions to be treated with the same statistical rigor as quantitative analyses. It presents practical examples illustrating how these approaches can be implemented, recognizing that the outcome of a qualitative method, such as the identification of a compound or the confirmation of a substance's presence, is subject to errors and uncertainties that must be systematically evaluated and communicated.[2]

Defining the types of criteria used in qualitative analysis involves distinguishing between quantitative criteria, which entail converting a numerical value into a category (e.g., compliant or non-compliant based

Cite: Rocha, A. M. S.; Lourenço, F. R.; de Oliveira, E. C. Assessment of Performance and Uncertainty in Qualitative Analytical Chemistry: A Metrological Approach based on the Eurachem/CITAC Guide. *Braz. J. Anal. Chem.* (Forthcoming). http://dx.doi.org/10.30744/brjac.2179-3425.letter.oliveira.et.al

Submitted: October 29, 2025; **Revised:** December 22, 2025; **Accepted:** December 26, 2025; **Published online:** February 2026.

on a threshold), and qualitative criteria, such as color change, visual observation, or other indications of presence or absence. Although the guide primarily focuses on binary nominal classification (e.g., yes or no, present or absent), it acknowledges that, in certain categorical cases, classification can be reduced to correct or incorrect to apply the same principles. By explicitly defining the decision criteria and the statistical or probabilistic thresholds that delineate the boundary between categories, the guide ensures metrological traceability and transparency in the interpretation of results.

As a central tool for characterizing the performance of qualitative methods, the guide introduces metrics based on false results. Laboratories are advised to collect samples with known or reference results, apply the method under evaluation to these cases, and estimate the frequency with which the process fails to correctly classify an item.

From these results, the proportions of true positives, true negatives, false positives, and false negatives are calculated. These proportions allow the estimation of error probabilities and the construction of confidence intervals, thereby expressing statistical uncertainty. Treated as direct measures of uncertainty, these error rates provide qualitative analyses with a quantitative dimension of reliability, which is important when analytical results affect technical, regulatory, or scientific decisions.[3,4]

A crucial aspect mentioned by the guide is the representativeness and diversity of the cases used in performance estimation. If the test cases are too homogeneous or not representative of real conditions, the estimated error rates may underestimate or overestimate the actual uncertainty. Therefore, it is recommended to use test samples that cover the expected range of conditions, including different matrices, varying levels of interferents, and operational variations, to ensure the reliability estimate is robust.

This recommendation becomes especially relevant in methodologies based on spectroscopy and measurement instruments, where spectral variability and signal overlap among similar constituents pose practical challenges. By requiring sample sets that are representative of real analysis conditions, the guide promotes validation that reflects the natural variability of matrices and the presence of interferents, an essential element to realistically estimating the probability of incorrect classifications.[5-7]

In addition to representativeness, the number of samples used for performance estimation plays a critical role in the reliability and stability of validation results. Although no universally fixed minimum sample size can be defined since this number depends on the analytical objective, system complexity, and data variability, previous validation studies conducted under simple or simulated scenarios have shown that insufficient sample numbers can lead to unstable or overly optimistic estimates of error rates, particularly in qualitative and classification problems. As a general indication, validation studies require sample sizes on the order of several tens per class to achieve stable and convergent performance estimates, with larger sample sizes becoming necessary as model complexity increases.[8]

Therefore, representativeness and sample size should be considered jointly to ensure that estimated probabilities of incorrect classification realistically reflect the uncertainty associated with real analytical conditions.

