



Cite this: DOI: 10.1039/d6rp00078a

Learning to trust experimental data: a validation-based approach to experiment design in pre-service chemistry teacher education

Saule Zhunissova, * Leilya Zhussupova,  Gulmira Abyzbekova  and Bakytbek Islambekuly 

In laboratory chemistry education, experimentation is still often employed as a means of obtaining a “correct” result, while the grounds for trusting data and the conditions under which it is interpreted remain implicit for learners. Even in project-oriented formats, students often lack a methodological language that would allow them to comprehend experimental results as justified and context-dependent knowledge. To address this gap, the present study explores how key features of analytical method validation, specificity, linearity, precision, and trueness, can be pedagogically reinterpreted to support students’ reasoning about experimental data and design. This approach was implemented in an educational workshop with pre-service chemistry teachers, during which participants developed and justified a spectrophotometric method for measuring the total flavonoid content in a plant matrix. A qualitative analysis of students’ written reflections was conducted to examine how engagement with validation characteristics influenced their reasoning about experimental results and the justification of methodological choices. The findings suggest that introducing analytical validation as a conceptual framework for analysing experimental data encourages students to interpret results through the lens of measurement quality, relate method features to the reliability of analytical outcomes, and apply validation criteria when designing experimental procedures. These results imply that analytical validation can serve not only as a professional analytical process but also as a methodological foundation for fostering experimental reasoning and making trust in experimental data an explicit focus in chemistry education.

Received 13th February 2026,
Accepted 12th May 2026

DOI: 10.1039/d6rp00078a

rsc.li/cerp

Introduction

In international chemical education practice, a notable trend has emerged: an increased focus on developing competencies related to designing and independently executing scientific experiments (Viera *et al.*, 2017). Integrating elements of experimental design is considered crucial for cultivating research-oriented thinking (Bain *et al.*, 2019; Mitarlis *et al.*, 2020; Kwitonda *et al.*, 2021; Alrasheed and Alghamdi, 2023; Suradika *et al.*, 2023). However, the specific methods of integration vary. Some studies demonstrate the successful implementation of project-based learning, leading to greater student autonomy and improved academic performance (Ayaz and Söylemez, 2015; Viera *et al.*, 2017; Bain *et al.*, 2019; Tümay, 2023; Deffner and Hermanns, 2025). Meta-analyses, in particular, confirm the sustained advantages of these approaches compared to traditional instruction (Ayaz and Söylemez, 2015). Conversely, other

research indicates that short-term or isolated project assignments often maintain a recipe-driven approach to experimentation, failing to foster the development of robust experiment design skills (Szalay and Tóth, 2016; Szalay *et al.*, 2024). Further research reveals that even when project elements are present, students may still struggle with data interpretation and constructing evidence-based arguments (Bicak *et al.*, 2021). This suggests that the critical factor is not simply including project elements, but incorporating pedagogical strategies that genuinely promote the development of experimental thinking. Consequently, there is growing interest in instructional approaches that encourage engagement in experiment design (Márquez-Barreto *et al.*, 2023; Zhao *et al.*, 2023), the ability to independently develop experimental procedures (Farley *et al.*, 2021; Scoggin and Smith, 2023), and, in some instances, the creation of original methods (Viera *et al.*, 2017; Farley *et al.*, 2021; Scoggin and Smith, 2023; van Wyk *et al.*, 2025).

The Next Generation Science Standards (NGSS) documents (National Research Council, 2013) and the OECD Programme for International Student Assessment (PISA) scientific literacy

Korkyt Ata Kyzylorda University, Institute of Natural Sciences, 29A Aiteke Bi Street, Kyzylorda, 120014, Republic of Kazakhstan. E-mail: zhunissovasaule7@gmail.com



framework (Organisation for Economic Co-operation and Development (OECD, 2023) both emphasise that scientific and pedagogical thinking relies on the ability to plan investigations, control variables, ensure appropriate experimental conditions, and interpret data. These skills are particularly significant for future chemistry teachers, as they determine their readiness to teach meaningful experimentation and evidence-based analysis of results (Bicak *et al.*, 2021; Deffner and Hermanns, 2025; Riemer *et al.*, 2025). Given the increasing importance of experimentation in fostering students' research culture, aspiring chemistry teachers have a particularly acute need for reliable methods for analysing their own data. However, research indicates that students often experience uncertainty even when laboratory work is carefully executed. Despite gaining experience, they tend to focus on the mechanical execution of experimental procedures while remaining unsure of their results (Robinson, 2013; Hamer *et al.*, 2024). Insufficient preparation has also been noted in core research skills such as problem-solving (Ferreira and Trudel, 2012; McLaughlin *et al.*, 2024), data evaluation (Worrall *et al.*, 2020; Bicak *et al.*, 2021) and argument construction (Sampson *et al.*, 2011; Stephenson *et al.*, 2020). Thus, a tension arises between the recognised importance of experimental design and the actual capacity of students and future teachers to confidently analyse data. This highlights the need for methods and tools that not only enable the acquisition of measurements but also support a structured understanding of experimental quality and trustworthiness, thereby forming stable research foundations for subsequent professional practice.

Contemporary instructional formats in chemistry actively develop skills related to experimental design (Zhunisova *et al.*, 2025), however, these frameworks could benefit from additional support for students' engagement with data trustworthiness. While project-based activities promote the active development of skills in planning and conducting experiments, they do not always include structured means for evaluating data trustworthiness, creating a didactic gap between procedural execution and critical analysis. As noted by Talanquer, scientifically grounded chemical reasoning relies on structured cognitive methods that enable students to explain system behaviour, interpret data, and justify conclusions (Talanquer, 2018). Studies emphasise the importance of discussing differences in results (Viera *et al.*, 2017; Eymur and Çetin, 2024; McLaughlin *et al.*, 2024; Moutsakis *et al.*, 2025), sample quality (Araujo, 2009; Riemer *et al.*, 2025), relationships between variables (Bicak *et al.*, 2021; Reith and Nehring, 2024), and the logic of conclusions (Bicak *et al.*, 2021; Reith and Nehring, 2024). At the same time, the following question remains open: which specific tools can help students and aspiring chemistry teachers analyse experimental data with greater confidence and recognise appropriate bases for planning and interpreting experimental results?

In professional experimental chemistry, analytical method validation is a crucial procedure for assessing measurement consistency and accuracy (Taverniers *et al.*, 2004; Cantwell, 2025). While validation does not require complex statistical calculations, it provides a clear framework for understanding

the trustworthiness of experimental data and ensuring proper interpretation of results. In education, validation offers a way to translate professional metrological language into design-oriented thinking, structuring students' reasoning during experiment planning and data analysis. This raises the question of whether specific validation characteristics can be adapted as pedagogical tools to foster understanding of experimental data's reliability (European Medicines Agency, 2023). These characteristics, in an educational context, may deepen understanding of the experiment's internal logic and the quality of the results.

In the scientific literature, however, comprehensive guidance is still lacking on how elements of validation – such as specificity, linearity, precision, and trueness – can be pedagogically adapted to support students in working with experimental data. The potential of these characteristics to foster confidence in result interpretation and understanding of the basis for trusting experimental outcomes remains underexplored. Incorporating validation characteristics into students' reasoning may offer a mechanism for meaningful data analysis, enabling reliance on method quality in experimental design. Therefore, this study examines how integrating validation elements into an educational workshop can cultivate more reflective engagement with experimentation among future chemistry teachers, increase their confidence in data analysis, and improve their understanding of the foundations of trust in experimental results.

Given the lack of guidance on pedagogical adaptations of analytical method validation, this study examines how integrating key validation characteristics (specificity, linearity, precision, and trueness) into an educational workshop influences future chemistry teachers' ability to work with experimental data:

- meaningfully analyse experimental data based on clear criteria for measurement quality;
- establish relationships between method characteristics and result quality;
- use validation characteristics to guide the design of experimental procedures.

In this study, students engaged in learning activities centred on designing and justifying a spectrophotometric method to determine the total flavonoid content in plant material using a reference standard. The goal was not to obtain a definitive quantitative value, but rather to systematically justify the trustworthiness of results. This was achieved through analysis of the analytical system's behaviour and meaningful engagement with experimental data, based on key validation characteristics: specificity, linearity, repeatability, and trueness.

Theoretical foundations for the pedagogical adaptation of validation characteristics

To substantiate the pedagogical applicability of validation characteristics and to demonstrate their utility as tools for data



analysis, it is pertinent to examine the theoretical underpinnings that connect inquiry-based experimentation with the role of validation in assessing the trustworthiness of results (Abrahams and Millar, 2008; Reith and Nehring, 2024; Sanchez, 2024). The outcomes of chemical measurements are contingent upon the properties of the analyte, the methodology employed, and the conditions under which the procedure is executed. Consequently, variance among obtained values is an inherent attribute of experimentation (European Medicines Agency, 2023; ISO, 2023: 5725-1; Cantwell, 2025) and serves as an indicator of a method's capacity to yield a justified result. From a pedagogical perspective, this implies that a result should be considered not merely as a numerical value but also as a foundation for critical reasoning: students should comprehend the potential sources of data variability and the determinants of a result's justification (Stephenson *et al.*, 2020; Moju *et al.*, 2025). In the context of chemical experimentation, this principle is reflected in the understanding that a result constitutes an estimate whose trustworthiness is contingent upon the characteristics of the method and the conditions of its application (European Medicines Agency, 2023; Hibbert, 2023; Cantwell, 2025). Thus, pedagogical and chemical reasoning converge on the notion that confidence in a result is cultivated through an understanding of the method characteristics that influence the obtained data and determine its trustworthiness (Hamer *et al.*, 2024). Validation characteristics – specificity, linearity, precision, and trueness – delineate the aspects by which the reliability of a method and the level of confidence in a result can be evaluated. Each characteristic elucidates the measurement's reflection of the analyte, the relationship between data and variations in analyte quantity, the consistency of repeated results, and the correspondence to the true quantity being measured. Collectively, these characteristics constitute a framework for data analysis and can facilitate students' comprehension of method behaviour during the design of an experimental procedure (Araujo, 2009; Bicak *et al.*, 2021). In this section, these characteristics are considered as didactic elements capable of structuring students' reasoning during the analysis of experimental data and supporting the link between experiment design and the evaluation of result quality (Araujo, 2009; Cantwell, 2025).

