

Article

Process-Oriented Assessment of Project-Based Learning in Analytical Chemistry: Tracking Students' Experimental Decision-Making Pathways

Galiya Madybekova and Tansholpan Zabynbekova *

Department of Chemistry, Zhanibekov University, Shymkent 160000, Kazakhstan; galiya56@list.ru

* Correspondence: tzabynbekova@gmail.com

Abstract

Learning analytical chemistry requires not only mastery of analytical and calculation methods but also the ability to make consistent experimental decisions. This study tested a process-oriented assessment framework based on the experimental decision-making pathway (EDMP) construct and the analysis of student-specific decision-making trajectories. A quasi-experimental study involved 60 second-year students, 30 each in a project-based learning group and a control group with a traditional laboratory format. During an eight-week module on determining total iron in water, 2317 individual decisions were recorded; independence, justification, use of data, and revision patterns were coded for each event. Dynamics were analyzed using generalized linear mixed models and sequence analysis, and trajectory characteristics were compared with educational and analytical outcomes. In the experimental group, the proportion of independent decisions increased from 35.7 to 61.5%, and the proportion of data-driven decisions increased from 32.0 to 58.2%. In the control group, changes were less pronounced. The between-group difference was greater for analytical problem solving ($d = 1.26$) than for subject knowledge ($d = 0.84$). The proportion of data-driven solutions correlated positively with the final problem-solving score ($\rho = 0.52$) and negatively with the relative error in iron determination ($\rho = -0.44$). Overall, the results supported the hypotheses and demonstrated that process analysis complements the final indicators, allowing us to trace the transition from dependent procedure execution to more independent and adaptive experimental reasoning.

Keywords: project-based learning; analytical chemistry; process-oriented assessment; experimental decision making; decision trajectories

Academic Editor: Michael DeBraga

Received: 21 August 2026

Revised: 3 October 2026

Accepted: 5 October 2026

Published: 8 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Analytical chemistry occupies a unique position among chemical disciplines, as its content is inextricably linked to the acquisition, evaluation, and interpretation of quantitative information about the composition of a substance. Mastering this discipline requires knowledge of chemical reactions, analytical methods, calculation procedures, and laboratory techniques. However, this knowledge alone is far from exhaustive of an analyst's competence. Equally important is the student's ability to make consistent experimental decisions, which determine the course and outcome of the entire analysis.

At virtually every stage of the study, the student is faced with a choice. They must decide which analytical method is appropriate for a specific task, how to collect and prepare the sample, whether pretreatment is required, and over what concentration range to construct the calibration curve. If an anomalous result occurs, the student must determine its probable cause, determine whether the measurement should be repeated, and evaluate whether the obtained data can be considered sufficiently reliable to draw a conclusion. Each of these decisions builds on previous actions and, in turn, sets the conditions for subsequent steps. Therefore, it is appropriate to view an analytical experiment not as a linear execution of instructions, but as a sequence of interdependent decisions. The quality of an analysis is determined not only by adherence to the procedure but also by the student's ability to select a method, interpret data, recognize deviations, and reasonably modify the experimental strategy (Hofstein & Lunetta, 2004; Seery et al., 2024).

Despite the central role of the laboratory in the training of chemists, the prevailing model of laboratory instruction poorly develops this decision-making ability. A typical laboratory lesson follows a fixed pattern in which the student receives instructions, carries out a prescribed procedure, performs a calculation, and prepares a report. This format, referred to in the literature as expository or prescription-based, communicates virtually all essential parameters of the work to the student in advance, namely, the component to be determined, the analytical method, the list of reagents, the required volumes, the sequence of operations, calculation formulas, and criteria for the correctness of the result (Domin, 1999).

With this type of organization, students are able to successfully complete an analysis without actually making a single independent analytical decision. These concerns are consistent with the broader laboratory-education literature emphasizing inquiry, student decision making, and higher-order reasoning rather than procedural completion alone (Hofstein & Lunetta, 2004; Seery et al., 2024). Recent work in analytical chemistry has likewise emphasized the incorporation of active-learning approaches into undergraduate courses (Kovarik et al., 2022).

More broadly, assessment research has emphasized the value of using evidence about learning processes in addition to end-point performance (Black & Wiliam, 1998). In laboratory education, recent work similarly emphasizes the need to align assessment with authentic scientific practices, including decision making, interpretation, and working with uncertainty (Seery et al., 2024; Agustian, 2025).

One way to increase the number of meaningful choice points is project-based learning (PjBL), in which work is organized around an open-ended task and unfolds over several interrelated stages (Blumenfeld et al., 1991; Krajcik & Blumenfeld, 2006). In analytical chemistry, this approach has already been used in multi-stage research projects and courses, where students participated in planning, selecting procedures, interpreting data, and presenting results (Matilainen et al., 2021; Muna, 2021).

The advantage of PjBL for analytical chemistry lies not only in the greater openness of the assignment (Thomas, 2000). It creates situations in which the student must decide which method is appropriate for the task, how to organize calibration and quality control, what constitutes an anomalous result, and what procedural modifications are justified by the available data. Research in university chemistry laboratories shows that it is the ability to independently determine procedural and analytical elements that distinguishes more inquiry-oriented formats from recipe-based assignments (Zammit et al., 2024). In analytical chemistry, the transition to more authentic inquiry-based assignments remains a relevant direction for updating laboratory courses (Van Wyk et al., 2025; Kovarik, 2026).

Existing studies, however, often focus on project design, inquiry levels, student experiences, or final learning outcomes. Much less frequently, the analysis focuses on the temporal organization of experimental reasoning: what decisions follow one another,

when students consult data or the instructor, and how they change their strategy after an unexpected result. Therefore, the final score alone is insufficient to answer the question of how independence develops over the course of a project.

In this study, product-oriented and process-oriented assessments are viewed as complementary. The former characterizes the final knowledge and quality of the completed analytical task; the latter reconstructs the sequence of decisions, the sources of their justification, requests for support, and revisions to the experimental strategy. This approach is close to process data analysis, where the unit of information is not only the outcome but also the ordered sequence of actions (Bannert et al., 2014).

This is particularly important for laboratory learning, as the same end result can be achieved in different ways: through meaningful diagnosis and adjustment or through mechanical repetition and direct assistance from the instructor. Process recording allows us to distinguish between these cases and link changes in reasoning to the objective quality of the completed analysis.

To describe this dynamic, we propose the experimental decision-making pathway (EDMP) as a domain-specific, process-oriented assessment framework for representing how students make, justify, evaluate, and revise experimental decisions during an analytical task. EDMP is not intended as an alternative theory of scientific reasoning or metacognition; rather, it operationalizes selected elements of these broader constructs in the specific context of analytical laboratory activity.

The definition of EDMP draws on the notion of scientific reasoning as a coordinated search between the space of hypotheses and the space of experiments (Klahr & Dunbar, 1988), as well as on the idea of epistemically valid inquiry, in which the student plans the procedure and evaluates evidence, rather than simply reproducing a given algorithm (Chinn & Malhotra, 2002). In chemistry education, similar processes are examined through metacognitive regulation and critical reflection on experimental actions (Cooper & Sandi-Urena, 2009). The EDMP framework therefore shifts the focus from isolated decisions to the structure and development of decision trajectories over time.

EDMP differs from related constructs primarily in its level of operationalization and unit of analysis. Scientific reasoning describes broad cognitive processes involved in coordinating hypotheses, experiments, and evidence, whereas EDMP represents these processes as observable sequences of domain-specific analytical decisions. Metacognitive regulation concerns planning, monitoring, and evaluation of one's activity, but EDMP additionally records what experimental decision was made, who initiated it, what evidence supported it, whether it was analytically correct, and whether subsequent data led to revision. Inquiry frameworks primarily characterize the structure and openness of the learning environment, whereas EDMP focuses on the student-specific pathway that unfolds within that environment. Finally, process mining and sequence analysis are methodological approaches for examining ordered event data; EDMP provides the analytical-chemistry-specific event ontology and attributes to which such methods can be applied.

The distinct contribution of EDMP therefore lies not in introducing entirely new cognitive processes, but in integrating them into a two-level representation of laboratory reasoning. The first level captures the temporal sequence of analytical decision states, such as method selection, calibration, measurement, anomaly detection, troubleshooting, revision, and interpretation. The second level characterizes each decision in terms of autonomy, justification, evidence use, correctness, teacher support, and revision. This combination makes it possible to distinguish, for example, an analytically correct action performed after direct instruction from an independently generated and evidence-based decision, even when both lead to the same final result.

Thus, the research gap lies not in the absence of theoretical accounts of scientific reasoning, metacognition, or inquiry, but in the limited availability of domain-specific

frameworks that translate these processes into temporally ordered, observable, and quantitatively analyzable decision trajectories in analytical chemistry laboratories. EDMP is proposed to address this methodological and conceptual gap. Recent work demonstrates growing interest in authenticity, inquiry, and laboratory learning practices (Zammit et al., 2024; Van Wyk et al., 2025; Agustian, 2025), but analytical chemistry students' experimental decision trajectories rarely serve as a standalone, quantitatively descriptive unit of analysis. This aspect is the focus of this paper.