When it comes to selectivity, the guide allows for a reinterpretation of this traditional concept. Instead of remaining a purely qualitative attribute, selectivity is viewed as a measurable parameter linked to the probability of correctly distinguishing between samples containing similar compounds or interferents. Thus, evaluating selectivity becomes an exercise in pattern recognition and probabilistic classification, where performance is expressed by the method's ability to assign each sample to its category correctly. This perspective broadens the understanding of selectivity, associating it with the statistical reliability of the response rather than solely the absence of visual or instrumental interference.[9-13]

The practical application of this methodology is especially relevant in non-targeted analyses, where the goal is to detect complex patterns and identify substances across diverse matrices without predefined targets. In such cases, the method must correctly differentiate signals corresponding to distinct compounds, even in the presence of instrumental noise and spectral overlap. Recent studies illustrate this by developing an automated approach for identifying microplastics using Raman spectroscopy, addressing the challenges of spectral variability and signal overlap among similar polymers. Instead of relying on an arbitrary similarity value, the method used a correlation distribution obtained via bootstrap sampling to determine the practical

acceptance threshold, aligning with the guide's recommendation to base qualitative decisions on probabilistic metrics and explicit performance assessments. This statistically controlled approach has been shown to significantly reduce classification errors, providing known confidence levels for each decision and bringing the analytical process closer to a metrologically traceable system where each decision is supported by uncertainty and performance estimates.[14]

When Raman spectroscopy is applied to quantify species in reactive mixtures, as in studies of urea and thiourea, adherence to the principles outlined in the guide is essential. Although the primary objective of such studies is to estimate concentrations from Raman signals quantitatively, a qualitative component remains inherent in spectral assignment, band identification, and the differentiation of interfering signals, steps that carry a risk of interpretative error. The guide emphasizes that any implicit qualitative decision, for instance, assigning a peak to a specific vibrational mode or determining whether a signal belongs to the analyte or to noise, is subject to uncertainty, and this uncertainty should be expressed in terms of error probabilities. In the urea/thiourea system, this involves evaluating the likelihood of misinterpretations—such as mistaking an interfering band for the analyte or overlooking weak peaks buried in noise—and designing validation experiments that account for varying compound ratios, noise levels, and instrumental conditions.[15]

Moreover, the guide recommends that qualitative conclusions, such as band assignments or confirmation of analyte presence, be accompanied by a confidence statement or an estimate of the local error probability derived from previously assessed spectral error rates, thereby enhancing transparency and discouraging absolute interpretations of spectroscopic signals. The systematic implementation of these guidelines also fosters a culture of continuous performance monitoring in laboratories. By acknowledging that even seemingly unambiguous spectral assignments may fail, researchers are encouraged to establish controls, retest protocols, and periodically review the criteria used to discriminate signals.[16,17]

Beyond individual method performance, interlaboratory comparability of qualitative decisions is equally critical to ensure reproducibility across different analytical contexts. In this regard, the joint IUPAC/CITAC (2025) guide enhances metrological capacity by proposing a statistical framework specifically designed for analyzing agreement in categorical results obtained across laboratories, operators, or instruments. Tools such as CATANOVA (Categorical Analysis of Variance) and ORDANOVA (Ordinal Analysis of Variance) enable the treatment of nominal and ordinal variables, respectively, quantifying the degree of agreement among classifications and identifying sources of systematic variability between laboratories. This statistical evaluation is crucial for validating automated decision systems in qualitative methods. Consequently, it ensures that a method not only performs reliably within a single laboratory but also yields equivalent, traceable results across diverse analytical environments—an indispensable requirement for the international recognition of qualitative measurements.[18]

Furthermore, the consolidation of metrological approaches to qualitative methods has expanded across various domains, reflecting the need to transform descriptive judgments into traceable, comparable decisions. In environmental monitoring systems, the use of portable devices, assay kits, and continuous sensors highlights the importance of incorporating performance and uncertainty considerations into the interpretation of detection or non-detection results.[19]

In toxicological, forensic, and genetic sequencing contexts, harmonised procedures and the application of statistical metrics enable quantification of interlaboratory variability and reduction of classification errors.[20,21] In the assessment of pharmaceutical equivalence and clinical diagnostic testing, risk models and sensitivity and specificity metrics reinforce the need to track the reliability of qualitative decisions.[22,23]