Specificity as a mechanism for defining the object of measurement

Specificity refers to the ability to unequivocally assess the analyte in the presence of other components, such as impurities, degradants, and matrix components (European Medicines Agency, 2023). In other words, a method's result must reflect what it is actually intended to measure. In a chemical experiment, specificity determines which properties of the sample influence the result and under what conditions the measurement remains valid. From a pedagogical perspective, specificity is important because it establishes the initial analytical question: what exactly does the method measure? This understanding helps

students interpret the nature of the result, explain discrepancies in data, and account for procedural limitations during experimental design. Thus, specificity defines the starting point of analytical reasoning and precedes further consideration of method behaviour, allowing for subsequent analysis of characteristics such as linearity – a mechanism for describing quantitative changes in the signal.

Linearity as a mechanism for establishing a quantitative relationship

The linearity of an analytical procedure is its ability to obtain test results that are directly proportional to the concentration of analyte in the sample, within a specified range (European Medicines Agency, 2023). This indicates that the measurement result changes predictably with variations in the amount of substance, allowing for quantitative interpretation. In chemical experimentation, linearity reflects how accurately a method describes the relationship between concentration and response, enabling the construction and use of a calibration curve. For educational purposes, understanding linearity is crucial because it helps students view data as an expression of a systematic relationship, rather than as a collection of isolated numbers. It reinforces the concept that measurement is based on a model, and that the result reflects the method's behaviour. Linearity establishes the foundation for quantitative analysis, while precision is used to assess the stability of the data.

Precision as a mechanism for assessing measurement consistency

The precision of an analytical procedure expresses the closeness of agreement, defined as the degree of scatter, among a series of measurements obtained by repeatedly sampling the same homogeneous sample under prescribed conditions (European Medicines Agency, 2023). Precision indicates the extent to which identical experimental conditions yield similar values, reflecting the method's stability under repeatability conditions. In chemical experimentation, precision indicates method robustness and helps distinguish random data fluctuations from systematic errors. In an educational context, understanding precision reframes discrepancies between measurements as predictable method behaviour, helping students understand random deviations and reducing the tendency to attribute differences to personal error. However, consistent measurements do not guarantee correspondence to the true value, which requires evaluating trueness.

Trueness as a mechanism for assessing result trustworthiness

Trueness, as defined in ISO, 2023: 5725-1, is the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value (ISO, 2023:



5725-1). In essence, trueness indicates whether a method is fundamentally sound. Assessing trueness involves using a sample with a known composition. A significant difference between the method's result and the known value suggests an error in the procedure or execution, even if all measurements are consistent. In chemical analysis, trueness helps identify instances where a method produces "consistent" but incorrect results due to equipment, methodology, or experimental conditions. For future educators, trueness is a critical concept, demonstrating that careful and consistent experimental execution does not, by itself, guarantee accurate results. This understanding emphasises the need to verify the experimental design itself and to interpret data critically rather than accept it mechanically. Trueness complements specificity, linearity, and precision, together forming an integrated foundation for data analysis and meaningful experimental design.

A model for data analysis

Specificity, linearity, precision, and trueness collectively form an integrated data analysis model where each characteristic has a distinct function and reinforces other aspects of experimental quality assessment. Their pedagogical value lies in shifting result interpretation from isolated numerical values to structured reasoning, incorporating an understanding of method construction and experimental design logic. Table 1 presents

this system, describing each characteristic in both scientific and pedagogical terms. Note that the table does not describe empirically identified student difficulties, nor does it claim a diagnostic function. Instead, its purpose is to delineate aspects of design-oriented thinking that can be strengthened by including analytical method validation characteristics in a learning activity. The described "gaps" reflect elements of experimental understanding that can be supported through engagement with the corresponding validation characteristics, rather than actual difficulties of specific students. The pedagogical adaptation of each characteristic demonstrates how it was transformed into a supporting element for students' reasoning during experimental data analysis. Thus, the table serves as a conceptual scheme that defines the logic for integrating validation characteristics into the structure of the workshop.

Methodology

This qualitative study explored how incorporating validation characteristics into an instructional experiment fosters future chemistry teachers' understanding of experimental data quality and experimental design logic. Using thematic analysis (Braun and Clarke, 2021), the study traced changes in students' explanations and data interpretation approaches over time. The choice of a qualitative method was motivated by the desire to explore the nature of the transformation in students' reasoning

Table 1 Conceptual framework for the adaptation of validation characteristics

Characteristic	Scientific meaning	Gap	Pedagogical adaptation	Expected design skill
Specificity	Measurement refers precisely to the substance the method is intended to determine and remains valid under defined sample properties and procedural conditions	Lack of awareness that result validity is determined not only by procedural accuracy, but also by sample properties and measurement conditions	Use of the question "what exactly does the method measure?" as a means of reflecting on the influence of sample properties and measurement conditions on the result	Ability to justify the selection of experimental conditions and to take into account sample properties that determine result validity
Linearity	Reflects the ability of a method to reproduce a quantitative relationship between analyte content and the measured signal within a defined range	Lack of awareness that quantitative conclusions are possible only when a stable relationship exists within the working range, rather than from a set of unrelated values	Use of the question "which range ensures a stable relationship of the method?" as a tool for analysing method behaviour when key parameters change and for understanding when a quantitative conclusion becomes justified	Ability to justify the selection of the quantitative range and to use it to determine unknown values
Precision	Degree of agreement among results obtained under defined conditions	Underestimation of the need to plan replicate measurements and to understand that data spread reflects condition stability rather than "operator error"	Creation of a situation involving analysis of a series of replicate measurements, enabling students to recognise the pattern of dispersion as an indicator of condition stability and the quality of design decisions	Ability to design the number of replicates and ensure condition stability in order to obtain robust data that support justified conclusions
Trueness	Degree of agreement between the mean result and the accepted reference value; reflects whether the method yields a correct result overall, not merely consistent measurements	Lack of awareness that stable data may be systematically biased and fail to reflect the true value even when measurements are carefully performed	Use of a sample with known composition as a tool for evaluating stable data: students learn to distinguish results that correspond to the true value from systematically biased data and, on this basis, to adjust experimental design	Ability to verify a method using a sample with known composition, identify whether stable data are true or systematically biased, and adjust the experiment based on this analysis



that occurs as validation procedures are integrated into the structure of an instructional experiment.

The study included five third-year students from the Chemistry programme enrolled in the Institute of Natural Sciences, Korkyt Ata Kyzylorda University. It focused on a small group of five students, aligning with the project's aim to conduct a detailed examination of how understanding of the logic of analytical validation and the reliability of experimental data develops during validation procedures. The small sample size reflects the pilot nature of the study, which is aimed at an in-depth analysis of reasoning development rather than producing statistically generalisable results. Such small groups are commonly used in educational research during the initial stages of designing and testing instructional interventions, where the main goal is to identify features of learners' cognitive and conceptual development (Lawrie, 2021; Taber, 2024). The group formed a stable learning cohort participating in joint training within the pedagogical track "Chemistry Teacher Education". Participants had prior experience in performing basic spectrophotometric determinations from an analytical chemistry course, but lacked experience in validating analytical methods, allowing observation of the formation of understanding about data trustworthiness and the logic of experimental design from the ground up. Participation in the workshop was voluntary, not linked to any formal course, and conducted as extracurricular research activity, thereby excluding any influence on academic assessment and enabling students to freely express their reasoning.

The leaves and flowers of *Malva sylvestris* L., which contain flavonoids, were selected for this study (Irfan *et al.*, 2021). This choice was based on three factors: (1) the plant material is readily available regionally and simple to prepare in a teaching laboratory; (2) spectrophotometric methods described in the literature (Harborne, 1998; Fabjan *et al.*, 2003; Kreft *et al.*, 2006; Ramos *et al.*, 2017; Smyslova *et al.*, 2019; Alara *et al.*, 2020) allow for the reproduction of key validation steps under educational conditions; and (3) working with a real natural sample offers valuable didactic opportunities to discuss factors related to obtaining reliable analytical results with students. The quantitative determination of flavonoids was performed using a spectrophotometric approach based on general methodological recommendations (Harborne, 1998). As this monograph does not provide a specific experimental model for our raw material, students independently designed the experiment, guided by the research logic of contemporary studies. This included selecting parameters for raw material preparation and the concentration

of the alcoholic extractant (Alara *et al.*, 2020), the extraction temperature regime and the reference sample (Ramos *et al.*, 2017), and the conditions for alcoholic extraction and the measurement procedure (Smyslova *et al.*, 2019).

The study was conducted as an eight-day educational workshop (Table 2), with each day representing a practice-oriented research session. The structure of the workshop followed a sequence of stages that reflected the logic of experiment design and the acquisition of key validation characteristics:

Following the completion of experimental measurements at each of the four stages, students were asked to provide written explanations of how the conducted experiments influenced their interpretation of analytical results. Written reflection is widely used in science education research to examine the interpretation of experimental data and the justification of observed differences (Xu and Talanquer, 2012). Reflective responses were collected using open-ended questionnaires administered *via* Google Forms immediately after the experimental sessions, focusing on key validation characteristics: specificity (Day 4), linearity (Day 5), repeatability (Day 6), and trueness (Day 7). The questions aimed to elicit explanations of experimental decisions, interpretations of analytical signals, and assessments of result reliability. Additionally, a final written reflection was gathered on Day 8. All five participants completed the questionnaires at each stage, enabling the creation of a complete set of responses and the tracing of their reasoning development throughout the sequence of experiments. The written format was chosen because it captures individual reasoning, minimises group influence, and produces a textual corpus suitable for qualitative analysis (Martins and Macagno, 2022).