The proposed assessment framework is based on a conceptual model that describes experimental decision making as a cyclical process. The starting point is the analytical problem, followed by problem definition, method selection, experimental design, measurement, data evaluation, decision making, and, finally, decision revision or confirmation, leading to an analytical conclusion. Crucially, this cycle is not strictly linear. It allows for backtracking, in which an unexpected measurement result triggers a search for a cause and leads to a new measurement before the work moves forward (measurement → unexpected result → problem resolution → new measurement). It is the presence of such backtracking and its internal structure that provide diagnostically valuable information about the quality of experimental reasoning. This understanding is consistent with the model of scientific knowledge as a coordinated search in two spaces (Klahr & Dunbar, 1988) and with the criteria for epistemically valid research (Chinn & Malhotra, 2002). To operationalize the cycle, the solution was considered a multidimensional event. Its quality was described by individual characteristics: level of independence, validity, use of experimental or theoretical evidence, correctness, and the nature of revision. This distinction is fundamental: an independent solution can be erroneous, while a correct action can be performed following direct instructions. Therefore, the analysis did not use a single, summary scale of "solution level"; instead, individual dimensions were retained, allowing for the reconstruction of the trajectory structure.

The diagnostic model thus includes two interconnected levels: the sequence of solution types over time and the set of attributes for each solution. The first level demonstrates how the student moves between problem formulation, measurement, anomaly detection, diagnosis, and revision; the second demonstrates the independence and evidence-based nature of each transition. This representation makes the EDMP compatible with both sequence analysis and the assessment of metacognitive regulation in chemical problem solving (Cooper & Sandi-Urena, 2009; Sandi-Urena et al., 2011).

The aim of this study is to develop and test a process-oriented assessment framework for project-based learning in analytical chemistry using the EDMP construct and the tracking, coding, and analysis of student-specific decision-making trajectories.

The study is structured around four research questions. The first question asks how students' experimental decision-making trajectories change during project-based learning in analytical chemistry. The second question asks whether project-based learning increases independence, validity, and reliance on experimental data in decision-making compared to traditional laboratory classes. The third question addresses the types of trajectories that distinguish students with high and low levels of success in solving analytical problems. The fourth question asks how trajectory characteristics relate to subject knowledge in analytical chemistry and competence in solving experimental problems.

The formulated questions are specified in a system of testable hypotheses. According to the first hypothesis, after project-based learning, the proportion of independent, experimental solutions will increase. The second hypothesis predicts an increase in the proportion of solutions supported by analytical justification and experimental data. According to the third hypothesis, the number of unjustified actions, mechanical repetitions, and solutions initiated entirely by the instructor will decrease. The fourth hypothesis posits that students in the upper tertile of the final analytical problem-solving test will exhibit

trajectories characterized by a greater frequency of “data → assessment → justified revision” cycles and a lower frequency of “error → instructor contact → direct correction” cycles than students with lower final problem-solving performance. Finally, according to the fifth hypothesis, the quality of the solution trajectory will be positively related to the results of assessment based on the completion of real analytical tasks.

2. Materials and Methods

2.1. Study Design

The study utilized a mixed-methods explanatory design, combining a quasi-experimental comparison of two groups with pre- and post-intervention measurements with an integrated process analysis of experimental decision making. The quantitative approach examines whether subject knowledge, analytical problem-solving competence, and trajectory characteristics differ over time between the project-based and traditional laboratory conditions, while the process and qualitative approaches explore how decision making unfolds during laboratory work. The unit of process analysis is not the student as a whole, but rather a single, recorded experimental decision, allowing us to describe the dynamics of reasoning at the level of specific actions and reconstruct individual decision sequences (Bannert et al., 2014).

Participants were assigned to conditions within two established study groups, making the design quasi-experimental rather than randomized at the individual student level. Equivalence of the groups at baseline was tested for age, gender, grade point average (GPA), previous chemistry grade, and pretest scores, and any baseline differences found were then controlled for as covariates.

2.2. Participants and Study Context

The sample consisted of 60 second-year students at [blinded] University enrolled in programs 6B01504 “Chemistry Teacher Training” and 6B01507 “Chemistry and Biology Teacher Training.” The average age of the participants was 19.4 years (SD = 0.7; range 18–22 years), including 41 female students (68.3%) and 19 male students (31.7%). The average GPA on a four-point scale was 3.21 (SD = 0.32). The sample represented the full available sample of two established study groups; program assignment was recorded during a baseline comparability check.

The experimental group (EXP) consisted of 30 students enrolled in a project-based learning format, whereas the control group (CTRL) consisted of 30 students enrolled in a traditional laboratory format. Inclusion criteria for the analysis were attendance of at least 80% of classes and completion of all course stages. All 60 participants met this criterion and were retained in the final analysis; no attrition was recorded. Each student was assigned an anonymous identifier from S001 to S060, and identifying information was not entered into the analytical tables.

2.3. Experimental Treatment and Control Condition

Both conditions were taught by the same instructor throughout the eight-week module. Using the same instructor was intended to reduce variability associated with teaching experience, instructional style, assessment practices, and familiarity with the analytical content. The instructor implemented the pedagogical format assigned to each condition while maintaining comparable laboratory time, analytical content, and assessment requirements. The experimental group completed a research project around an open-ended analytical problem: students formulated a working strategy, selected an analytical method, planned the experiment, prepared samples, constructed calibration, organized quality control, troubleshooted, and interpreted the results. Instructor support in the

project condition was primarily guiding and focused on questioning and discussing the rationale for their choices, rather than conveying a completed algorithm. The control group studied the same analytical material in a recipe-based, or expository, format, in which the main procedural parameters and operational sequence were predetermined (Domin, 1999). The amount of laboratory time, key analytical concepts, and final assessment were the same in both groups. After completing the module, all participants completed a single, standardized final analytical practical assignment, allowing for comparison of the quality of analysis regardless of differences in the organization of the learning activities.

2.4. Descriptive Check of Instructional Implementation

A descriptive implementation check was conducted to document whether the instructional characteristics intended to distinguish the two conditions were observable during the laboratory sessions. An independent observer used a structured lesson-observation protocol containing eight predefined characteristics: open-ended problem setting, independent planning, student choice of method, independent decision making, iterative revision, teamwork, final product, and reflection. Each characteristic was coded at the lesson level as present (1) when it was observably incorporated into the session and absent (0) when it was not observed. For each condition, the percentage reported for a characteristic represents the proportion of observed lessons in which that characteristic was coded as present.

This procedure was used as a descriptive implementation check rather than as a psychometric fidelity scale; the binary codes were not intended to quantify the quality or intensity of implementation. The same independent observer coded the observed lessons. Because a second observer did not independently code the eight implementation characteristics, inter-rater reliability was not available for this measure. Accordingly, the implementation percentages should be interpreted descriptively. A detailed breakdown is provided in Appendix A.1, Table A4.

2.5. Analytical Task in the Experimental and Control Conditions

The analytical topic was the same in both learning conditions: determination of total iron in natural and drinking water samples. The principal difference between the conditions concerned the degree of procedural openness and student decision making rather than the analytical content itself.

In the experimental group (EXP), the task was organized as an open-ended analytical project. Students were required to formulate a working strategy and justify key methodological and procedural decisions. In particular, they could make a justified choice between spectrophotometric determination with 1,10-phenanthroline and flame atomic absorption spectrometry (AAS). In the photometric variant, sample preparation included the conversion of iron to Fe(II) prior to complexation with 1,10-phenanthroline, whereas the AAS variant involved acidic sample preparation.

In the control group (CTRL), students worked with the same analytical topic and underlying concepts, but the laboratory activity followed a recipe-based format. The analytical method, principal procedural parameters, and sequence of operations were predetermined rather than selected and justified by the students. Thus, the two conditions differed primarily in the extent to which students were required to plan, select, justify, and revise experimental actions. Laboratory time and the key analytical concepts addressed were comparable across conditions.

Reference concentrations for assessing analytical performance were established by the teaching laboratory using a standardized procedure on independent aliquots. At the end of the eight-week module, both groups completed the same standardized final

practical assignment under identical conditions. Results from this common task were used for between-group comparisons of relative error and recovery so that differences in analytical performance were not confounded with the different instructional structures used during the module.

2.6. Measurement Instruments

Learning outcomes were measured before and after the module. The analytical chemistry subject knowledge test included 30 questions, each scored dichotomously (0/1); short answers were scored using a predetermined key. Therefore, internal consistency was estimated using the KR-20 coefficient and was 0.82. The primary outcome measure was the analytical problem-solving test, which focused on the quality of experimental reasoning (Table 1).

Table 1. Measurement instruments and their characteristics.

Tool	What Does it Measure?	Structure	Maximum	Reliability
Knowledge test	Subject knowledge in analytical chemistry	30 tasks, before and after	30	KR-20 = 0.82
Problem-solving test	Quality of experimental reasoning	6 situational tasks, 0–4 points	24	$\alpha = 0.78$; ICC(2,1) = 0.89, 95% CI [0.84, 0.93]
Section of the final practical assignment	Quality of execution of the final analytical task	8 domains	90	ICC(2,1) = 0.86, 95% CI [0.79, 0.91]
Decision log	Individual experimental solutions	multidimensional event coding	-	$\kappa = 0.84$ –0.88
Reflection and interview	Grounds and review of decisions	3 questions per lesson; 7 after the project	-	-

The analytical problem-solving test was developed for this study to assess students' ability to reason through common decision points encountered in analytical chemistry rather than reproduce factual knowledge or routine calculations. Its structure was aligned with the study's conceptualization of experimental reasoning as a sequence of decisions involving problem recognition, selection of an analytical strategy, evaluation of evidence, quality control, diagnosis of unexpected results, and interpretation of analytical data. The instrument therefore focused on six domains that represent recurrent decision-making demands in undergraduate analytical chemistry: method selection, sample collection, calibration, quality control, troubleshooting, and data interpretation.