Finally, studies involving bottom-up uncertainty estimation in complex matrices and interlaboratory comparisons demonstrate that selectivity and agreement among categorical results can be addressed with statistical rigor, thereby consolidating qualitative metrology as a structured discipline applicable across multiple analytical fields.[24-26]

The relevance of these principles and tools transcends specific domains, extending well beyond laboratory contexts. In sectors such as food and beverages, authentication and origin traceability depend strongly on spectroscopic and multivariate methods, in which trace elements and spectral profiles are used as markers

to distinguish regions or products. Despite the high performance of supervised classification models, the natural variability of elements and the overlap of compositional features can lead to false positives or negatives, which are precisely the central metrics discussed in the Eurachem and IUPAC/CITAC guides.

The explicit incorporation of performance and uncertainty measures could therefore enhance the validation of these methods, making results more comparable across laboratories and legally more defensible in cases of fraud or commercial dispute.[27]

A comparison among these contexts generally reveals significant conceptual convergence, with all cases exhibiting a clear transition from descriptive approaches to quantitatively validated systems for qualitative decision-making. The central principle that emerges is that every qualitative decision is, in essence, a statistical inference and must therefore be accompanied by explicit measures of performance, uncertainty, and reproducibility. This perspective establishes a new paradigm for analytical chemistry, in which qualitative methods are no longer merely screening tools but rather integral components of the metrological domain, characterized by traceability, comparability, and transparency.

The scope of the guide is broad, reflecting a modern approach to the determination of neurotransmitters. The determination and correct interpretation of the limits of detection (LoD) and quantification (LoQ) are essential aspects for ensuring the reliability and sensitivity of the proposed analytical method. In the study, achieving extremely low LoD and LoQ values demonstrates not only the system's high instrumental performance but also its capability to detect and quantify analytes at trace levels, which is fundamental for biological and neurochemical applications.

Thus, defining LoD and LoQ within the context of this work not only confirms the procedure's sensitivity but also reinforces the method's metrological robustness and suitability for its analytical purpose, in accordance with internationally advocated principles of performance and uncertainty.[28]

Finally, incorporating the recommendations from the guides promotes a culture of quality within laboratories: it does not blindly assume that a qualitative method is infallible. Instead, it acknowledges that it can and should be accompanied by reasonable estimates of reliability. This stance strengthens the credibility of results, supports analytical risk management, guides decisions on confirming or investigating borderline cases, and encourages continuous improvement of procedures, periodic method validation, and quality control strategies adapted to non-targeted analyses.

In this way, selectivity ceases to be merely a technical property and becomes a performance metric with direct implications for uncertainty assessment and the reliability of analytical decisions. This consolidation elevates qualitative metrology to the status of a mature scientific discipline, indispensable for ensuring confidence in analytical decisions at both local and international levels.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The author E. C. de Oliveira is grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant No. 305479/2021-0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 001).

REFERENCES

(1) da Silva, R. B.; Ellison, S. L. R. (Eds.) Eurachem/CITAC Guide: Assessment of performance and uncertainty in qualitative chemical analysis. 1st Edition, Eurachem (2021). ISBN 978-0-948926-39-6.

(2) Carmen, A. Quantitative and qualitative methods, and primary methods. In: Botana, L. M. (Ed.) *Environmental Toxicology: Non-bacterial Toxins*. De Gruyter, Berlin, Boston, 2024. Chapter 3, pp 65-102. https://doi.org/10.1515/9783111014449-003

(3) Bertanha, M. L. G.; Lourenço, F. R. Total combined risk of false conformity decisions in pharmaceutical equivalence of acetaminophen oral solution. *Microchem. J.* **2024**, *201*. https://doi.org/10.1016/j.microc.2024.110610