Data analysis was conducted using reflexive thematic analysis, which ensured transparency and avoided reliance on pre-determined categories. The unit of analysis was a segment of text where a student explained an experimental decision, interpreted data, or justified the reliability of the results. To maintain anonymity, participants were assigned coded labels (S1–S5), which were used throughout the analysis and in reporting excerpts. During coding, each segment was given an initial descriptive code reflecting the type of reasoning. Through iterative comparison, recurring features were identified and organised into broader reasoning patterns associated with specific validation stages. At the familiarisation stage, responses were read repeatedly and analysed manually to develop an overall understanding of the dataset, followed by the identification of fragments containing explanations of

Table 2 Structure of the eight-day educational workshop

Day	Session focus	Experimental activity	Validation characteristic
1	Conceptual introduction	Discussion of experimental data quality and the validation procedure	
2	Experiment design	Selection of the method and experimental variables	
3	Sample preparation	Preparation of extracts and standard solutions	
4	Spectral analysis	Selection of λ_{max} and measurement parameters	Specificity
5	Calibration	Construction of the calibration relationship	Linearity
6	Quantitative determination	Determination of flavonoid content	Repeatability
7	Trueness assessment	Application of the standard addition method	Trueness
8	Reflection	Interpretation of results and pedagogical discussion	



experimental decisions, data interpretation, and reasoning about result reliability. This was followed by inductive coding; meaningful fragments were labelled with descriptive codes, which were refined and consolidated through iterative analytical work. In the next stage, the codes were grouped into broader thematic categories reflecting features of students' reasoning and the dynamics of their approaches to experimental design. The thematic analysis focused not on quantitative comparison but on identifying qualitative changes, specifically, the strengthening of students' ability to justify the selection of measurement conditions and understand how these conditions influence the attainment of reliable results. To ensure analytical transparency, a portion of the dataset was independently coded by two researchers, and any differences in coding were discussed until consensus was reached.

The local ethics committee approved the study. All participants provided written informed consent, were guaranteed data confidentiality, and were assured that their participation would not affect their academic assessment. All experimental work was conducted in accordance with chemical safety requirements, utilising personal protective equipment and a fume hood. Standard safety precautions were observed when working with heating devices and ethanol.

Results and discussion

This section presents the results of an instructional workshop where students performed a spectrophotometric determination

of the total flavonoid content in a plant sample and engaged with elements of experimental design and data analysis through the validation features of an analytical method. The focus is on how experimental decisions at different stages influence data interpretation and the justification of conclusions, rather than on the phytochemical characterisation of the sample. The analytical experiment and its outcomes are provided in the supplementary materials. The plant sample was treated as a didactically meaningful object with a complex matrix, allowing discussion of method specificity, the choice of analytical wavelength, extraction conditions, and sources of measurement variability. Students' written reflections, collected after each stage and in a final integrative reflection, were analysed to identify how they explained observed differences and justified their analytical choices. The analysis revealed recurring reasoning patterns associated with stages of analytical validation, which formed the basis for the qualitative analysis presented in the following sections. Their relationship to the validation stages is schematically illustrated in Fig. 1.

Specificity

The assessment of specificity marked the initial stage in the validation cycle of the student project and established the foundation for reevaluating the connection between analytical measurement and experimental conditions. In analytical chemistry, specificity is defined as a method's ability to selectively identify the target analyte amid matrix components, thereby

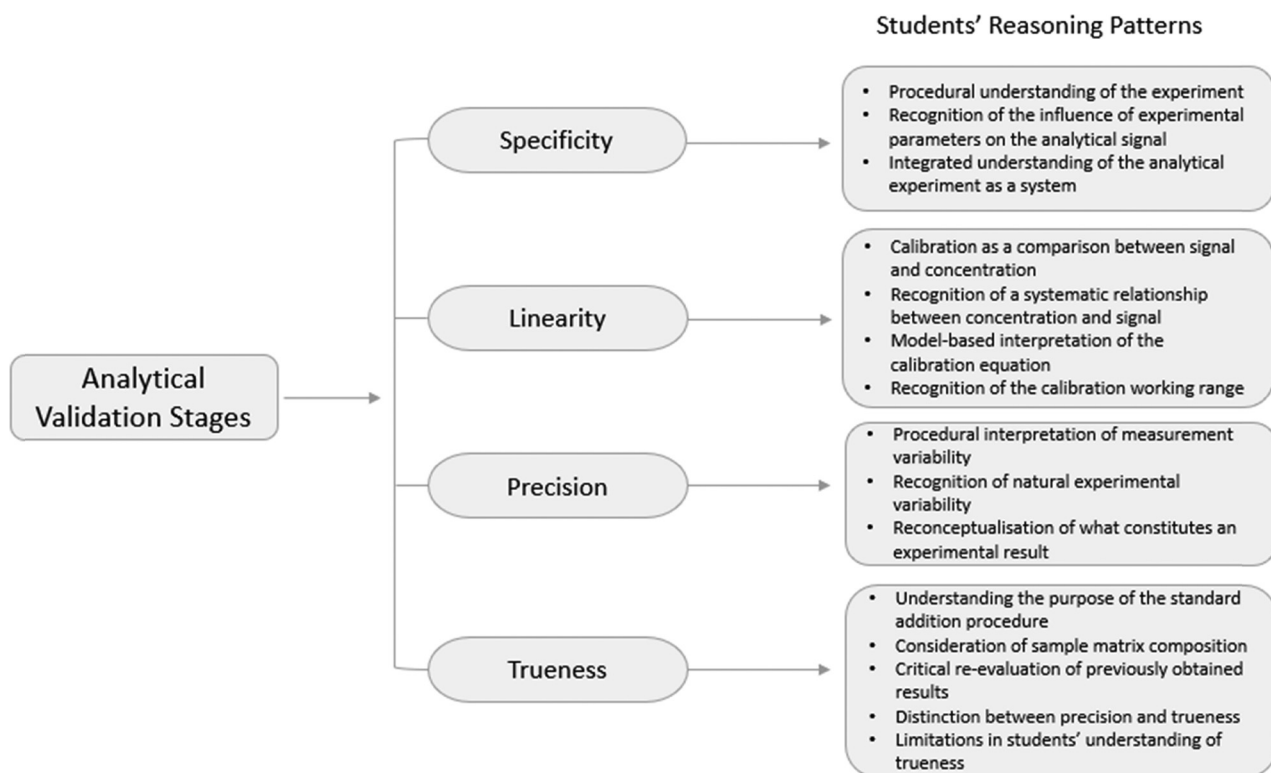


Fig. 1 Analytical validation stages and the patterns of students' reasoning identified in the study.



enabling the signal to be attributed to the compound of interest (Skoog *et al.*, 2017; European Medicines Agency, 2023). Within the pedagogical framework of the course, this phase focused on a reflective question: what exactly does the analytical method measure, and which experimental conditions ensure the reliability of the analytical signal?

The analytical work began with measuring the UV-Vis spectrum of the rutin–AlCl₃ complex, which served as a reference for flavonoid analysis. Full spectral data for the standard solution and plant extracts prepared under different extraction conditions are included in the SI (Fig. S1 and S2). The spectrum of the rutin complex showed an absorption maximum at 415 nm, which was selected as the analytical wavelength. Extracts of *Malva sylvestris* L., obtained with ethanol at different concentrations (40%, 70%, and 95%), exhibited varying absorbance values (0.163, 0.136, and 0.053), with the differences following a clear pattern. This required explanation and indicated that the analytical signal depends not only on the presence of the analyte but also on sample preparation methods (Harris, 2016; Skoog *et al.*, 2017). Comparison of the extract spectra with that of the rutin complex revealed overlapping absorption bands in the 410–415 nm range, with no significant additional peaks. This allowed the signal at 415 nm to be interpreted as related to flavonoid compounds, and the spectral correspondence was regarded as an indicator of the method's analytical specificity.

After the measurements, students were asked to provide written explanations of how extraction conditions affected the interpretation of results. Analysis of their responses showed three distinct reasoning patterns, differing in how experimental variations were understood and linked to the experimental conditions.

The first pattern depicts a procedural understanding of the experiment, where laboratory work is seen as a series of operations resulting in a measurable signal. Such interpretations are well documented in chemistry education research, where experimentation is often reduced to following set instructions (Domin, 1999; Bretz, 2019). This type of reasoning is demonstrated in the response of student S1: “*After comparing the conditions obtained, my understanding of the experiment changed; I can now choose experimental conditions myself*”. This statement shows an initial conceptual shift: despite the dominance of procedural logic, the student starts to view experimental parameters as variables that can be deliberately chosen. Such understanding can be seen as an early step in the shift from a procedural to an inquiry-based perception of laboratory work (Amadeus and Lorenz, 2023).

The second pattern shows recognition of how experimental parameters affect the analytical signal. Four of five students (S2–S5) saw differences in absorbance values as resulting from extraction conditions rather than random fluctuations. In their explanations, they related these differences to ethanol concentration and its impact on flavonoid extraction efficiency. For instance, student S3 stated: “*The experiment showed that ethanol concentration influences the yield of rutin. . . At 40% ethanol, the extraction value was highest, whereas at 95% it was lowest*”. This explanation was based not on prior assumptions but on the

need to align absorbance values with extraction conditions. Similar interpretations appeared in other responses, where a higher analytical signal for the 40% ethanol extract was linked to greater solvent polarity and better solubility of flavonoid glycosides. In this reasoning pattern, the analytical signal is not just seen as an instrument reading but as a result of interactions between the chemical properties of the analyte and the experimental conditions. This type of reasoning indicates a move toward experimental thinking, where students start to analyse how experimental parameters influence measurement outcomes (Xu and Talanquer, 2012; Brederode *et al.*, 2020).

The reflections also showed that interpreting experimental data at this stage might involve some uncertainty. In one case, the interpretation of the results seemed partly influenced by expectations based on previously studied scientific literature. Student S4 linked the expected maximum signal to 70% ethanol, but the experimental data showed the highest absorbance for the 40% solution. This assumption related to the idea that, in the literature reviewed, optimal extraction conditions often occur at intermediate solvent concentrations, and the student applied this pattern to the current experiment. Such interpretation through prior knowledge becomes especially clear when results do not match expectations. In this case, the gap between what was expected and what was observed required either a revision of the initial assumption or a reinterpretation of the data (Brewer and Lambert, 2001; Talanquer, 2018).

Another observation relates to how students perceived their role in choosing experimental parameters. Despite the collaborative discussion of extraction conditions during the planning stage, student S2 distanced themselves from the analytical justification of these choices, noting that the parameters were not selected personally and therefore could not be fully evaluated. This stance suggests an incomplete ownership of responsibility for methodological decisions. In laboratory education research, such cases are described as diffused ownership of experimental design, where responsibility for parameter selection is seen as belonging to the instructional context or the instructor rather than the students. This attitude is often linked to prior experience in so-called “cookbook laboratories”, where experimental conditions are predetermined and do not require independent design (Seery *et al.*, 2019; Areljung *et al.*, 2021).