The test consisted of six open-ended situational tasks, with one task representing each domain. Students were required not only to propose an appropriate analytical action but also to justify their decision using relevant analytical principles or information provided in the task. Each response was scored using a common five-level rubric: 0 = the analytical problem was not recognized; 1 = the problem was recognized, but no viable solution was proposed, or the proposed solution was incorrect; 2 = an analytically acceptable solution was proposed without justification; 3 = a correct solution was accompanied by a partial or incomplete analytical justification; and 4 = the solution was correct and fully justified using relevant analytical evidence or principles. Each task therefore contributed 0–4 points, yielding a total possible score of 0–24. Higher scores indicated stronger analytical problem-solving performance.

Content validity was evaluated by the same five-member expert panel described in Section 2.9, representing analytical chemistry, chemical education, physical chemistry, chemistry teaching methodology, inorganic chemistry, and general chemistry. The

experts independently evaluated the relevance and correspondence of each task to its intended analytical problem-solving domain using a four-point relevance scale. Their judgments were used to verify domain coverage, clarity of task formulation, and correspondence between the task requirements and scoring rubric.

Inter-rater reliability was evaluated using a two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient, ICC(2,1), because the two raters were treated as representatives of a broader population of potential raters and agreement in absolute scores was of interest. The ICC for the total analytical problem-solving score was 0.89, 95% CI [0.84, 0.93], and the ICC for the final practical assignment was 0.86, 95% CI [0.79, 0.91]. According to conventional reliability criteria, both values indicated good inter-rater reliability (Koo & Li, 2016/2017).

For the final practical assignment, the following technical characteristics were retained: the calibration coefficient of determination, number of calibration points, relative standard deviation, relative error, recovery, number of replicates, use of blank and control samples, obtained and reference concentrations, and number of rejected measurements. These indicators were used as external analytical criteria for assessing the quality of the assignment.

Reflexive tools were used to reconstruct the process of reasoning. After each lesson, students briefly answered three questions about the main decision they made, the data or knowledge on which it was based, and whether they changed their original plan and why. After completing the project, each student was interviewed in a semi-structured seven-question interview focused on the most difficult decisions, the reasons for their choices, instances of plan revision, requests for help, and changes in reactions to unexpected results. Additionally, an independent observer maintained a lesson log. Each observation-log entry was linked to the corresponding event number to facilitate alignment with the student decision log.

2.7. Experimental Solution Log and Coding Scheme

The central array was a structured journal of experimental solutions, in which one line corresponded to one recorded student solution. An event was defined as the moment when a student chose between two or more possible actions or assessed the need to maintain or change a current strategy. The initial recording of the event was carried out directly during laboratory work in an individual decision journal and included a brief description of the choice and its basis; After the lesson, notes were synchronized with the laboratory notebook and observation protocol. This predetermined criterion was applied to all events without subsequent selection of only “interesting” episodes. A total of 2317 decisions were registered; an average of 38.6 per student (SD = 9.2; range 21–59).

Each event included a student ID, event number, date/week, project stage, solution type, and a short text comment. The serial number made it possible to restore the individual sequence. The command code was retained as a contextual feature of group work, but decision-level statistical inferences accounted for repeated observations within a student. Project stage was coded into ten stages, solution type into twelve categories; Complete lists are provided in the Appendix A.1 (Tables A1 and A2).

Key qualitative characteristics of the decision were coded on ordinal scales reflecting autonomy, validity, and reliance on data. Decision independence was assessed from a complete lack of independence, when the teacher directly indicated what to do, to completely independent detection of the problem and choice of action. Validity was separated from the use of evidence, as the student is able to convincingly explain an incorrect decision as well as randomly select the correct one without understanding. For the same reason, the scientific correctness of the decision was coded as a separate indicator. A complete list of ordinal scales is provided in Appendix A.1 (Table A3).

Additionally, for each event, the reason for its occurrence was coded, which made it possible to distinguish between decisions dictated by the task, an unexpected result, data inconsistency, experimental error, a classmate's suggestion, a teacher's question, a student's own initiative, information from the literature, or limitations of equipment and reagents. The source of the evidence used, the initiator of the decision, the fact of detection of an anomaly and its type, as well as the reason for the revision were separately recorded. This set of fields made it possible to trace how the share of decisions initiated by the student himself changes over time, and how the source of support shifts from instructions and the teacher to one's own data and theoretical knowledge.

2.8. Data Collection Procedure

Data collection was organized across baseline, intervention, and post-intervention stages. Table 2 summarizes the temporal sequence of instructional activities and data collection across the eight-week module. Baseline measures were collected before the beginning of the module. Process data were collected continuously throughout Weeks 1–8, whereas outcome measures and the standardized practical assignment were administered after completion of the module. For the trajectory analyses, Weeks 1–3 were defined as the early period and Weeks 6–8 as the late period.

Table 2. Research chronogram and data collection across the eight-week analytical chemistry module.

Time	EXP Condition	CTRL Condition	Data Collected
Pre-intervention/baseline	Preparation for the analytical chemistry module	Preparation for the same analytical chemistry module	Demographic and academic characteristics; GPA; previous chemistry score; knowledge pre-test; analytical problem-solving pretest
Weeks 1–3: Early period	Open-ended project-based laboratory work with student planning, methodological choice, and decision making	Recipe-based laboratory work with predetermined procedures and operational sequence	Individual decision logs during laboratory work; observer records; brief reflective responses after each session
Weeks 4–5: Intermediate period	Continued project work, including experimental implementation, evaluation of results, and revision of decisions	Continued traditional laboratory work on the same analytical content	Individual decision logs; observer records; brief reflective responses after each session
Weeks 6–8: Late period	Continued project work with increasing emphasis on evaluation, troubleshooting, revision, and interpretation	Continued recipe-based laboratory work and completion of the corresponding analytical activities	Individual decision logs; observer records; brief reflective responses after each session
Post-intervention	Completion of the eight-week module	Completion of the eight-week module	Knowledge post-test; analytical problem-solving post-test; common standardized final practical assignment; analytical performance indicators; final reflective responses; semi-structured interview
After data collection	—	—	Coding of decision events using the predefined coding guide; construction of individual decision sequences; statistical and sequence analyses

Process data were therefore collected repeatedly throughout the intervention rather than only at fixed pre- and post-test points. Decision logs provided the primary event-level data, observer records were used to verify the context and sequencing of selected episodes, and reflective responses provided supplementary information about the reasoning underlying students' decisions. Coding was conducted after data collection using the predefined coding guide described in Sections 2.7 and 2.9.

2.9. Reliability and Validity of Coding

The content validity of the coding scheme was assessed by a panel of five experts selected to provide complementary disciplinary and pedagogical expertise relevant to experimental decision making in chemistry. The panel comprised a PhD Associate Professor specializing in analytical chemistry and chemical education with 18 years of professional experience; a Candidate of Chemical Sciences and Associate Professor specializing in physical chemistry with 21 years of experience; a PhD specialist in chemistry teaching methodology with 7 years of experience; a Doctor of Chemical Sciences and Professor specializing in inorganic chemistry with 37 years of experience; and a Candidate of Chemical Sciences specializing in general chemistry with 25 years of experience. The experts were selected based on the relevance of their disciplinary or pedagogical expertise to the coding framework, advanced academic qualifications, and substantial research and/or teaching experience in chemistry. The composition of the panel was intended to combine expertise in analytical and fundamental chemistry with expertise in chemistry education and instructional methodology. Each expert independently rated the relevance of every coding category using a four-point scale. With a five-member expert panel, an I-CVI of 0.80 corresponded to agreement on relevance by four of the five experts and was therefore used as the minimum criterion for item-level content adequacy. Unanimous agreement corresponded to an I-CVI of 1.00. To account for agreement expected by chance, modified kappa (κ) was also considered following Polit et al. (2007); values ranged from approximately 0.76 for four-of-five agreement to 1.00 for unanimous agreement. Across the coding categories, I-CVI values ranged from 0.80 to 1.00, while S-CVI/Ave was 0.93 (Polit & Beck, 2006; Polit et al., 2007).

The reliability of process coding was verified by independently re-coding 636 events (27.4% of the dataset). Cohen's κ coefficient was calculated for nominal variables (Cohen, 1960), whereas weighted κ was used for ordinal scales to account for the degree of disagreement between categories (Cohen, 1968). Agreement was $\kappa = 0.88$ for decision type and $\kappa = 0.87$ for project stage; for autonomy, validity, and use of evidence, the weighted κ ranged from 0.84 to 0.86. Disagreements were discussed until consensus was reached, after which the consensus version was used in further analysis.

2.10. Solution Trajectory Metrics

Based on the solution log, a set of derived metrics characterizing the student's trajectory was calculated for each student. These included solution density, the proportion of independent solutions, the proportion of data-driven solutions, the frequency of adaptive revisions, the proportion of solutions with direct instructor assistance, the diversity of solution types, independent troubleshooting, the proportion of successful revisions, and the complexity of the trajectory, measured by the number of unique states, transitions, and return cycles. These metrics were not entered manually but calculated programmatically from the primary dataset, reducing the risk of errors and ensuring reproducibility.

2.11. Data Analysis

Data processing was performed in IBM SPSS Statistics 29 and R 4.3 using the lme4 and TraMineR packages (Bates et al., 2015; Gabadinho et al., 2011). The significance level was set at 0.05. Holm's correction was applied to the domain-specific ANCOVA family; the prespecified main tests (week $\times$ group interactions, overall problem-solving score, and the two main H5 relationships) were interpreted as confirmatory. Additional correlations reported in the Appendix A were treated as exploratory and are presented uncorrected for multiplicity. Baseline comparability between groups was tested using independent-samples t-tests and χ^2 .

For the first and second research questions, outcome measures were analyzed by ANCOVA: final score served as the dependent variable, learning condition as the factor, and the corresponding entry score and GPA as covariates. The effect size of the ANCOVA was expressed as the partial η^2 ; Cohen's d was calculated separately for unadjusted post-test means and pooled standard deviations (Cohen, 1988). At the individual decision level, generalized linear mixed models with a logistic link were used. The probability of independent or data-driven decisions was predicted by week, condition, and their interaction; a random intercept for the student was included to account for repeated decisions by the same participant.