(4) Couto, A. R.; Passarin, P. B. S.; Lourenço, F. R. Risk of false conformity decisions in LAL gel-clot test due to false-negative results. *Chemom. Intell. Lab. Syst*. **2023**, *234.* https://doi.org/10.1016/j.chemolab.2023.104760
(5) Pradella, M. New Guides for Uncertainty of Qualitative Results. *J. Appl. Lab. Med.* **2023**, *8* (1), 217-218. https://doi.org/10.1093/jalm/jfac087
(6) Pradella, M. Criticalities of Eurachem/CITAC guide uncertainty of qualitative results. *Measurement* **2022**, *193*, 110911. https://doi.org/10.1016/j.measurement.2022.110911
(7) Pradella, M. Positive and negative: guides to qualitative examinations. *Riv. Ital. Med. Lab.* **2022**, *8*. https://doi.org/10.23736/S1825-859X.22.00139-6
(8) Brereton, R. G. Consequences of sample size, variable selection, and model validation and optimization, for predicting classification ability from analytical data. *TrAC, Trends Anal. Chem.* **2006**, *25* (11), 1103-1111. https://doi.org/10.1016/j.trac.2006.10.005
(9) Ohn, H. M. Practical aspect of measurement uncertainty in a conventional testing microbiology laboratory. *Accredit. Qual. Assur.* **2025**. https://doi.org/10.1007/s00769-025-01674-6
(10) Pomerantsev, A. L.; Rodionova, O. Ye. Selectivity in Nontargeted Qualitative Analysis. *Anal. Chim. Acta* **2024**, *1332*. https://doi.org/10.1016/j.aca.2024.343352
(11) Kalichkina, L. E.; Krivolapenko, P. K.; Tuguldurova, V. P. Quantitative Raman Spectroscopy of Urea and Thiourea in the Reaction Mixtures of Allantoin and 4,5-Dihydroxyimidazolidine-2-Tione Formation. *J. Raman Spectrosc*. **2024**, *55* (11), 1165-1178. https://doi.org/10.1002/jrs.6729
(12) Pradella, M. The SIPMeL recommendations for ISO accreditation of validity monitoring of examination results. *Riv. Ital. Med. Lab*. **2024**. https://doi.org/10.23736/S1825-859X.24.00244-5
(13) Pereira, P.; Seghatchian, J. Spotlights on the trends in performance assessment of qualitative in vitro diagnostic medical devices in transfusion medicine. *Transfus and Apheresis Science* **2024**, *63* (2). https://doi.org/10.1016/j.transci.2024.103887
(14) Kalichkina, L. E.; Krivolapenko, P. K.; Tuguldurova, V. P. Quantitative Raman Spectroscopy of Urea and Thiourea in the Reaction Mixtures of Allantoin and 4,5-Dihydroxyimidazolidine-2-Tione Formation. *J. Raman Spectrosc.* **2024**, *55* (11), 1165-1178. https://doi.org/10.1002/jrs.6729
(15) Trapmann, S.; Ellison, S. L. R.; Davis, W. C.; Possolo, A.; Zheng, J.; Botha, A. Approaches for the production of reference materials with qualitative properties—The new International Standard ISO 33406. *Accredit. Qual. Assur.* **2024**, *29*, 269–276. https://doi.org/10.1007/s00769-024-01602-0
(16) Pradella, M. New ISO standards for medical biology laboratories, prescriptions and deviations. *Ann. Biol. Clin.* **2022**, *80* (5), 451–453. https://doi.org/10.1684/abc.2022.1755
(17) Kuselman, L.; Gadrich, T.; Pennecchi, F. R.; Hibbert, D. B.; Semenova, A. A.; Botha, A. IUPAC/CITAC guide: interlaboratory comparison of categorical characteristics of a substance, material, or object (IUPAC Technical Report). *Pure Appl. Chem*. **2025**, *97* (7) 715-750. https://doi.org/10.1515/pac-2025-0408
(18) Guigues, N.; Uysal, E.; Raveau, S.; Magar, J.; Billon, G.; Kahiluoto, J.; Schneider, R. J.; Liv, L.; Tosun, S. G.; Canteau, G.; Lalere, B. Insights on the validation of alternative tools for water quality monitoring: the case of on-site test kits, portable devices and continuous measuring devices. *Accredit. Qual. Assur.* **2024**, *29*, 163–173. https://doi.org/10.1007/s00769-023-01570-x
(19) Pradella, M. Toxicological-forensic and medicolegal examinations: proposals for harmonizati. *Riv. Ital. Med. Lab.* **2024**, *20* (1), 62-66. https://doi.org/10.23736/S1825-859X.23.00218-9
(20) Pradella, M. Quality of massively parallel sequencing/next-generation sequencing (MPS/NGS) methods. *Riv. Ital. Med. Lab.* **2023**, *19* (2), 138-144. https://doi.org/10.23736/S1825-859X.23.00190-1
(21) Caffaro, A. M.; Lourenço, F. R. Total combined global risk assessment applied to pharmaceutical equivalence – A case study of ofloxacin medicines. *Chemom. Intell. Lab. Syst.* **2023**, *241*. https://doi.org/10.1016/j.chemolab.2023.104935
(22) Rodrigues, C. O.; Olivares, I. R. B. Application of Quality Statistical Tools for the Evaluation of Diagnostic Tests for SARS-CoV-2 Detection. In: Crespilho, F. N. (Ed.). *COVID-19 Metabolomics and Diagnosis*. Springer, Cham. 2023, pp 175–192. https://doi.org/10.1007/978-3-031-15889-6_8