The third reasoning pattern demonstrates an integrated understanding of the analytical experiment as a system, where sample preparation, extraction conditions, and instrumental measurement are viewed as interconnected elements. This type of reasoning is best exemplified by student S5, who remarked: “*This stage showed me that the measurement result depends not only on the substance itself, but also on how the sample is prepared. In other words, extraction conditions effectively become part of the analytical model of the experiment and can significantly influence the final analytical signal*”.

In this case, spectrophotometric measurement is understood not as a direct indicator of substance content, but as the result of a series of experimental choices that influence the composition of the extract and the development of the



analytical signal. This interpretation reflects a systemic view of the analytical experiment, where laboratory procedures, measurement methods, and analytical models are seen as parts of a unified experimental system (Petritis *et al.*, 2022). Overall, the patterns identified suggest that discussing specificity became a crucial step in students' rethinking of the structure of the analytical experiment. At first, the laboratory procedure was mainly seen as a measurement task; however, the need to explain differences in spectral features and absorbance values under different extraction conditions led students to see the analytical signal as the outcome of specific experimental decisions related to sample preparation and parameter choices. Therefore, the characteristic of specificity served not only a methodological purpose, confirming the connection between the analytical signal and the target analyte, but also an educational purpose, guiding students' attention to the link between experimental design and analytical measurement.

It is important to note that instances of uncertainty and partial awareness of experimental decisions were observed mainly at the early stage of the project, when students were just beginning to understand the logic of experimental design and method validation. As the project advanced, such signs became less noticeable: in later reflections, students more often viewed experimental parameters as variables that needed intentional selection and justification rather than as fixed elements of a procedure. Therefore, initial uncertainty can be seen not as a sign of a lack of understanding, but as a typical feature of the early phase of inquiry-based laboratory learning, where the clash between expectations and experimental results helps foster scientific thinking (Seery *et al.*, 2019; Wellhöfer and Lühken, 2021).

Linearity

Following the discussion of specificity, the next step in validation was linearity, which in analytical chemistry describes a method's ability to produce a signal proportional to the analyte concentration within a specified working range (Raposo and Barceló, 2021; European Medicines Agency, 2023). While specificity confirms that the signal corresponds to the target analyte, linearity assesses whether this signal can be interpreted quantitatively. Linearity was evaluated by establishing a calibration relationship between standard solution concentrations and the resulting analytical signal, forming a mathematical model that converts instrument response into analyte concentration (Analytical Methods Committee, 1994). However, in the context of the project, constructing the calibration curve was not viewed as a mere formal step. The need to interpret this relationship prompted questions about how numerical signal values gain quantitative significance. Therefore, the calibration stage was regarded not only as an analytical procedure but also as an educational opportunity to explore how measured signals relate to the amount of substance.

The experimental task involved creating a calibration curve using rutin as a reference flavonoid. The complete set of calibration data, including absorbance readings, the calibration

curve, and linear regression details, is provided in the SI (Fig. S3). Standard solutions of rutin reacted with AlCl_3 to form a complex with an absorption maximum at 415 nm, after which absorbance was measured and the calibration relationship was established. Importantly, the absorbance values themselves did not have immediate quantitative significance for students; interpretation became possible only after determining the link between signal and concentration, which required an understanding of the calibration model. The regression equation was then used to calculate the flavonoid content in the plant extracts.

The analysis of students' written reflections demonstrated that their interpretation of the calibration relationship was influenced by their understanding of the relationship between signal values and the concentrations of standard solutions. Three related types of reasoning can be identified in the responses, each representing a different way of understanding this relationship.

The first type of explanation can be viewed as a procedural interpretation of calibration, where standard solutions act as reference points to compare the signal from an unknown sample with signals from solutions of known concentrations. This understanding was evident in the responses of students S1, S2, and S4. For example, S1 stated: "A standard sample is needed to understand what the signal looks like for a known amount of substance. First, the standards are measured to see how the signal varies with flavonoid concentration. This creates a 'comparison scale'. Then, the unknown sample is measured, and its signal is compared with this scale". Likewise, S4 described the role of the standard solution as a way to translate the instrument signal into a quantitative concentration value: "We need a standard model to translate the language of the instrument into the context of concentration". Such explanations demonstrate an understanding of the purpose of standard solutions; however, they also imply that the calibration relationship is mainly viewed as a tool for signal comparison. The connection between signal values and concentrations is not always understood as a functional relationship but rather as a set of discrete reference points. These procedural interpretations are well documented in chemistry education research (Reid and Shah, 2007; Raposo and Barceló, 2021).

Alongside this, students' reflections also revealed elements of a more advanced understanding of calibration as a functional relationship between analyte concentration and the analytical signal. This interpretation can be seen in the responses of students S1, S2, S4, and S5, who noted that changes in concentration lead to systematic changes in the signal. For example, S1 wrote: "The quantitative determination of flavonoids became possible because the optical density of the solution is directly related to the concentration of the substance". Similarly, S4 linked the increase in signal to the absorption of light by the molecules: "The more rutin there is, the more light is absorbed. As a result, the stronger the signal". Such explanations reflect an understanding of the systematic variation of the signal with concentration and emerged as students attempted to reconcile the observed relationship between absorbance and



the concentrations of standard solutions. The shift from comparing signals to interpreting the relationship between variables is considered an important stage in the development of quantitative reasoning in chemistry (Brandriet *et al.*, 2018).

The most advanced understanding of the calibration relationship was demonstrated in student S5's response, who described the regression equation as a model linking the instrument signal to the amount of substance. The student noted: "When we measure only one sample, we simply obtain an instrument signal value, but by itself it does not directly indicate the amount of substance. To relate the signal to concentration, it is necessary to first observe how the signal changes for known concentrations of a standard substance". He further added: "This equation is perceived not merely as a formula for calculation, but as a model that links the physically measured instrument signal to the amount of substance". This understanding evolved as the student moved from considering individual data points to recognising them as part of a unified relationship, reflecting a view of the calibration equation as a model of the analytical system. Research in chemistry education indicates that interpreting graphs and equations as models of physical or chemical relationships is a key stage in the development of scientific thinking (Sevian and Talanquer, 2014).

Simultaneously, the analysis of written responses uncovered several conceptual issues concerning the conditions under which the calibration relationship applies, most clearly illustrated in the reflections of student S3. For example, the student questioned: "It is unclear why a particular concentration is chosen. Does it matter which one we take if we are going to determine it anyway?" This question often arises when the choice of concentration range cannot be justified without considering the form of the relationship and highlights the need to match experimental conditions with the characteristics of the calibration model. It indicates some conceptual uncertainty in selecting an appropriate concentration range for developing the calibration curve. In another case, the student remarked: "It is unclear why the graph sometimes bends as the concentration increases". Observations of deviations from linearity challenge the universality of the relationship and call for an explanation of factors that disturb the proportionality between signal and concentration. These questions point to real limitations of analytical methods, as non-linearity can result from chemical interactions, optical effects, or instrumental factors, especially at higher concentrations where the Beer-Lambert law is no longer strictly valid.

Furthermore, the student showed an emerging understanding of measurement variability, noting that "if a blank sample is measured several times, it may give different values", which indicates awareness of random error and the need for statistical analysis of data. Likewise, several students (S1, S2, S4, and S5) emphasised that the validity of quantitative calculations relies on maintaining consistent experimental conditions and operating within the concentration range used to build the calibration curve. For instance, S5 noted that calculations are only reliable when the signal from the unknown sample lies within the calibration range. Overall, students' reflections suggest a

shift from mainly procedural views of calibration towards a more conceptual understanding of the relationship between analytical signal and analyte concentration. The process of constructing the calibration curve encouraged students to see the instrumental signal not just as a numerical output but as part of a quantitative relationship. In this context, the calibration equation is increasingly understood not merely as a calculation tool but also as a model representing the analytical connection between signal and concentration. This transition from using equations procedurally to interpreting them as models of chemical measurement is recognised as a vital step in developing quantitative reasoning in chemistry (Sevian and Talanquer, 2014).

Although establishing a linear relationship between signal and concentration allows for quantitative interpretation, a calibration curve alone does not guarantee the reliability of results. Even with good linearity, results may still be unreliable if measurement variability is high. In analytical chemistry, the validity of results depends not only on calibration but also on measurement reproducibility. Therefore, the next validation step focused on assessing precision, aimed at analysing the repeatability of the analytical signal.

Precision: assessment of measurement repeatability

Precision was assessed under repeatability conditions by measuring the same flavonoid-AlCl₃ complex solution multiple times. Absorbance was measured at 415 nm using UV-Vis spectrophotometry. The complete dataset, including signal values and statistical parameters, is provided in the SI (Table S12 and Fig. S4). Under stable conditions, the identical solution was measured repeatedly, allowing students to observe that, even with the same method, instrument, and sample, the values can fluctuate within a certain range. In analytical chemistry, such consistency is described as precision under repeatability conditions, which indicates the stability of the signal across multiple short-term measurements (Konieczka, 2007; Miller and Miller, 2018). The repeated measurements showed small but noticeable fluctuations in absorbance values, which could not be ignored and prompted students to consider how to interpret differences between measurements of the same sample. Students were then asked to reflect on the causes of these differences, the importance of repeated measurements, and the implications of the results.

Analysis of the responses from the five participants identified several recurring modes of reasoning, differing in how students conceptualised the observed variability between measurements.

The first type of reasoning involves a procedural view of measurement variability. Three participants (S1, S3, and S4) primarily explained differences between repeated measurements in terms of operational aspects of the procedure, attributing variations to technical factors such as inaccuracies in solution preparation, cuvette contamination, pipetting errors,



or instability in cuvette positioning within the spectrophotometer. For example, S1 explained: “Differences between repeated measurements of the same sample are understood by me as operator error... for example: incorrect pouring, improper cuvette placement... or background instrument noise or instability of the light source”. Similarly, S3 pointed to inaccuracies in solution preparation, AlCl_3 dosing, and cuvette contamination. Such explanations indicate that the observed variability was not yet conceptualised as an inherent feature of the measurement process; instead, differences were perceived as deviations from an expected norm. This reflects a common tendency whereby students initially attribute variability to errors rather than to the probabilistic nature of measurements (Talanquer, 2018; Ferreira *et al.*, 2025).