Trajectory dynamics were examined by analyzing the D1–D12 sequences. For sequences ordered by event number, state frequencies, transition matrices, and repeated subsequences were calculated (Bannert et al., 2014; Gabadinho et al., 2011). The Teacher and Evidence attributes were not considered as separate states of the sequence; they were analyzed conditionally for events after D8/D10. Particular attention was paid to the probability of the D8→D9 transition, the probability of D10 after D8, and the proportion of D8 events accompanied by a request for direct support. The early period included weeks 1–3, and the late period included weeks 6–8.

For the third research question, pairwise distances between sequences D1–D12 were calculated using the optimal matching method, followed by Ward's hierarchical clustering. Candidate solutions containing two to six clusters were examined. The number of clusters was selected using a combination of internal partition quality, assessed by the average silhouette width (ASW), cluster size balance, and substantive interpretability of the resulting trajectory profiles. The four-cluster solution was retained because it provided a favorable balance between within-cluster homogeneity, between-cluster separation, and theoretically meaningful distinctions among the observed decision-making patterns. For H4, student success was defined independently of trajectory characteristics. Students whose final analytical problem-solving test scores fell within the upper tertile of the total sample distribution were classified as high-success students. This classification was based exclusively on the final problem-solving outcome and was not used in sequence construction, optimal matching, or cluster formation. Trajectory clusters were derived solely from the D1–D12 sequences, after which the proportion of high-success students within each trajectory profile was examined. For transparency, ASW values for all candidate solutions from $k = 2$ to $k = 6$ are reported in Appendix A.1, Table A11. For H5, Spearman correlations were calculated between predefined trajectory characteristics and the final problem-solving score/relative error; additional relationships with other indicators were considered exploratory. Reflective notes were used to qualitatively illustrate quantitative patterns without constructing a separate, independent thematic model.

2.12. Ethical Aspects

The study was approved by the Zhanibekov University ethics committee. All participants were informed of the study's aims and procedures and provided written informed consent. They were also informed of their right to discontinue participation at any time without consequences for their academic performance. Personal data were stored separately from the analytical dataset, which used only anonymous identifiers, and access to matching codes and identities was restricted to a limited number of researchers.

3. Results

All 60 students were included in the analysis: EXP, $n = 30$, and CTRL, $n = 30$. The experimental module included an open-ended project to determine total iron in water, while the control condition used a recipe-based format; the unit of process analysis was an individual experimental solution (Tables 3 and 4).

Table 3. Group characteristics and initial equivalence testing.

Indicator	EXP (n = 30)	CTRL (n = 30)	Criterion	<i>p</i>
Age, M (SD)	19.3 (0.6)	19.5 (0.7)	$t(58) = -1.12$	0.268
Gender, f/m	21/9	20/10	$\chi^2(1) = 0.08$	0.781
Program 6B01504/6B01507	16/14	16/14	$\chi^2(1) = 0.00$	1.000
Grade Point Average (GPA), M (SD)	3.23 (0.31)	3.19 (0.34)	$t(58) = 0.48$	0.634
Previous Chemistry Score, M (SD)	79.8 (6.7)	78.9 (7.2)	$t(58) = 0.50$	0.620
Knowledge test before, M (SD)	16.7 (3.1)	16.3 (3.0)	$t(58) = 0.51$	0.614
Problem-solving test before, M (SD)	10.8 (2.6)	10.5 (2.5)	$t(58) = 0.46$	0.650

Table 4. Descriptive statistics of the main indicators before and after, M (SD).

Indicator (Range)	EXP Before	EXP After	CTRL Before	CTRL After
Knowledge test (0–30)	16.7 (3.1)	23.0 (3.0)	16.3 (3.0)	20.4 (3.2)
Problem-Solving Test, General (0–24)	10.8 (2.6)	18.5 (2.7)	10.5 (2.5)	14.9 (3.0)
Select method (0–4)	1.9 (0.8)	3.2 (0.7)	1.8 (0.8)	2.6 (0.8)
Sampling (0–4)	1.8 (0.7)	3.0 (0.7)	1.8 (0.7)	2.5 (0.8)
Calibration (0–4)	1.7 (0.8)	3.1 (0.7)	1.7 (0.8)	2.4 (0.8)
Quality control (0–4)	1.8 (0.8)	3.0 (0.8)	1.7 (0.8)	2.5 (0.8)
Troubleshooting (0–4)	1.6 (0.8)	3.0 (0.8)	1.6 (0.7)	2.3 (0.9)
Interpretation (0–4)	2.0 (0.8)	3.2 (0.7)	1.9 (0.8)	2.6 (0.8)
Final practical assignment (0–90)	-	76.4 (7.8)	-	68.2 (8.6)

Before the start of the eight-week module, no statistically significant differences were found between EXP and CTRL in terms of age, gender, program assignment, GPA, previous chemistry grade, and both entrance tests (all $p > 0.05$). Thus, the baseline characteristics of the groups were comparable, and the respective entrance grade and GPA were subsequently used as covariates in the analysis of covariance.

A fidelity check revealed a clear contrast between the learning conditions. In EXP, features of the project-based format were observed in an average of 90.6% of lessons, while in CTRL, they were observed in 29.7%; the most pronounced differences concerned open-ended problem setting, independent method selection, planning, and iterative revision of the experimental strategy (see Appendix A, Table A4).

The psychometric properties of the instruments were acceptable. Inter-rater reliability was good for both the problem-solving test, ICC(2,1) = 0.89, 95% CI [0.84, 0.93], and the final practical assignment, ICC(2,1) = 0.86, 95% CI [0.79, 0.91]. A total of 2317 solutions were recorded (1268 in EXP and 1049 in CTRL), with an average of 38.6 solutions per student; 636 events (27.4%) were independently recoded by a second expert (Table A5).

3.1. RQ1: Changes in Experimental Decision-Making Trajectories over Time

To address the first research question, changes in experimental decision-making trajectories were examined across the eight-week module. In the experimental group, the number of recorded choice points increased over time, accompanied by a progressive shift toward independently initiated and data-driven decisions and a reduction in direct reliance on instructor support (Figure 1). Weekly descriptive values are provided in Table A6.

In the experimental group, as the eight-week project progressed, not only did the number of recorded choice points increase but also the quality of the solutions changed:

the proportion of independently initiated and experimentally based actions increased, while direct dependence on the instructor decreased (Figure 1). In the control group, the dynamics of these same indicators were significantly weaker (weekly values are presented in Appendix A.1, Table A6).

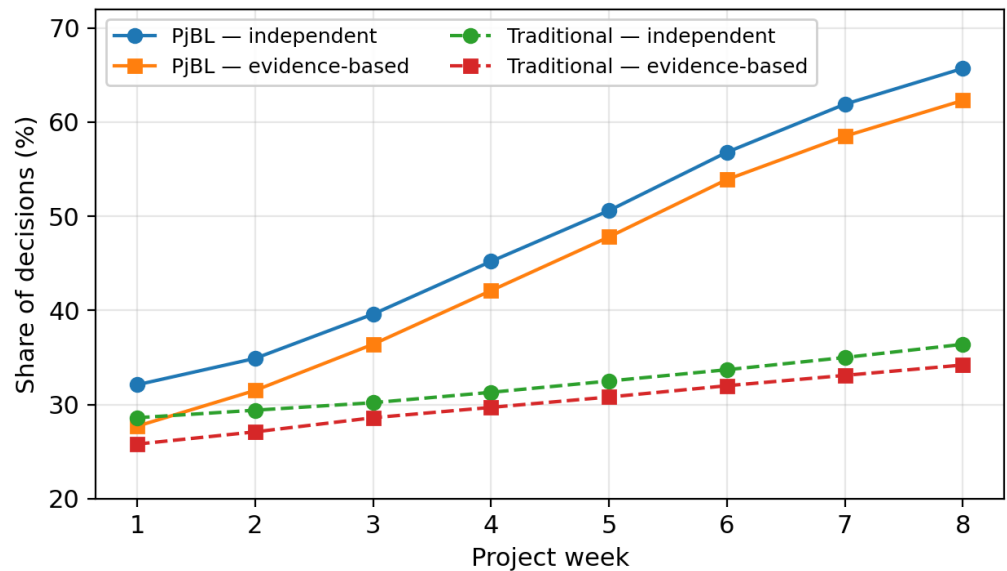


Figure 1. Dynamics of the proportion of independent and data-driven decisions by week for two groups.

In EXP, the proportion of independent decisions increased from 35.7% in the early period to 61.5% in the late period, while the proportion of decisions with a high level of data reliance and analytical justification increased from 32.0% to 58.2%.

After anomaly detection (D8), self-correction (D9) and procedural revision (D10) increased from early to late in EXP, while the proportion of D8 events requiring direct teacher assistance decreased; in CTRL, the changes were significantly weaker (Table A7). Teacher and Evidence were considered here as event attributes rather than separate states of the D-sequence (Figure 2).