(23) Morgado, V.; Palma, C.; Da Silva, R. J. N. B. Bottom-Up Evaluation of the Uncertainty of the Quantification of Microplastics Contamination in Sediment Samples. *Environ. Sci. Technol*. **2022**, *56* (15), 11080–11090. https://doi.org/10.1021/acs.est.2c01828
(24) Jobbágy, V.; Dupuis, E.; Emteborg, H.; Hult, M. Performance evaluation of a European wide proficiency test on gross alpha/beta activity determination in drinking water. *Appl. Radiat. Isot.* **2022**, *186*, 110304. https://doi.org/10.1016/j.apradiso.2022.110304
(25) Gadrich, T.; Kuselman. I.; Pennecchi, F. R.; Hibbert, D. B.; Semenova, A. A.; Cheow, P. S.; Naidenko, V. N. Interlaboratory comparison of the intensity of drinking water odor and taste by two-way ordinal analysis of variation without replication. *J. Water Health* **2022**, *20* (6), 1005–1016. https://doi.org/10.2166/wh.2022.060
(26) Albu, P. C.; Tanczos, S-K.; Ferencz, A.; Pîrțac, A.; Grosu, A. R.; Pașcu, D.; Grosu, V-A.; Bungău, C.; Nechifor, A. C. pH and Design on n–Alkyl Alcohol Bulk Liquid Membranes for Improving Phenol Derivative Transport and Separation. *Membranes* **2022**, *12* (4), 365. https://doi.org/10.3390/membranes12040365
(27) Ghidotti, M.; Papoci, S.; Respaldiza, A.; Emteborg, H.; Ulberth, F.; Guntiñas, M. B. L. C. Use of energy dispersive X-ray fluorescence to authenticate European wines with protected designation of origin. Challenges of a successful control system based on modelling. *Food Chem*. **2025**, *465* (Part 1). https://doi.org/10.1016/j.foodchem.2024.141989
(28) Baranyi, M.; Karácsony, O.; Sperlágh, B. Analytical Method for Gradient Separation of Monoamine, Adenine Nucleotide, and β-Endorphin Neurotransmitters Using Amperometric and Spectrometric Detection. *Sep. Sci. plus.* **2025***, 8 (10).* https://doi.org/10.1002/sscp.70137