However, the reflections also revealed a second model, linked to recognising natural experimental variability. Four participants (S2–S5) noted that their interpretation of differences changed during the experiment. Initially, deviations were often seen as errors; however, as multiple measurements were taken and similar yet non-identical results were repeatedly observed, these differences were understood as a normal part of experimental work. For example, S5 stated: “When we obtained the first results, the differences between values initially raised doubts... it seemed that if the sample is the same, the values should be almost identical”. He later added: “As we performed a series of measurements, it became clear that small differences between results are a normal part of experimental work”. In this case, the rethinking emerged not from abstract discussions of precision, but from repeated encounters with non-identical results under unchanged conditions. A similar shift was seen in S2, who initially interpreted differences as errors and later as normal variation. This transition indicates the development of an understanding of variability in experimental data and the need to evaluate reliability based on a series of observations rather than a single measurement. Such findings align with research indicating that engagement with repeated measurements supports understanding of uncertainty and the probabilistic nature of data (Szalay and Tóth, 2016).

The third approach involves a reconceptualisation of what constitutes an experimental result. Four participants (S2–S5) noted that repeated measurements altered their understanding of the result: rather than representing a single instrument reading, it came to be viewed as a characteristic of the entire dataset, since individual measurements did not match and could not serve as a sufficient basis for conclusions. For example, S5 explained: “Now the result is perceived more as a characteristic of the whole series of measurements... usually as an average value that better reflects the actual signal”. Similarly, S4 noted: “Now the result is considered as a characteristic of the entire dataset—a mean value taking variability into account”. This shift towards interpreting results as statistical characteristics reflects the development of a quantitative approach to experimental measurement and aligns with research showing that analysing series of measurements supports understanding of the roles of averaging and variability (Davidowitz *et al.*, 2001). One participant (S1) demonstrated a more advanced attempt to

distinguish between accuracy and precision: “Accuracy is how close the measurements are to the true value, while precision is how close the measurements are to each other”. The student also emphasised the importance of confidence intervals and statistical interpretation, noting that results should be expressed not as single values but as estimates that include uncertainty.

Despite this shift in how repeated measurements are interpreted, the reflections reveal several conceptual limitations in students’ understanding of data variability. First, although students increasingly recognised variability as inevitable, they still primarily interpreted its origin in procedural terms, such as operator actions, technical inaccuracies, or instrument instability. While these factors do influence results, such explanations do not fully capture the probabilistic nature of measurement, where random fluctuations occur even under controlled conditions. Second, in several reflections, an expectation of perfect reproducibility persisted, especially at earlier stages. As S1 noted, deviations were seen as “a sign that something went wrong”. These expectations are well documented in laboratory education research, where it is often assumed that correct procedures should produce identical results (Talanquer, 2006). This shows the ongoing belief that a “correct” experiment should yield identical values, even when empirical evidence suggests otherwise. Third, although students began to recognise the importance of a statistical approach, their explanations remained largely qualitative. For example, they emphasised the importance of averaging, but did not explicitly consider quantitative parameters such as standard deviation or relative standard deviation, which are commonly used to assess precision (Miller and Miller, 2018).

Another notable aspect relates to students’ reflections on the role of data dispersion in evaluating reliability. Four participants (S2–S5) emphasised that the proximity of repeated measurements is more significant than their exact sameness. For example, S2 observed that perfectly identical values can raise doubts, as small variations are expected in real experimental conditions. This suggests that differences between measurements are interpreted not as faults but as a basis for assessing the reliability of results. In scientific practice, such variations are seen as inherent properties of measurement processes, requiring statistical analysis rather than procedural elimination (Cantwell, 2025). Overall, the analysis reveals that repeated measurements played a crucial role in transforming students’ understanding of experimental results. They created conditions where variability could be observed and interpreted as a phenomenon needing explanation. The repeated mismatch of results under unchanged conditions provided an empirical basis for revising the idea of a single “correct” value. Consequently, four participants (S2–S5) shifted from viewing results as individual instrument readings to understanding them as statistical properties of a measurement series, encompassing both central tendency and dispersion.

From the perspective of the conceptual model (Table 1), these observations illustrate a pedagogical adaptation linked to the characteristic of precision: analysing repeated measurements allowed students to recognise variation as an indicator



of stability rather than as operator error. At the same time, the findings suggest that recognising variability does not equate to a full understanding of precision. Awareness of measurement consistency prompts a further question: whether the result corresponds to the true value, thereby introducing the next validation characteristic: trueness.

Trueness: assessment by the standard addition method

At the final stage of the sequence, students' attention shifted from measurement repeatability to trueness, as the previous stage had shown that consistency of results alone does not guarantee their validity. In analytical chemistry, trueness is usually evaluated through recovery experiments, where a known amount of analyte is added to a sample and the resulting change in signal is analysed. In this study, trueness was assessed using the standard addition method, a widely used procedure to verify result accuracy in complex matrices (Cantwell, 2025; Skoog *et al.*, 2017). During the experiment, a known amount of rutin standard was added to the plant extract before spectrophotometric measurement of the flavonoid–AlCl₃ complex at 415 nm. Detailed results, including absorbance values and recovery calculations, are provided in the SI (Table S13). The addition of the standard caused an increase in absorbance proportional to the amount of analyte added, and the calculated recovery values were close to the expected range.

However, these values required interpretation: agreement with expected values was not obvious and raised questions about how well the result reflects the actual analyte content in the sample. This relationship illustrates a fundamental principle of spectrophotometry, the proportionality between absorbance and concentration as described by the Beer–Lambert law (Harris, 2016; Skoog *et al.*, 2017). From an analytical perspective, the recovery values obtained suggest an adequate estimation of flavonoid content and a low risk of significant systematic error. From a pedagogical perspective, this stage was especially important because it encouraged students to reconsider the assumption that stable measurements automatically indicate correct results, demonstrating that consistency does not guarantee the absence of systematic bias.

The analysis of written reflections uncovered several distinctive features in students' interpretation of this stage.

The first pattern involves recognising the distinction between measurement repeatability and accuracy to the true value. Four participants (S1, S2, S4, and S5) noted that consistent results do not always indicate correctness. For example, S2 wrote: “Stable results do not always mean correct results. They may also be the result of systematic error”. Similarly, S4 pointed out that “repeatability indicates the stability of measurements, but does not imply their correspondence to the true value”. These conclusions arose from comparing previously obtained stable values with the results of standard additions, which demonstrated that consistency alone does not guarantee correctness. Such statements reflect an important conceptual shift, as

students in laboratory learning often interpret consistent measurements as evidence of correctness without considering the possibility of systematic deviation (Gao and Lloyd, 2020; Patel, 2022). In this case, the standard addition method challenged this assumption by showing that even stable data can remain systematically biased.

The second type of interpretation relates to students' attempts to explain the sources of systematic error through the chemical characteristics of the system. Many responses explicitly mentioned the influence of solution composition and the presence of other substances. For example, S1 wrote: “If the solution contains other substances, they may affect the signal”. Similarly, S4 identified matrix effects, noting that “the matrix structure of the plant extract may influence the signal”. In the same vein, S5 attributed deviations to solution composition and impurities: “The problem may lie in the composition of the solution, since impurities affect the signal or the reaction”. These interpretations show that students started to consider the impact of sample composition on the analytical signal, recognising that it depends not only on analyte concentration but also on other components. In analytical chemistry, such matrix effects are regarded as a primary source of systematic error, and the standard addition method is specifically designed to compensate for them (Harris, 2016; Skoog *et al.*, 2017). Therefore, explanations of deviations emerged as efforts to interpret signal behaviour within a complex matrix.

The third interpretation relates to understanding the standard addition procedure as a way to validate the analytical method. Some students began to see the experiment not merely as another measurement step, but as an opportunity to confirm that the analytical signal genuinely correlates to the target compound. For example, S1 noted: “If the signal increases in accordance with the added analyte, this confirms that the measured signal is truly associated with that substance”. Likewise, S3 wrote that “comparing the expected and observed increase in signal helps to determine whether the result can be trusted”. A particularly illustrative case is the interpretation provided by S5, who described the purpose of the procedure as testing whether “the method works correctly” and whether “the instrument detects all of the rutin or whether part of it is masked by impurities”. This statement reflects a shift from simply obtaining a value to evaluating the conditions under which that value is valid. These explanations suggest that students began to view the experiment as a tool for verifying the reliability of the analytical method. The ability to interpret laboratory procedures as means for assessing analytical reliability is recognised as an important aspect of scientific thinking in chemistry education (Talanquer, 2011).

Simultaneously, students' reflections revealed several conceptual challenges in interpreting the results. This was most evident in the response of S5: “Is a result of 95% a good result or a large error? How should it be interpreted?” Such a question arises when the obtained value differs from the expected one and needs assessment. It shows an expectation of exact agreement between experimental and theoretical values and indicates that a numerical result alone is inadequate without



understanding the criteria for its interpretation. However, in analytical practice, recovery values are evaluated within acceptable limits of uncertainty rather than through strict correspondence with theoretical values (Miller and Miller, 2018; Cantwell, 2025). This suggests that, although students may grasp the procedure, the criteria for judging result acceptability remain unclear.

In several responses, the introduction of standard additions prompted a re-evaluation of previously obtained data, as results once considered reliable due to their stability now required further validation. For example, S3 observed that earlier measurements seemed dependable, but it later became evident that repeatability alone was inadequate to confirm accuracy. Similarly, S5 noted that comparing results with the standard led to a reassessment of initial confidence in measurement precision. In the final question, students were asked to explain what trust in analytical data should be based on. Their responses emphasised the importance of independent verification of the method. For instance, S1 stated that confidence in results should rely not only on measurement stability but also on validation through standards and additional experiments. S3 expressed a similar view: *“To trust a result, reproducibility alone is not sufficient; its accuracy must also be verified”*. S5 summarized this reasoning by noting that the reliability of a method depends on multiple factors, including correct instrument response, the relationship between signal and concentration, measurement reproducibility, and potential matrix effects.

Taken together, these reflections suggest that the standard addition experiment created a situation where even reproducible measurements could not automatically be regarded as valid without independent verification. Analytical reliability, therefore, was understood not as a trait from a single characteristic, but as the outcome of integrating multiple aspects of the data. In this sense, this stage shifted the discussion from simply “obtaining a value” to justifying trust in that value. At the same time, students’ responses indicate that their understanding of trueness remained partly incomplete, especially in interpreting recovery values within acceptable uncertainty limits. Thus, this stage not only expanded students’ ideas of validation but also laid the foundation for exploring a broader question: on what basis can an analytical result be considered trustworthy?