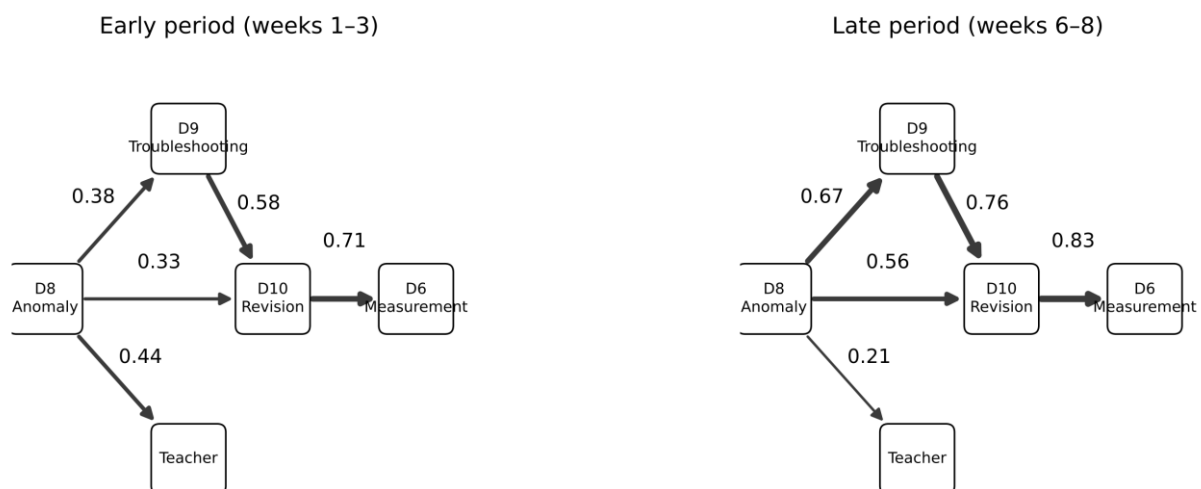


Figure 2. Changes in key transitions after an anomaly in EXP was detected.

3.2. RQ2: Differences Between Project-Based and Traditional Laboratory Conditions

To address the second research question, differences in the temporal development of independent and data-driven decisions between EXP and CTRL were examined using generalized linear mixed models (Table 5).

Table 5. Generalized linear mixed models at the individual decision level.

Outcome	Predictor	OR	95% CI	<i>p</i>
Independent decision	Week	1.04	1.00–1.08	0.038
	Group (EXP vs. CTRL)	1.09	0.78–1.52	0.620
	Week × Group	1.12	1.07–1.17	<0.001
Data-driven decision	Week	1.03	0.99–1.08	0.124
	Group (EXP vs. CTRL)	1.06	0.75–1.49	0.744
	Week × Group	1.13	1.08–1.18	<0.001

Logistic models included a random intercept for students to account for repeated decisions within a single participant. Week was centered on the first week; the primary interaction of interest was the week × group interaction. In CTRL, the dynamics were significantly weaker. The week × group interaction was significant for both independent decisions (OR = 1.12; 95% CI 1.07–1.17; $p < 0.001$) and data-driven decisions (OR = 1.13; 95% CI 1.08–1.18; $p < 0.001$). In EXP, there was a simultaneous decrease in unreasonable actions, rote repetition, and teacher-initiated solutions; the decrease in rote repetition was the most moderate. These results support H1–H3.

Between-group differences in post-intervention outcome measures were additionally assessed using ANCOVA while controlling for the corresponding baseline score and GPA (Table 6).

Table 6. Analysis of Covariance for Outcome Measures.

Dependent Variable	Adjusted M EXP	Adjusted M CTRL	F (df)	<i>p</i>	η^2p	<i>d</i>
Knowledge test	22.9	20.5	10.8 (1.56)	0.002	0.162	0.84
Problem-Solving Test, General	18.4	15.0	22.9 (1.56)	<0.001	0.290	1.26
Selecting a method	3.17	2.61	11.7 (1.56)	0.001	0.173	0.80
Sampling	3.01	2.56	6.2 (1.56)	0.016	0.100	0.67
Calibration	3.08	2.43	15.4 (1.56)	<0.001	0.216	0.93
Quality control	2.97	2.51	7.3 (1.56)	0.009	0.115	0.63
Troubleshooting	3.03	2.32	18.6 (1.56)	<0.001	0.249	0.82
Interpretation	3.17	2.62	10.9 (1.56)	0.002	0.163	0.80

Cohen's *d* values were recalculated from the unadjusted post-test means and pooled standard deviations of the two groups ($n = 30$ per group). ANCOVA inference is based on adjusted means and partial η^2 . On average, across the EXP module, students reported a higher proportion of independent and data-driven solutions, a greater frequency of adaptive revisions, and a lower reliance on direct instructor assistance (full descriptive measures are provided in Appendix A.1, Table A10). In the ANCOVA, the adjusted mean of the post-test problem-solving scores was 18.4 in EXP and 15.0 in CTRL, $F(1, 56) = 22.9$, $p < 0.001$, $\eta^2p = 0.290$. The difference in subject knowledge was also significant, but the effect size was smaller, consistent with a more specific advantage of the design condition for experimental problem solving.

Covariates are the corresponding entry score and GPA; for all models, $df = 1, 56$. Cohen's *d* is calculated on unadjusted post-test means and pooled SD; ANCOVA inference is based on adjusted means and η^2p .

3.3. RQ3: Trajectory Profiles Associated with Analytical Problem-Solving Success

To address the third research question, individual decision sequences were clustered according to their trajectory structure, resulting in four interpretable profiles (Table 7 and Figure 3).

Table 7. Trajectory typologies and their relationship with success.

Cluster	n	Key Features of the Trajectory	Problem Solving Score, M	Highly Successful, %
Cluster 1: Procedural Performers	15	Low autonomy and reliance on data; high frequency of direct teacher assistance	13.6	6.7
Cluster 2: Trial and Error	14	High action density, but frequent mechanical repetitions and weak justification	15.2	14.3
Cluster 3: Guided Analytical Reasoning	17	Good justification with continued reliance on instructor support	17.4	35.3
Cluster 4: Adaptive Analytics Solutions	14	High autonomy, data → assessment → revision; low teacher dependence	20.1	78.6

Clusters were identified using Ward’s method applied to the optimal-matching distance matrix derived exclusively from the D1–D12 sequences. High-success students were defined independently as those whose final analytical problem-solving test scores fell within the upper tertile of the total sample distribution. The success classification was not used in constructing the clusters. The “High-success students, %” column therefore indicates the proportion of independently defined high-success students represented within each trajectory profile (Figure 3). Candidate solutions containing two to six clusters were compared quantitatively using average silhouette width (ASW). The ASW values were 0.31 for $k = 2$, 0.39 for $k = 3$, 0.46 for $k = 4$, 0.38 for $k = 5$, and 0.34 for $k = 6$. The four-cluster solution showed the highest ASW and also yielded balanced cluster sizes ($n = 15, 14, 17$, and 14) and substantively distinct trajectory profiles. Solutions with five or six clusters produced smaller, less interpretable subdivisions without improving partition quality (Table A11).

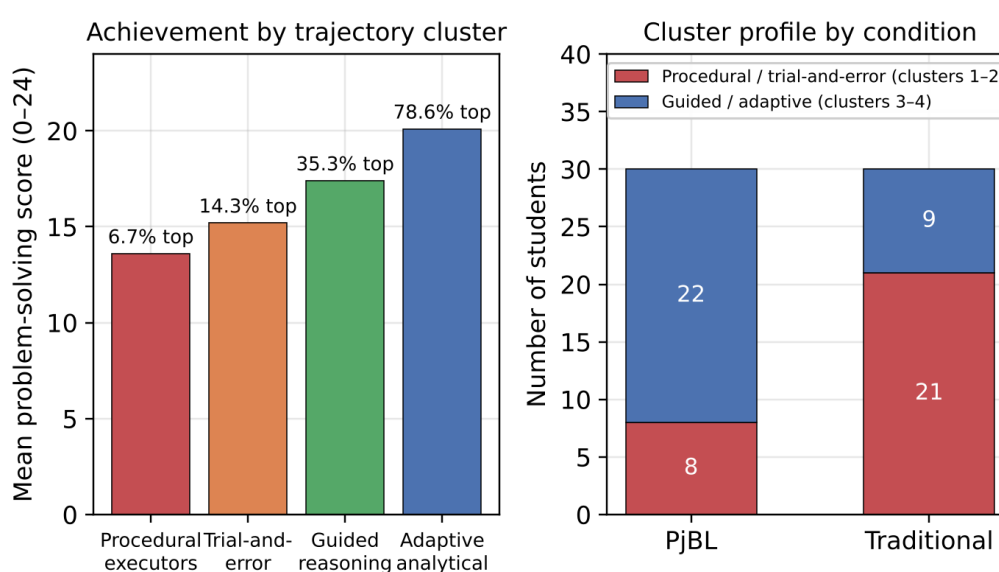


Figure 3. Trajectory cluster profiles, their relationship with success, and distribution across learning conditions.

More successful students more often demonstrated adaptive cycles involving problem detection, diagnosis, and revision, and were less likely to limit themselves to mechanical repetition or direct following of prompts (process patterns are presented in the Appendix A.1, Table A8). In EXP, the probability of a D8→D9 transition increased from 0.38 to 0.67, and the proportion of D8 events involving a request to the instructor decreased from 0.44 to 0.21. Clustering identified four interpretable profiles; EXP was more frequently represented in the two more autonomous profiles (22 out of 30 students), and CTRL in the two more dependent profiles (21 out of 30), $\chi^2(3) = 12.58$, $p = 0.006$. Using the independently defined performance criterion, 78.6% of students in the Adaptive Analytical Solutions profile belonged to the upper tertile of the final problem-solving score distribution, compared with 6.7%, 14.3%, and 35.3% in the Procedural Performers, Trial-and-Error, and Guided Analytical Reasoning profiles, respectively. Because performance classification was defined independently of the trajectory variables and was not used for cluster construction, this pattern provides evidence consistent with H4. However, the association does not establish that trajectory profile determines subsequent performance.