Final reflection: formation of trust in experimental results and the limits of their justification

The final stage of reflection focused on analysing how students interpreted the analytical result after completing the full validation cycle. Unlike the previous subsections, which examined individual method characteristics, this stage addressed a broader question: what constitutes a trustworthy result.

Analysis of the written reflections indicates that participation in validation procedures transformed students’ approaches to interpreting results. At the initial stage, some participants perceived the result as a numerical value obtained from the

instrument. As S1 noted: *“At the beginning, I perceived the result simply as a single number shown by the instrument... it seemed that once a measurement was performed and a value obtained, that was the result of the experiment”*. Such an understanding arises when individual measurements are not questioned and do not require interpretation, and it aligns with a well-documented tendency to treat results as direct observations rather than claims needing justification (Keiner and Graulich, 2020). After completing the full validation cycle, this interpretation changed. Comparison of results across different stages demonstrated that no single measurement could serve as a sufficient basis for a conclusion. The same student (S1) later described the result as an outcome of the analytical process: *“Now, for me, a reliable result is not just a single measurement, but one obtained after a series of measurements and various checks”*.

Similar shifts were observed among other participants. S2 stated: *“Now I understand that a reliable experimental result is one that has been verified and repeatedly confirmed”*. S5 also linked trust in the result to a combination of checks: *“A reliable experimental result is a value that can be trusted... it must be supported by a series of repeated measurements and a stable signal”*. These responses suggest that students began to see results not merely as instrument readings, but as claims requiring justification through a system of analytical validations. The importance of such validation became especially clear when results obtained *via* different approaches could not be interpreted without comparison. Therefore, participation in validation procedures helped develop justified trust in the experimental data, based on the method’s verifiability.

At the same time, the reflections uncovered persistent contradictions. Some participants still linked the reliability of results mainly to measurement stability or instrument performance. For example, S3 remarked: *“Repeated measurements and consistency of results are key steps in increasing confidence in the obtained data”*. Similarly, S4 highlighted the importance of correct procedural execution: *“The instruments must operate without errors, and operator error must be eliminated”*. These responses reflect a well-documented phenomenon in science education: new interpretations do not fully replace previous ones, but rather coexist with them, especially when learners are required to choose between multiple explanatory frameworks (Talanquer, 2011). These differences show that the transition towards a methodological understanding of results is not straightforward, but involves the coexistence of multiple models of explanation. The overall process of this transition is illustrated in Fig. 2.

The figure illustrates the shift from viewing results as instrument outputs to understanding them as method-supported claims. Initially, students considered results as mere instrument readings, trusting their validity based on measurement consistency. After engaging in validation processes, some began to see results as claims backed by the method, since data from different stages needed comparison and integration. This change is reflected in S1’s reflection: *“First, it is necessary to understand how the signal is related to the quantity of substance... then to verify that the measurements are repeatable... and after*



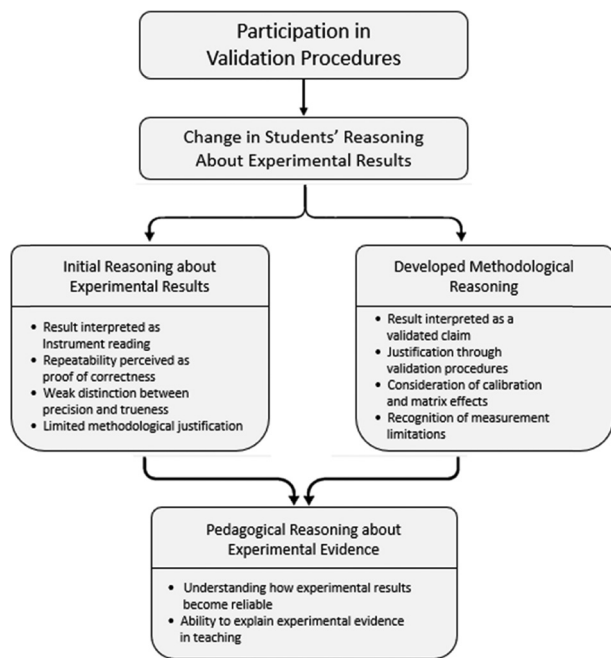


Fig. 2 Change in students' reasoning regarding experimental results after engaging in validation procedures.

that to ensure that the result is indeed correct". This sequence is more than just a set of steps; it embodies a logical process where validity is established by aligning different types of evidence. However, the shift remains incomplete. For example, S1 noted: "The stability of results does not yet mean that they are correct. . . additional verification is still required". Other participants expressed this with less certainty: S4 suggested that the correct procedure only offers partial reliability, estimating it at about "30–40%". These responses demonstrate that students recognise the importance of methodological checks, but their understanding has not yet formed a clear system for evaluating analytical evidence.

An important aspect of the reflections was their pedagogical interpretation. S3 related this experience to future teaching practice, noting that "in school laboratories, students often think that there must be a 'correct answer'. . . whereas the main purpose of an experiment is to understand why the result turned out the way it did". Overall, the findings demonstrate that participation in validation procedures facilitated a shift from viewing experimental data as simple numerical readings to understanding them as methodologically justified scientific claims. This transition was accompanied by the development of a more critical attitude towards data: trust in results came to be seen not as a consequence of measurement, but as the outcome of a system of analytical validations. At the same time, the persistence of procedural interpretations indicates that this shift remains incomplete, as criteria for the sufficiency of validation and interpretation of results have not yet formed a stable framework for evaluating analytical evidence. In this context, the validation experience functions not only as an analytical process but also as a model for explaining the nature of experimental results in

education. This is especially important for future chemistry teachers, who must be able to explain to students the necessity of critically evaluating and methodologically justifying results. Thus, participating in the full validation cycle enabled students not only to master analytical procedures but also to develop an understanding that trust in data is established through critical interpretation and comparison, rather than by obtaining a single "correct" value.

Conclusions

This study aimed to investigate an alternative pedagogical strategy for fostering students' experimental reasoning within chemistry laboratory education. Specifically, it proposes and evaluates the application of key analytical validation characteristics – specificity, linearity, precision, and trueness – as a methodological framework for instructing experiment design and data interpretation.

The findings suggest that validation characteristics possess a utility extending beyond the assessment of analytical method quality. Within an educational context, they furnish a robust logical framework for reasoning about the trustworthiness of results, conditions of applicability, and limitations of interpretation, thereby facilitating the development of well-grounded experimental judgment. This enables a reconceptualisation of experimentation not merely as a process of data acquisition, but as a platform for cultivating a professional disposition toward the quality and reliability of empirical inferences.

The principal contribution of this work resides in a methodological reconceptualisation of the role of validation in chemistry education. Validation logic is posited as a pedagogical instrument for explicating the epistemic structure of experimentation, rendering the processes of justification, identification of limitations, and the establishment of trust in data as explicit targets of purposeful instruction. In this sense, validation serves as an epistemic "scaffold", supporting the transition from procedural execution of laboratory tasks to meaningful experimental reasoning.

From a pedagogical standpoint, the proposed validation-oriented approach constitutes a reproducible and conceptually coherent model for structuring laboratory instruction that does not necessitate more sophisticated instrumentation or methodologies. Reframing professional analytical concepts as educational tools creates opportunities for systematic engagement with uncertainty, argumentation, and the contextual dependence of experimental conclusions as central facets of experimental practice. Consequently, validation transcends its status as an exclusively professional procedure, becoming an accessible and productive language for teaching experimental thinking.

It is crucial to emphasise that the proposed validation-oriented approach is not intended to impart formal metrological validation in its entirety. Within an educational context, not all validation characteristics can be implemented in a strictly analytical sense; their selection and depth of treatment are



contingent upon instructional objectives, available resources, and the nature of the experimental task. Thus, validation is conceived not as a universal checklist, but as a flexible logic of experimental reasoning, the pedagogical merit of which lies in supporting justification, interpretation, and context-dependent trust in data.

Future research may explore the applicability of this approach across diverse analytical contexts, at varying levels of chemistry education, and in the longitudinal development of experimental judgment.

Author contributions

S. Z. conceptualized the project and authored the manuscript. S. Z., L. Z., G. A., and B. I. conducted data analysis and discussed the interpretation of the results and implications. All authors reviewed, edited, and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this study are not publicly available. Ethics approval for this study did not permit public data sharing, and participants did not consent to their data being shared; releasing the data could compromise participant privacy.

Supplementary information (SI): experimental details, UV-Vis spectra, calibration data, repeatability measurements, recovery calculations, and supporting analytical validation data related to the spectrophotometric determination of total flavonoid content. See DOI: <https://doi.org/10.1039/d6rp00078a>.

Acknowledgements

The authors would like to thank Gulzhan Karibayeva for her valuable support during the preparation of the manuscript.