3.4. RQ4: Associations Between Trajectory Characteristics and Learning and Analytical Outcomes

To address the fourth research question, trajectory characteristics were examined in relation to students' learning and analytical outcomes. The prespecified relationships involved the final problem-solving test and the relative error of total iron, whereas relationships with the final practical assignment and recovery were treated as exploratory.

For H5, the prespecified relationships were correlations of trajectory characteristics with the final problem-solving test and the relative error of total iron; additional relationships with the final practical task and recovery were treated as exploratory (Figure 4). The median relative error was 5.6% (IQR 4.2–7.8) in EXP and 7.1% (IQR 5.0–9.6) in CTRL, $U = 307$, $p = 0.034$. The mean recovery values were 97.1% (SD = 5.0) and 94.7% (SD = 6.2), respectively; the difference did not reach statistical significance ($p = 0.101$).

Figure 4 summarizes the Spearman rank correlations between trajectory indicators and the principal learning and analytical outcomes.

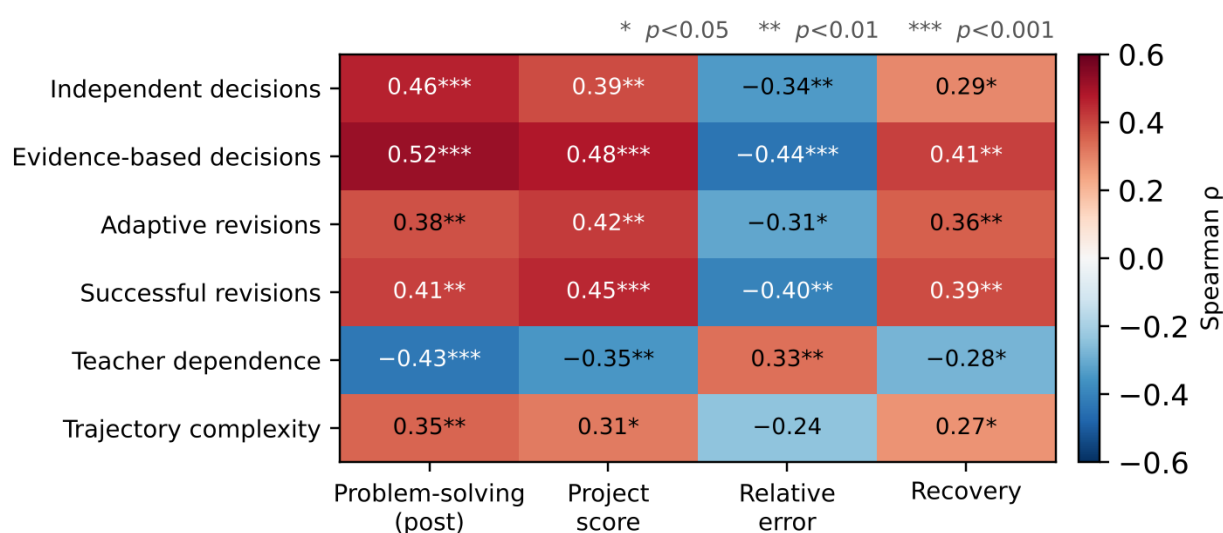


Figure 4. Rank correlations of trajectory indicators with learning outcomes and quality of analysis (according to Spearman).

The strongest relationship was observed between the proportion of data-driven solutions and the final problem-solving score ($\rho = 0.52$, $p < 0.001$); the same indicator was

negatively correlated with the relative error of iron determination ($\rho = -0.44$, $p < 0.001$). The relationship with the instructor was in the opposite direction: $\rho = -0.43$ with the final score and $\rho = 0.33$ with the relative error. The full exploratory matrix is provided in the Appendix A (Table A9). The resulting coefficients indicate a moderate, rather than deterministic, relationship between process trajectory maturity and analytical task performance, supporting H5.

Reflexive notes were used as a qualitative illustration of the observed process changes. In the late EXP period, students more often cited the calibration range, consistency of replicates, and control measurements as reasons for procedural changes; An unexpected result often prompted an investigation into a possible cause rather than automatically repeating the measurement. The recordings also noted a change in the function of teacher support: instead of requesting a predefined action, students more often formulated their own and used the teacher's question to check the logic. These observations are consistent with the quantitative dynamics of H1–H4, but are used for illustrative purposes only.

4. Discussion

4.1. RQ1: Development of Experimental Decision-Making Trajectories over Time

The first research question concerned how students' experimental decision-making trajectories changed over the course of the eight-week module. The main finding of the study is that differences between conditions were evident not only in the final scores but also in the structure of the solution process. In the PjBL group, the likelihood of independent and data-driven solutions increased over the course of eight weeks, while in the control condition these changes were significantly weaker. The temporal pattern is important because it suggests that the development of experimental reasoning was not limited to an increase in the number of decisions made by students. The nature of these decisions also changed. Students increasingly moved from externally prompted actions toward decisions based on their own observations and experimental evidence. At the same time, mechanical repetition declined more slowly than other dependent behaviors. This suggests that greater autonomy does not automatically eliminate less productive laboratory strategies. A student may become more willing to make an independent decision while still relying on repetition rather than diagnosis when an unexpected result occurs. From an instructional perspective, autonomy and diagnostic reasoning therefore appear to require partly distinct forms of support and feedback. This interpretation is consistent with previous work emphasizing that authentic laboratory learning requires students not only to perform procedures but also to make decisions about experimental design, evidence, and interpretation (Zammit et al., 2024; Seery et al., 2024).

4.2. RQ2: Differences Between Project-Based and Traditional Laboratory Conditions

The second research question focused on differences between the project-based and traditional laboratory conditions.

This pattern is consistent with the current literature on laboratory learning, which associates authenticity with students' ability to make decisions about procedure, analysis, and interpretation, rather than simply following a prescribed algorithm (Zammit et al., 2024; Seery et al., 2024). In analytical chemistry, projects and CURE approaches have already shown potential for engaging students in planning and independent research activities (Matilainen et al., 2021; Muna, 2021). Our data complement these results at the process level: differences are evident in the solutions that follow a problem and what the student relies on when reviewing.

The final results also indicate that the difference is more pronounced in analytical problem solving than in subject knowledge. This is important for interpreting the effect: PjBL does not replace course content, but rather creates more opportunities to apply it in method selection, calibration, diagnosis, and interpretation. The larger difference in analytical problem solving than in subject knowledge is theoretically meaningful. The two conditions addressed the same core analytical content, but differed in the extent to which students were required to select, justify, evaluate, and revise experimental actions. Therefore, the observed pattern is consistent with the idea that PjBL may primarily influence the application and coordination of knowledge rather than factual or conceptual knowledge alone. In this sense, the intervention appears more closely aligned with competencies required for method selection, calibration, troubleshooting, and interpretation than with the acquisition of disciplinary content itself. Because the design was quasi-experimental and each condition corresponded to an established learning group, it is more accurate to interpret this pattern as a stable association with the project condition rather than as an isolated causal effect of PjBL.

This emphasis is consistent with contemporary discussions of the analytical-chemical curriculum, where, alongside fundamental knowledge, research and professional practices, data handling, and decision making are increasingly important (Van Wyk et al., 2025; Kovarik, 2026).

4.3. RQ3: Trajectory Profiles Associated with Analytical Problem-Solving Success

The third research question examined which types of decision trajectories characterized students with different levels of analytical problem-solving success. Clustering revealed that higher performance was more frequently associated with trajectories characterized by adaptive backtracking, evidence-based revision, and lower dependence on direct teacher support.

An important observation is that trajectory quality cannot be reduced to the sheer number of actions performed. The “Trial and Error” profile was characterized by relatively high action density, yet its problem-solving performance remained lower than that of the more adaptive profiles. This suggests that frequent activity by itself should not be interpreted as productive inquiry. What appears to matter more is the organization of actions into diagnostically meaningful cycles in which an unexpected result is followed by problem identification, evaluation of evidence, and a justified procedural revision. The intermediate position of the “Guided Analytical Reasoning” profile is also informative. These students demonstrated relatively strong justification but still relied on external support. This pattern suggests that the development of experimental competence may not follow a simple transition from dependence to independence; students may first learn to reason analytically with scaffolding before they can consistently regulate the same process autonomously.

Nevertheless, cluster membership should not be interpreted as a determinant of achievement. The profiles are descriptive patterns that summarize recurring forms of decision organization and their association with performance.

4.4. RQ4: Associations Between Trajectory Characteristics and Learning and Analytical Outcomes

The fourth research question addressed the relationship between trajectory characteristics and learning and analytical outcomes. The strongest association was observed between the proportion of data-driven decisions and analytical problem-solving performance, while greater reliance on data was also associated with lower relative error in the standardized iron determination task.

These relationships are important because they connect a process-oriented educational construct with an external indicator of analytical performance. In other words, the

EDMP indicators were not related only to another educational test; they were also associated with the quality of an actual analytical measurement. At the same time, the correlations were moderate rather than strong enough to imply that decision-making trajectory alone determines analytical accuracy. Measurement quality also depends on technical execution, instrumentation, sample preparation, calibration quality, and other factors that were not represented by the trajectory indicators. The absence of a statistically significant difference in recovery is equally informative. It indicates that not all analytical-quality indicators responded in the same way to the instructional condition. This reduces the risk of interpreting the EDMP framework as a universal predictor of laboratory performance and instead suggests that its strongest contribution lies in explaining selected aspects of experimental reasoning and decision quality.

Taken together, the findings clarify the theoretical and methodological contribution of the EDMP framework. The theoretical contribution of EDMP should therefore be understood as integrative rather than substitutive. EDMP does not propose a new general theory of scientific reasoning; instead, it provides a domain-specific operational bridge between scientific reasoning, metacognitive regulation, inquiry activity, and process-data analysis in the context of analytical chemistry. Its distinctive feature is the simultaneous representation of the temporal structure of experimental decisions and the epistemic attributes of each decision, allowing laboratory reasoning to be analysed as an evolving trajectory rather than inferred solely from final performance.

From a practical perspective, process recording can be used not as an additional “assessment for each action,” but as a diagnostic tool. It is more important for the instructor to see at what stage the student regularly seeks direct instruction, skips data checks, or changes several parameters at once without a hypothesis. A slower decline in mechanical repetitions is particularly indicative: this strategy can persist even when overall independence is already increasing and therefore requires separate feedback.

The study has several limitations. The conditions were assigned to two established study groups, so the effect of a specific study section cannot be completely separated from the learning format, despite the initial comparability. The sample is limited to one university, two related educational programs, and one subject-matter task. Process events were partly based on student decision logs and therefore depend on the completeness of self-recording, although context was verified by the observation protocol. Finally, eight weeks do not allow for assessing the stability of trajectories and the transfer of skills to a new analytical tool or other instrumental method. These limitations require caution with causal and generalizing formulations. Further research should be conducted at multiple sites and, if possible, with multiple training sections per condition. Transfer to new analytical tasks, delayed measurement, and more automated process data recording are also important. Testing the reproducibility of the proposed coding scheme by other research groups and comparing manual coding with digital laboratory activity logs deserve special attention.

The descriptive implementation check was completed by a single independent observer, and the eight binary instructional-format characteristics were not independently double-coded; therefore, inter-rater reliability could not be established for this measure. The corresponding percentages should be interpreted as descriptive indicators of implementation rather than as a validated fidelity score.

5. Conclusions

This study examined process-oriented assessment of project-based learning in analytical chemistry through the concept of experimental decision-making pathways (EDMP). Rather than evaluating laboratory learning only through final scores, the

proposed approach made it possible to reconstruct how students selected, justified, evaluated, and revised experimental actions over time.

With respect to the first research question, students in the project-based condition demonstrated a progressive shift toward more independent and data-driven decision making across the eight-week module. The development was also reflected in students' responses to anomalous results: diagnostic troubleshooting and reasoned procedural revision became more frequent, whereas reliance on direct instructor support decreased. At the same time, mechanical repetition declined more slowly, indicating that increasing autonomy does not necessarily imply an equally rapid development of diagnostic reasoning.

Regarding the second research question, the experimental and control conditions differed not only in final learning outcomes but also in the development of decision-making processes. The project-based group showed a stronger increase in independent and evidence-based decisions than the traditional laboratory group. Differences were particularly pronounced for analytical problem solving, calibration, and troubleshooting, whereas the difference in subject knowledge was comparatively smaller. Given the quasi-experimental design and the use of two established study groups, these findings should be interpreted as associations with the instructional condition rather than as isolated causal effects of project-based learning.

For the third research question, sequence analysis identified qualitatively different trajectory profiles. Higher analytical problem-solving performance was more frequently associated with adaptive trajectories characterized by problem detection, evidence evaluation, justified revision, and lower dependence on direct teacher support. By contrast, frequent experimental activity alone did not necessarily indicate stronger reasoning, as the trial-and-error profile combined relatively high action density with weaker justification and lower performance. These findings suggest that the organization and quality of experimental decisions may be more informative than the number of actions performed.

With respect to the fourth research question, trajectory characteristics were meaningfully associated with learning and analytical outcomes. In particular, greater reliance on experimental data was positively related to final problem-solving performance and negatively related to relative error in total iron determination. These relationships were moderate rather than deterministic, indicating that experimental decision making represents one component of analytical performance alongside technical execution, calibration quality, sample preparation, and other laboratory factors. The absence of a statistically significant between-group difference in recovery further shows that not all indicators of analytical quality are equally sensitive to differences in instructional format.

From an educational perspective, the findings support the use of process-oriented assessment as a complement to conventional laboratory reports, tests, and final analytical results. Decision logs and trajectory indicators can help instructors identify where students continue to depend on direct guidance, repeat measurements without diagnosis, fail to use available evidence, or revise procedures without a clear rationale. Such information may support more targeted formative feedback and allow laboratory instruction to focus not only on whether students obtain a correct result but also on how they arrive at that result. In this sense, EDMP can serve as a diagnostic framework for monitoring the development of experimental reasoning without reducing performance to a single composite score.

Further research should examine the reproducibility and transferability of the EDMP framework across different institutions, analytical tasks, and instrumental methods. Studies involving multiple teaching groups within each condition would help separate instructional effects from class-level effects and provide stronger evidence regarding the influence of project-based learning. Longer-term designs are also needed to determine whether changes in decision-making trajectories persist beyond a single course and transfer to

unfamiliar analytical problems. Future work may additionally compare manually coded decision logs with digital laboratory traces or automated process-data collection and examine whether trajectory-based feedback can itself improve students' experimental reasoning.

Author Contributions: Conceptualization, G.M. and T.Z.; methodology, G.M.; software, T.Z.; validation, G.M. and T.Z.; formal analysis, G.M.; investigation, T.Z.; resources, G.M.; data curation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, G.M.; visualization, T.Z.; supervision, G.M.; project administration, T.Z.; funding acquisition, T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Zhanibekov University (protocol code No. 12 and date of approval 5 February 2026).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy and ethical considerations related to the participation of university students.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAS	Atomic Absorption Spectrometry
ANCOVA	Analysis of Covariance
CI	Confidence Interval
CTRL	Control Group
CURE	Course-Based Undergraduate Research Experience
EDMP	Experimental Decision-Making Pathway
EXP	Experimental Group
GPA	Grade Point Average
ICC	Intraclass Correlation Coefficient
I-CVI	Item-Level Content Validity Index
IQR	Interquartile Range
KR-20	Kuder–Richardson Formula 20
OR	Odds Ratio
PjBL	Project-Based Learning
S-CVI/Ave	Scale-Level Content Validity Index, Average Method
SD	Standard Deviation

Appendix A

Appendix A.1

The appendix contains a coding scheme and supporting tables that complement the figures in the main text.

Table A1. Coding scheme for experimental decision types (D1–D12).

Code	Solution Type	Brief Description
D1	Problem Recognition	The student recognizes the analytical problem
D2	Sample Collection	Selecting the sampling method, sample volume, and number

D3	Method Selection	Selecting an analytical method
D4	Sample Preparation	Dilution, filtration, mineralization, etc.
D5	Calibration Strategy	Standards, concentration range, blank sample, calibration curve
D6	Measurement Conditions	Selecting analytical conditions
D7	Quality Control	Replicates, blank and control samples, recovery
D8	Anomaly Detection	Recognizing a suspicious result
D9	Troubleshooting	Determining the possible cause of the problem
D10	Procedure Review	Modifying the experimental procedure
D11	Data Interpretation	Interpreting the obtained results
D12	Final Conclusion	Formulating the final analytical conclusion

Table A2. Project stages at which decisions were recorded.

Code	Project Stage
PS1	Problem Definition
PS2	Method Selection
PS3	Experimental Design
PS4	Sample Preparation
PS5	Calibration
PS6	Measurement
PS7	Quality Control
PS8	Troubleshooting
PS9	Data Interpretation
PS10	Final Analytical Judgment

Table A3. Ordinal decision coding variables.

Variable	Range	Contents of Levels
Autonomy	0–3	0—direct instruction from the teacher; 1—solution after direct instruction; 2—solution after a leading question; 3—completely independent solution
Justification	0–3	0—no justification; 1—intuitive or weak; 2—partially scientific; 3—fully analytically justified
Evidence	0–3	0—no evidence; 1—reliance on subjective opinion; 2—one relevant observation; 3—several interrelated pieces of evidence
Teacher support	0–3	0—no help; 1—general question; 2—prompt; 3—direct instruction
Peer	0–3	0—individual solution; 1—shared idea of a colleague; 2—joint development; 3—solution proposed by another student
Revision	0–3	0—no revision; 1—change after teacher’s instruction; 2—change after independent problem discovery; 3—independent revision based on data
Accuracy	0–2	0—incorrect solution; 1—partially acceptable; 2—analytically correct
Alternatives	0–2	0—no alternatives were considered; 1—one alternative; 2—two or more alternatives
Outcome	0–2	0—problem not solved; 1—partially successful; 2—successful

Table A4. Descriptive occurrence of instructional-format characteristics across observed laboratory sessions (%).

Format Feature	EXP	CTRL
Open-ended Task	100.0	0.0
Independent Planning	87.5	12.5
Student Choice of Methods	87.5	0.0
Independent Decision Making	87.5	25.0
Iterative Revision	75.0	12.5
Teamwork	100.0	75.0
Final Product	100.0	100.0
Reflection	87.5	12.5
Mean proportion across the eight characteristics	90.6	29.7

Table A5. Coding reliability and content validity indicators.

Indicator	Meaning
Cohen's κ for solution type	0.88
Cohen's κ for the project stage	0.87
Weighted κ , independence scale	0.86
Weighted κ , validity scale	0.84
Weighted κ , evidence scale	0.85
I-CVI, range by category	0.80–1.00
S-CVI/Ave	0.93
Modified kappa (κ^*)	0.76–1.00

Table A6. Dynamics of trajectory indicators by week.