References

- Abrahams, I. and Millar, R. (2008), Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science, *Int. J. Sci. Educ.*, **30**(14), 1945–1969, DOI: [10.1080/09500690701749305](https://doi.org/10.1080/09500690701749305).
- Alara, O. R., Abdurahman, N. H. and Olalere, O. A. (2020), Ethanolic extraction of flavonoids, phenolics and antioxidants from *Vernonia amygdalina* leaf using two-level factorial design, *J. King Saud Univ. – Sci.*, **32**(1), 7–16, DOI: [10.1016/j.jksus.2017.08.001](https://doi.org/10.1016/j.jksus.2017.08.001).
- Alrasheed, H. S. and Alghamdi, A. K. H. (2023), Project-based learning at a Saudi university: Faculty and student feedback, *J. Teach. Educ. Sustainability*, **25**(1), 22–39, DOI: [10.2478/jtes-2023-0003](https://doi.org/10.2478/jtes-2023-0003).
- Amadeus, B. and Lorenz, M. (2023), Alles richtig – nichts verstanden? *Chimia*, DOI: [10.2533/chimia.2023.672](https://doi.org/10.2533/chimia.2023.672).
- Analytical Methods Committee. (1994), Is my calibration linear? *Analyst*, **119**(11), 2363–2366, DOI: [10.1039/AN9941902363](https://doi.org/10.1039/AN9941902363).
- Araujo, P. (2009), Key aspects of analytical method validation and linearity evaluation, *J. Chromatogr. B*, **877**(23), 2224–2234, DOI: [10.1016/j.jchromb.2008.09.030](https://doi.org/10.1016/j.jchromb.2008.09.030).
- Areljung, S., Leden, L. and Wiblom, J. (2021), Expanding the notion of “ownership” in participatory research involving teachers and researchers, *Int. J. Res. Method Educ.*, DOI: [10.1080/1743727x.2021.1892060](https://doi.org/10.1080/1743727x.2021.1892060).
- Ayaz, M. F. and Söylemez, M. (2015), The effect of the project-based learning approach on the academic achievements of the students in science classes in Turkey: A meta-analysis study, *Eğitim ve Bilim*, **40**(178), 255–283, DOI: [10.15390/EB.2015.4000](https://doi.org/10.15390/EB.2015.4000).
- Bain, K., Rodriguez, J.-M. G. and Towns, M. H. (2019), Chemistry and mathematics: Research and frameworks to explore student reasoning, *J. Chem. Educ.*, **96**(10), 2086–2096, DOI: [10.1021/acs.jchemed.9b00523](https://doi.org/10.1021/acs.jchemed.9b00523).
- Bicak, B. E., Borchert, C. E. and Höner, K. (2021), Measuring and fostering preservice chemistry teachers' scientific reasoning competency, *Educ. Sci.*, **11**(9), 496, DOI: [10.3390/educsci11090496](https://doi.org/10.3390/educsci11090496).
- Brandriet, A., Rupp, C. A., Lazenby, K. and Becker, N. M. (2018), Evaluating students' abilities to construct mathematical models from data using latent class analysis, *Chem. Educ. Res. Pract.*, DOI: [10.1039/c7rp00126f](https://doi.org/10.1039/c7rp00126f).
- Braun, V. and Clarke, V. (2021), One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qual. Res. Psychol.*, **18**(3), 328–352, DOI: [10.1080/14780887.2020.1769238](https://doi.org/10.1080/14780887.2020.1769238).
- Brederode, M. E., van Zoon, S. A. and Meeter, M. (2020), Examining the effect of lab instructions on students' critical thinking during a chemical inquiry practical, *Chem. Educ. Res. Pract.*, DOI: [10.1039/d0rp00020e](https://doi.org/10.1039/d0rp00020e).
- Bretz, S. L. (2019), Evidence for the importance of laboratory courses, *J. Chem. Educ.*, **96**(2), 193–195, DOI: [10.1021/acs.jchemed.8b00874](https://doi.org/10.1021/acs.jchemed.8b00874).
- Brewer, W. F. and Lambert, B. L. (2001), The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process, *Philos. Sci.*, DOI: [10.1086/392907](https://doi.org/10.1086/392907).
- Cantwell, H. (ed.) (2025), *Eurachem Guide: The fitness for purpose of analytical methods—A laboratory guide to method validation and related topics*, Eurachem.
- Davidowitz, B., Lubben, F. and Rollnick, M. (2001), Undergraduate science and engineering students' understanding of the reliability of chemical data, *J. Chem. Educ.*, DOI: [10.1021/ed078p247](https://doi.org/10.1021/ed078p247).
- Deffner, S. and Hermanns, J. (2025), Inquiry-based learning in a newly designed laboratory course for preservice chemistry teachers using a construction kit for planning experiments, *J. Chem. Educ.*, **102**(8), 3207–3217, DOI: [10.1021/acs.jchemed.4c01057](https://doi.org/10.1021/acs.jchemed.4c01057).
- Domin, D. S. (1999), A review of laboratory instruction styles, *J. Chem. Educ.*, **76**(4), 543, DOI: [10.1021/ed076p543](https://doi.org/10.1021/ed076p543).



- European Medicines Agency. (2023) *ICH Q2(R2): Validation of analytical procedures*. Available at: <https://www.ema.europa.eu/en/ich-q2r2-validation-analytical-procedures-scientific-guideline>.
- Eymur, G. and Çetin, P. S. (2024), Investigating the role of an inquiry-based science lab on students' scientific literacy, *Instruct. Sci.*, **52**(5), 743–760, DOI: [10.1007/s11251-024-09672-w](https://doi.org/10.1007/s11251-024-09672-w).
- Fabjan, N. et al. (2003), Tartary buckwheat as a source of dietary rutin and quercitrin, *J. Agric. Food Chem.*, **51**(22), 6452–6455, DOI: [10.1021/jf034543e](https://doi.org/10.1021/jf034543e).
- Farley, E. R., Fringer, V. and Wainman, J. W. (2021), Simple approach to incorporating experimental design into a general chemistry lab, *J. Chem. Educ.*, **98**(2), 350–356, DOI: [10.1021/acs.jchemed.0c00921](https://doi.org/10.1021/acs.jchemed.0c00921).
- Ferreira, M. M. and Trudel, A. R. (2012), The impact of problem-based learning on student attitudes toward science, problem-solving skills, and sense of community, *J. Classroom Interact.*, **47**(1), 23–30.
- Ferreira, B. D. et al. (2025), Is everything wrong in analytical chemistry? A study on reproducibility, *Accredit. Qual. Assur.*, DOI: [10.1007/s00769-025-01649-7](https://doi.org/10.1007/s00769-025-01649-7).
- Gao, R. and Lloyd, J. (2020), Precision and accuracy: Knowledge transformation through conceptual learning and inquiry-based practices, *J. Chem. Educ.*, DOI: [10.1021/acs.jchemed.9b00563](https://doi.org/10.1021/acs.jchemed.9b00563).
- Hamer, M. et al. (2024), Real-world sample-based teaching in analytical chemistry, *J. Chem. Educ.*, **101**(1), 205–214, DOI: [10.1021/acs.jchemed.3c00817](https://doi.org/10.1021/acs.jchemed.3c00817).
- Harborne, A. J. (1998), *Phytochemical methods: A guide to modern techniques of plant analysis*, Springer.
- Harris, D. C. (2016), *Quantitative chemical analysis*, W.H. Freeman.
- Hibbert, D. B. (2023), *Compendium of terminology in analytical chemistry (IUPAC Orange Book)*, Royal Society of Chemistry.
- Ifan, A., Imran, M., Khalid, M., Sami Ullah, M., Khalid, N., Assiri, M. A., Thomas, R., Muthu, S., Raza Basra, M. A., Hussein, M., Al-Sehemi, A. G. and Shahzad, M. (2021), Phenolic and flavonoid contents in *Malva sylvestris* and exploration of active drugs as antioxidant and anti-COVID-19 by quantum chemical and molecular docking studies, *J. Saudi Chem. Soc.*, **25**(8), 101277, DOI: [10.1016/j.jscs.2021.101277](https://doi.org/10.1016/j.jscs.2021.101277).
- ISO. (2023), *ISO 5725-1: Accuracy (trueness and precision) of measurement methods and results*, available at: <https://www.iso.org/standard/69418.html>.
- Keiner, L. and Graulich, N. (2020), Transitions between representational levels: Characterization of organic chemistry students' mechanistic features when reasoning about laboratory work-up procedures, *Chem. Educ. Res. Pract.*, **21**(2), 469–482, DOI: [10.1039/C9RP00241C](https://doi.org/10.1039/C9RP00241C).
- Konieczka, P. (2007), The role of method validation in QA/QC systems, *Crit. Rev. Anal. Chem.*, DOI: [10.1080/10408340701244649](https://doi.org/10.1080/10408340701244649).
- Kreft, I., Fabjan, N. and Yasumoto, K. (2006), Rutin content in buckwheat (*Fagopyrum esculentum* Moench) food materials and products, *Food Chem.*, **98**(3), 508–512, DOI: [10.1016/j.foodchem.2005.05.081](https://doi.org/10.1016/j.foodchem.2005.05.081).
- Kwitonda, J. de D., Sibomana, A., Gakuba, E. and Karegeya, C. (2021), Inquiry-based experimental design for enhancement of teaching and learning of chemistry concepts, *Afr. J. Educ. Stud. Math. Sci.*, **17**(2), 13–25, DOI: [10.4314/AJESMS.V17I2.2](https://doi.org/10.4314/AJESMS.V17I2.2).
- Lawrie, G. (2021), Considerations of sample size in chemistry education research, *Chem. Educ. Res. Pract.*, DOI: [10.1039/d1rp90009a](https://doi.org/10.1039/d1rp90009a).
- Márquez-Barreto, A. C., Quiñones-Flores, C. M., Ramírez-Alonso, G., Sámano-Lira, G. and Camarillo-Cisneros, J., (2023), Computational chemistry as an educational tool in health sciences, in Trujillo-Romero, C. J., Gonzalez-Landaeta, R., Chapa-González, C., Dorantes-Méndez, G., Flores, D.-L., Flores Cuautle, J. J. A., Ortiz-Posadas, M. R., Salido Ruiz, R. A. and Zuñiga-Aguilar, E. (eds), *XLV Mexican Conference on Biomedical Engineering*, Cham: Springer International Publishing, pp. 94–103.
- Martins, M. and Macagno, F. (2022), An analytical instrument for coding and assessing argumentative dialogues, *Sci. Educ.*, **106**(3), 573–609, DOI: [10.1002/scce.21708](https://doi.org/10.1002/scce.21708).
- McLaughlin, S., Amir, H., Garrido, N., Turnbull, C., Rouncefield-Swales, A., Swadźba-Kwaśny, M. and Morgan, K. (2024), Evaluating the impact of project-based learning in supporting students with the A-level chemistry curriculum in Northern Ireland, *J. Chem. Educ.*, **101**(2), 537–546, DOI: [10.1021/acs.jchemed.3c01184](https://doi.org/10.1021/acs.jchemed.3c01184).
- Miller, J. N. and Miller, J. C. (2018), *Statistics and chemometrics for analytical chemistry*, Pearson.
- Mitarlis, I., Rahayu, S. and Sutrisno, S. (2020), The effectiveness of new inquiry-based learning (NIBL) for improving multiple higher-order thinking skills (M-HOTS) of prospective chemistry teachers, *Eur. J. Educ. Res.*, **9**(3), 1309–1325, DOI: [10.12973/eu-jer.9.3.1309](https://doi.org/10.12973/eu-jer.9.3.1309).
- Moju, M., Taylor, L. and Iweuno, B. (2025), Tackling the challenge of chemical representations through sensemaking practices in chemistry education, *Discover Educ.*, **4**(1), 352, DOI: [10.1007/s44217-025-00703-3](https://doi.org/10.1007/s44217-025-00703-3).
- Moutsakis, G., Paschalidou, K. and Salta, K. (2025), Chemistry laboratory experiments focusing on students' engagement in scientific practices and central ideas of chemical practices, *Chem. Teach. Int.*, **7**(1), 173–182, DOI: [10.1515/cti-2024-0070](https://doi.org/10.1515/cti-2024-0070).
- National Research Council. (2013), *Next Generation Science Standards*, National Academies Press.
- OECD. (2023), *Agency in the Anthropocene: Supporting document to the PISA 2025 Science Framework*, OECD.
- Patel, B. A. (2022), A chemical analysis laboratory class assessed on accuracy and precision, *J. Chem. Educ.*, DOI: [10.1021/acs.jchemed.2c00583](https://doi.org/10.1021/acs.jchemed.2c00583).
- Petritis, S. J., Kelley, C. and Talanquer, V. (2022), Analysis of factors that affect the nature and quality of student laboratory argumentation, *Chem. Educ. Res. Pract.*, **23**(2), 408–420, DOI: [10.1039/D1RP00298H](https://doi.org/10.1039/D1RP00298H).
- Ramos, R. T. M., Bezerra, I. C. F., Ferreira, M. R. A. and Soares, L. A. L. (2017), Spectrophotometric quantification of