Week	Solution Density	Autonomy, %	Based on Data, %	Revisions	Teacher Support, %	Diversity of Types
EXP (n = 30; total of 1268 solutions)						
1	4.6	32.1	27.7	0.7	43.8	5.1
2	5.0	34.9	31.5	0.9	40.7	5.6
3	5.1	39.6	36.4	1.1	36.2	6.1
4	5.3	45.2	42.1	1.4	31.9	6.7
5	5.4	50.6	47.8	1.6	28.4	7.1
6	5.6	56.8	53.9	1.9	24.3	7.5
7	5.7	61.9	58.5	2.0	21.2	7.9
8	5.6	65.7	62.3	2.2	19.6	8.2
CTRL (n = 30; total of 1049 solutions)						
1	4.2	28.6	25.8	0.6	48.2	4.8
2	4.3	29.4	27.1	0.7	47.5	4.9
3	4.3	30.2	28.6	0.7	46.0	5.0
4	4.4	31.3	29.7	0.8	45.1	5.2
5	4.4	32.5	30.8	0.8	44.6	5.2
6	4.4	33.7	32.0	0.9	43.5	5.3
7	4.5	35.0	33.1	0.9	42.4	5.5
8	4.5	36.4	34.2	1.0	41.1	5.6

Table A7. Probabilities of key transitions after anomaly detection (D8).

Transition	EXP Early	EXP Late	CTRL Early	CTRL Late
P(D9 D8)—self-troubleshooting	0.38	0.67	0.34	0.40
P(Teacher support D8)—the proportion of D8 events with an appeal to the teacher	0.44	0.21	0.47	0.41
P(D10 D8)—procedure revision	0.33	0.56	0.29	0.34
P(Evidence-based D10 D8)—the proportion of revisions based on data after D8	0.31	0.61	0.28	0.36

Table A8. Most common process patterns.

Pattern	EXP Frequency	CTRL Frequency
D5 → D6 → D11 (calibration → measurement → interpretation)	96	81
D6 → D8 → D9 (measurement → anomaly → diagnosis)	87	39
D8 → D9 → D10 → D6 (adaptive cycle)	74	28
D8 → D10 with direct teacher support	31	67
D6 → D8 → D6 (mechanical repetition)	26	59

Table A9. Correlations of trajectory indicators with outcomes (Spearman ρ).

Trajectory Indicator	Problem-Solving Test (Post)	Final Practical Assignment	Relative Error	Recovery
Share of independent decisions	0.46 ***	0.39 **	−0.34 **	0.29 *
Share of decisions based on data	0.52 ***	0.48 ***	−0.44 ***	0.41 **
Frequency of adaptive revisions	0.38 **	0.42 **	−0.31 *	0.36 **
Proportion of successful revisions	0.41 **	0.45 ***	−0.40 **	0.39 **
Dependency on the teacher	−0.43 ***	−0.35 **	0.33 **	−0.28 *
Trajectory complexity	0.35 **	0.31 *	−0.24	0.27 *

Note: Additional coefficients in the table are presented as exploratory and are reported with unadjusted p values: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. The main pre-specified H5 relationships are described in the main text.

Table A10. Comparison of aggregated trajectory indicators between groups.

Solution Category	Week 1–3, %	Week 6–8, %	Change	p	
Unreasonable Actions	29.8	14.9	−14.9	<0.001	
Mechanical Repetitions	17.1	10.2	−6.9	0.014	
Teacher-Initiated Solutions	31.4	16.7	−14.7	<0.001	
Trajectory indicator	EXP, M (SD)	CTRL, M (SD)	Criterion	p	d
Percentage of independent solutions, %	49.1 (14.8)	32.2 (13.1)	$t(58) = 4.68$	<0.001	1.21
Percentage of data-based solutions, %	45.8 (14.5)	30.2 (13.2)	$t(58) = 4.36$	<0.001	1.13
Frequency of adaptive revisions	11.6 (4.3)	8.2 (4.0)	$t(58) = 3.17$	0.002	0.82
Instructor dependence, %	27.8 (13.7)	44.9 (14.2)	$t(58) = −4.75$	<0.001	−1.23
Independent troubleshooting, %	55.4 (16.8)	39.1 (17.4)	$t(58) = 3.69$	<0.001	0.95
Trajectory complexity (0–1)	0.64 (0.13)	0.53 (0.12)	$t(58) = 3.41$	0.001	0.88

Table A11. Comparison of candidate cluster solutions.

Number of Clusters	Average Silhouette Width (ASW)	Smallest Cluster, n	Interpretation
2	0.31	26	Broad separation; distinct intermediate profiles merged
3	0.39	17	Improved separation but adaptive and guided patterns partly merged
4	0.46	14	Best balance of separation, cluster size, and interpretability

5	0.38	7	Additional split without clear substantive distinction
6	0.34	5	Increased fragmentation and small clusters

Appendix A.2. Analytical Problem-Solving Test and Scoring Rubric

Task 1. Method Selection

A water sample is expected to contain approximately 0.20–0.50 mg/L total iron. The sample also contains several dissolved salts and has a slightly yellow color. You have access to spectrophotometric determination with 1,10-phenanthroline and flame atomic absorption spectrometry. Which analytical method would you select for this sample? Justify your choice by considering sensitivity, possible interferences, sample preparation, and reliability of the result.

Task 2. Sampling

You are asked to assess total iron in drinking water supplied through a university building. Describe how you would organize the sampling procedure, including the number and location of sampling points, container preparation, sample preservation, and measures required to obtain a representative sample.

Task 3. Calibration

The expected iron concentration in a sample is approximately 0.6 mg/L. Available standard solutions are 0.0, 0.1, 0.25, 0.50, 0.75, 1.0, and 2.0 mg/L. Select an appropriate calibration strategy and explain which standards you would use, whether a blank is required, and what you would do if the sample response falls outside the calibration range.

Task 4. Quality Control

During a series of iron determinations, all sample measurements appear internally consistent. However, no blank, duplicate, or recovery test has been performed. Which quality-control procedures should be introduced before accepting the analytical results? Explain what information each selected procedure provides.

Task 5. Troubleshooting

Three replicate measurements of the same water sample give iron concentrations of 0.43, 0.44, and 0.68 mg/L. Before repeating the entire analysis, what possible causes should be considered and what steps would you take to identify the source of the anomalous result?

Task 6. Data Interpretation

An iron analysis produced a calibration coefficient $R^2 = 0.998$, replicate concentrations of 0.47, 0.48, and 0.49 mg/L, and a spike recovery of 97%. The reference concentration is 0.50 mg/L. Evaluate the quality of the analytical result and formulate an appropriate analytical conclusion, indicating which evidence supports your interpretation.

Each of the six tasks was scored from 0 to 4 points, yielding a maximum total score of 24. The same scoring framework was applied across all six domains.

References

- Agustian, H. Y. (2025). Recent advances in laboratory education research. *Chemistry Teacher International*, 7(2), 217–224. <https://doi.org/10.1515/cti-2024-0071>.
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3–4), 369–398. <https://doi.org/10.1080/00461520.1991.9653139>.

- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. <https://doi.org/10.1002/sce.10001>.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cooper, M. M., & Sandi-Urena, S. (2009). Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving. *Journal of Chemical Education*, 86(2), 240–245. <https://doi.org/10.1021/ed086p240>.
- Domin, D. S. (1999). A review of laboratory instruction styles. *Journal of Chemical Education*, 76(4), 543–547. <https://doi.org/10.1021/ed076p543>.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37. <https://doi.org/10.18637/jss.v040.i04>.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54. <https://doi.org/10.1002/sce.10106>.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>. (Erratum in 2017, *Journal of Chiropractic Medicine*, 16(4), 346).
- Kovarik, M. L. (2026). The undergraduate analytical chemistry curriculum: How do we decide what to teach? *Analytical and Bioanalytical Chemistry*, 418, 9–14. <https://doi.org/10.1007/s00216-025-06191-6>.
- Kovarik, M. L., Robinson, J. K., & Wenzel, T. J. (2022). A new resource to help instructors incorporate active learning into analytical chemistry courses. *Analytical and Bioanalytical Chemistry*, 414, 4013–4017.
- Krajcik, J. S., & Blumenfeld, P. C. (2006). Project-based learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 317–334). Cambridge University Press.
- Matilainen, R., Nuora, P., & Valto, P. (2021). Student experiences of project-based learning in an analytical chemistry laboratory course in higher education. *Chemistry Teacher International*, 3(3), 229–238. <https://doi.org/10.1515/cti-2020-0032>.
- Muna, G. W. (2021). Stimulating students' learning in analytical chemistry through an environmental-based CURE project. *Journal of Chemical Education*, 98(4), 1221–1226. <https://doi.org/10.1021/acs.jchemed.0c01326>.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>.
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>.
- Sandi-Urena, S., Cooper, M. M., & Stevens, R. H. (2011). Enhancement of metacognition use and awareness by means of a collaborative intervention. *International Journal of Science Education*, 33(3), 323–340. <https://doi.org/10.1080/09500690903452922>.
- Seery, M. K., Agustian, H. Y., Christiansen, F. V., Gammelgaard, B., & Malm, R. H. (2024). 10 guiding principles for learning in the laboratory. *Chemistry Education Research and Practice*, 25(2), 383–402. <https://doi.org/10.1039/D3RP00245D>.
- Thomas, J. W. (2000). *A review of research on project-based learning*. Autodesk Foundation.
- Van Wyk, A. L., Frederick, K. A., Lieberman, M., & Cole, R. S. (2025). Increasing authenticity of the laboratory through the MICRO project: Analysis of analytical chemistry laboratory experiments for their level of inquiry. *Journal of Chemical Education*, 102(1), 3–14. <https://doi.org/10.1021/acs.jchemed.3c00945>.
- Zammit, K. M., Connor, M. C., & Raker, J. R. (2024). Evaluating the level of inquiry in postsecondary instructional laboratory experiments: Results of a national survey. *Chemistry Education Research and Practice*, 25, 79–91. <https://doi.org/10.1039/D3RP00154G>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.