- flavonoids in herbal material, crude extract, and fractions from leaves of *Eugenia uniflora* Linn., *Pharmacogn. Res.*, **9**(3), 253–260, DOI: [10.4103/pr.pr.143_16](https://doi.org/10.4103/pr.pr.143_16).
- Raposo, F. and Barceló, D. (2021), Assessment of goodness-of-fit for calibration models, *TrAC, Trends Anal. Chem.*, DOI: [10.1016/j.trac.2021.116373](https://doi.org/10.1016/j.trac.2021.116373).
- Reid, N. and Shah, I. (2007), The role of laboratory work in university chemistry, *Chem. Educ. Res. Pract.*, **8**(2), 172–185, DOI: [10.1039/B5RP90026C](https://doi.org/10.1039/B5RP90026C).
- Reith, M. and Nehring, A. (2024), Fostering scientific reasoning competencies: Investigating impacts of cross- and within-content area experimentation using the competency triad, *Int. J. Sci. Educ.*, **48**(1), 73–97, DOI: [10.1080/09500693.2024.2394708](https://doi.org/10.1080/09500693.2024.2394708).
- Riemer, N., Eidner, S. and Hermanns, J. (2025), Laboratory courses for pre-service chemistry teachers between acquisition of skills and didactic double decker, *Laboratories*, **2**(2), 1–16, DOI: [10.3390/laboratories2020010](https://doi.org/10.3390/laboratories2020010).
- Robinson, J. K. (2013), Project-based learning: Improving student engagement and performance in the laboratory, *Anal. Bioanal. Chem.*, **405**(1), 7–13, DOI: [10.1007/s00216-012-6473-x](https://doi.org/10.1007/s00216-012-6473-x).
- Sampson, V., Grooms, J. and Walker, J. P. (2011), Argument-driven inquiry, *Sci. Educ.*, **95**(2), 217–257, DOI: [10.1002/sce.20421](https://doi.org/10.1002/sce.20421).
- Sanchez, J. M. (2024), Integrating measurement uncertainty analysis into laboratory education, *J. Chem. Educ.*, **101**(11), 4783–4789, DOI: [10.1021/acs.jchemed.4c00583](https://doi.org/10.1021/acs.jchemed.4c00583).
- Scoggin, J. and Smith, K. C. (2023), Enabling general chemistry students to take part in experimental design activities, *Chem. Educ. Res. Pract.*, **24**(4), 1229–1242, DOI: [10.1039/D3RP00088E](https://doi.org/10.1039/D3RP00088E).
- Seery, M. K., Agustian, H. Y. and Zhang, X. (2019), A framework for learning in the chemistry laboratory, *Isr. J. Chem.*, **59**(6–7), 546–553, DOI: [10.1002/jch.201800093](https://doi.org/10.1002/jch.201800093).
- Sevian, H. and Talanquer, V. (2014), Rethinking chemistry: A learning progression on chemical thinking, *Chem. Educ. Res. Pract.*, **15**(1), 10–23, DOI: [10.1039/C3RP00111C](https://doi.org/10.1039/C3RP00111C).
- Skoog, D. A., Holler, F. J. and Crouch, S. R. (2017), *Principles of instrumental analysis*, Cengage.
- Smyslova, O. A., Bokov, D. O., Potanina, O. G., Litvinova, T. M. and Samylina, I. A. (2019), Development and validation of spectrophotometric procedure for quantitative determination of flavonoid content used to control the quality of mixture herbal product, *J. Adv. Pharm. Technol. Res.*, **10**(4), 155–162, DOI: [10.4103/japtr.JAPTR_61_19](https://doi.org/10.4103/japtr.JAPTR_61_19).
- Stephenson, N. S. et al. (2020), Development and validation of scientific practices assessment tasks, *J. Chem. Educ.*, **97**(4), 884–893, DOI: [10.1021/acs.jchemed.9b00897](https://doi.org/10.1021/acs.jchemed.9b00897).
- Suradika, A., Dewi, H. I. and Nasution, M. I. (2023), Project-based learning and problem-based learning models in critical and creative students, *J. Pendidikan IPA Indonesia*, **12**(1), 153–167, DOI: [10.15294/jpii.v12i1.39713](https://doi.org/10.15294/jpii.v12i1.39713).
- Szalay, L. and Tóth, Z. (2016), An inquiry-based approach of traditional “step-by-step” experiments, *Chem. Educ. Res. Pract.*, **17**(4), 923–961, DOI: [10.1039/C6RP00044D](https://doi.org/10.1039/C6RP00044D).
- Szalay, L., Tóth, Z., Borbás, R. and Füzesi, I. (2024), Progress in developing experimental design skills among junior high school learners, *J. Turk. Sci. Educ.*, **21**(3), 484–511, DOI: [10.36681/tused.2024.026](https://doi.org/10.36681/tused.2024.026).
- Taber, K. S. (2024), Is educational research science, superstition or confidence trick? *Nat. Rev. Chem.*, DOI: [10.1038/s41570-024-00582-6](https://doi.org/10.1038/s41570-024-00582-6).
- Talanquer, V. (2006), Commonsense chemistry: A model for understanding students’ alternative conceptions, *J. Chem. Educ.*, **83**(5), 811–816, DOI: [10.1021/ed083p811](https://doi.org/10.1021/ed083p811).
- Talanquer, V. (2011), School chemistry: The need for transgression, *Sci. Educ.*, **22**(1), 175–184, DOI: [10.1007/s11191-011-9392-x](https://doi.org/10.1007/s11191-011-9392-x).
- Talanquer, V. (2018), Progressions in reasoning about structure–property relationships, *Chem. Educ. Res. Pract.*, **19**(4), 998–1009, DOI: [10.1039/C7RP00187H](https://doi.org/10.1039/C7RP00187H).
- Taverniers, I., De Loose, M. and Van Bockstaele, E. (2004), Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance, *TrAC, Trends Anal. Chem.*, **23**(8), 535–552, DOI: [10.1016/j.trac.2004.04.001](https://doi.org/10.1016/j.trac.2004.04.001).
- Tümay, H. (2023), Systems thinking in chemistry education, *J. Chem. Educ.*, **100**(10), 3925–3933, DOI: [10.1021/acs.jchemed.3c00474](https://doi.org/10.1021/acs.jchemed.3c00474).
- van Wyk, A. L., Bhinu, A., Frederick, K. A., Lieberman, M. and Cole, R. S. (2025), Bridging the science practices gap: Analyzing laboratory materials for their opportunities for engagement in science practices, *J. Chem. Educ.*, **102**(3), 970–983, DOI: [10.1021/acs.jchemed.4c00744](https://doi.org/10.1021/acs.jchemed.4c00744).
- Viera, L. I., Ramírez, S. S. and Fleisner, A. (2017), El laboratorio en química orgánica: Una propuesta para la promoción de competencias científico-tecnológicas, *Educ. Quím.*, **28**(4), 262–268, DOI: [10.1016/j.eq.2017.04.002](https://doi.org/10.1016/j.eq.2017.04.002).
- Wellhöfer, L. and Lühken, A. (2021), Problem-based learning in an introductory inorganic laboratory: Identifying connections between learner motivation and implementation, *J. Chem. Educ.*, **98**(12), 3779–3787, DOI: [10.1021/acs.jchemed.1c00808](https://doi.org/10.1021/acs.jchemed.1c00808).
- Worrall, A. F., Bergstrom Mann, P. E., Young, D., Wormald, M. R., Cahill, S. T. and Stewart, M. I. (2020), Benefits of simulations as remote exercises during the COVID-19 pandemic: An enzyme kinetics case study, *J. Chem. Educ.*, **97**(9), 2733–2737, DOI: [10.1021/acs.jchemed.0c00607](https://doi.org/10.1021/acs.jchemed.0c00607).
- Xu, H. and Talanquer, V. (2012), Effect of the level of inquiry of lab experiments on general chemistry students’ written reflections, *J. Chem. Educ.*, **89**(7), 840–845, DOI: [10.1021/ed3002368](https://doi.org/10.1021/ed3002368).
- Zhao, L., Zhao, B. and Li, C. (2023), Alignment analysis of teaching–learning–assessment within the classroom: How teachers implement project-based learning under the curriculum standards, *Discipl. Interdiscipl. Sci. Educ. Res.*, **5**(1), 78, DOI: [10.1186/s43031-023-00078-1](https://doi.org/10.1186/s43031-023-00078-1).
- Zhunissova, S., Zhussupova, L., Abyzbekova, G. and Balyk-bayeva, G. (2025), A review of teaching experimental design in chemistry, *J. Chem. Educ.*, **102**(9), 3817–3827, DOI: [10.1021/acs.jchemed.5c00529](https://doi.org/10.1021/acs.jchemed.5c00529).

