Education Insights

Article

Design of Rasch-Based Evaluation Scale for Chemistry Experimental Teaching Literacy among Pre-Service Teachers

Yabin Wang [1] and Yuzhen Li [1,*]

[1] College of Chemistry and Environmental Sciences, Inner Mongolia Normal University, Hohhot, Inner Mongolia, 010022, China
* Correspondence: Yuzhen Li, College of Chemistry and Environmental Sciences, Inner Mongolia Normal University, Hohhot, Inner Mongolia, 010022, China

Abstract: Experimental teaching competence constitutes a fundamental component of professional literacy for pre-service chemistry teachers, reflecting their ability to integrate theoretical knowledge with practical laboratory skills, organize and present experimental procedures effectively, and adapt their instructional strategies to diverse learning situations. Systematically assessing the developmental level of this competence is critical for guiding teacher preparation programs, informing curriculum design, and promoting the overall quality of chemistry education. Based on the Rasch measurement model, this study developed a comprehensive evaluation scale tailored to pre-service chemistry teachers' experimental teaching literacy. By deconstructing the workflow of experimental teaching tasks, a four-dimensional framework was established, encompassing "knowledge understanding, skill execution, expression and organization, and thinking adjustment," with twelve initial indicators. These indicators were translated into 75 test items through a rigorous development process involving iterative expert review and multiple revisions. The scale was administered to 244 pre-service chemistry teachers from two universities, and Rasch modeling analysis was performed using Winsteps software. Items were evaluated and refined based on fit statistics, discrimination indices, and other relevant metrics, resulting in a final scale containing 56 well-functioning items. Empirical analysis indicates that the finalized scale demonstrates strong internal consistency, robust construct validity, and the capacity to provide multidimensional and precise assessment of experimental teaching literacy. By offering a reliable and empirically validated tool, this study supports both research and practical evaluation in chemistry teacher education, enabling educators and program designers to identify strengths and areas for improvement, and to tailor instructional interventions that enhance the development of high-quality experimental teaching competencies among pre-service teachers.

Keywords: chemistry experimental teaching literacy; pre-service teachers; Rasch model

Published: 27 December 2025

1. Introduction

Chemical experiments form the foundation of secondary school chemistry, serving as an essential means for students to engage with scientific phenomena and develop inquiry skills, while experiment-based teaching constitutes a defining characteristic of the discipline [1]. The General High School Chemistry Curriculum Standards (2017 Edition, Revised in 2020) emphasize that teachers should fully recognize the unique educational value of chemistry experiments and design experimental inquiry activities thoughtfully and systematically [2]. To achieve these goals, teachers must cultivate the dispositions, values, and practical abilities associated with experimental teaching, collectively referred to as experimental teaching literacy. At the level of teacher education, the Accreditation Standards for Chemistry Teacher Education Programs identify experimental teaching ability as a core graduation competency for pre-service teachers, underscoring its fundamental professional significance. In practical terms, this implies that contemporary chemistry teachers are expected not only to perform experiments skillfully, but also to integrate theoretical knowledge, technical skills, instructional design strategies, and scientific reasoning into a coherent and dynamic form of experimental teaching literacy.

Despite its recognized importance, the academic literature still lacks a unified and precise definition of chemistry teachers' experimental teaching literacy, and its internal components remain insufficiently articulated [3]. Existing research often emphasizes theoretical discussion while providing limited empirical validation, which hinders a comprehensive understanding of the current levels and developmental trajectories of pre-service teachers' experimental teaching competence. Without a systematic and reliable measurement framework, it is difficult to identify specific areas for improvement or to inform targeted interventions in teacher training programs.

Addressing these challenges requires the development of an evaluation instrument that is structurally transparent, psychometrically rigorous, and capable of generating detailed diagnostic information on multiple dimensions of competence. The Rasch measurement model, a cornerstone of modern educational assessment theory, offers notable advantages for such purposes. Its properties of parameter invariance and linear scale construction allow for the objective measurement of both examinee abilities and item difficulties on a shared metric [4]. Furthermore, the Rasch model provides methodological rigor in evaluating scale unidimensionality, optimizing item functioning, and enhancing measurement precision, making it particularly suitable for assessing complex, multidimensional constructs such as experimental teaching literacy.

Consequently, this study employs the Rasch model to develop a systematic and empirically validated evaluation scale for assessing experimental teaching literacy among chemistry pre-service teachers. By doing so, it aims to provide robust empirical evidence and theoretical support for understanding pre-service teachers' competency levels, informing curriculum design, and guiding professional development initiatives that foster high-quality experimental teaching practices. This approach not only addresses the current methodological gap in measuring experimental teaching literacy, but also offers a structured pathway for supporting the growth of future chemistry educators.

2. Definition of Concepts

2.1. Clarifying the Connotation of Chemistry Experiment Teaching Literacy

Research on experiment teaching literacy has primarily concentrated on conceptualizing the notion itself and exploring strategies for its cultivation. However, due to the inherent complexity of this concept, a universally accepted definition has yet to be established. Drawing upon relevant literature, chemistry experiment teaching literacy can be understood as a core dimension of teachers' professional competence. It encompasses a deep understanding of the pedagogical value of experiment-based instruction and the integrated application of multiple domains of expertise, including disciplinary knowledge, operational skills, instructional design and implementation capabilities, classroom management, and reflective improvement strategies. This literacy manifests in both real and simulated teaching environments, with the overarching aim of successfully completing experimental tasks and achieving instructional objectives [5-9].

Specifically, chemistry experiment teaching literacy spans the entire lifecycle of experiment-based teaching. Prior to conducting experiments, it involves analyzing underlying scientific principles, designing experimental procedures, and planning instructional strategies. During the experimental process, it requires the ability to perform standardized operations, guide student observations, manage laboratory safety, and facilitate active engagement. After experiments, it includes conducting effective explanations, evaluating experimental outcomes, optimizing instructional approaches, and fostering students' scientific reasoning. In addition, this literacy entails the judicious use of modern educational technologies, such as information and communication tools, to enhance teaching effectiveness, promote interactive learning, and support reflective practice. Altogether, chemistry experiment teaching literacy represents a holistic, multidimensional capability that integrates knowledge, skills, pedagogy, and technological proficiency in the service of high-quality chemistry education.

2.2. Basic Principles of the Rasch Model

The Rasch model is rooted in latent trait measurement theory and provides a rigorous framework for educational assessment. Its central concept is to map both the test-taker's ability and the item's difficulty onto a common linear scale, thereby estimating the probability that an individual will correctly respond to a given item. Mathematically, the model employs a logistic probability function, which ensures that the measurement process is independent of specific sample characteristics and allows for the comparability of results across different populations [10].

In the context of scale development, the Rasch model offers several methodological advantages. It enables simultaneous estimation of both examinee abilities and item difficulties, and it evaluates the consistency of each item with the underlying latent trait through fit statistics. Compared with traditional statistical approaches, the Rasch model emphasizes diagnosing individual item performance, including identifying misfitting items or those exhibiting insufficient discrimination. Using tools such as item characteristic curves, item information functions, and person-item mapping graphs, the Rasch model provides visual and quantitative insights into measurement precision across varying levels of ability. This approach not only supports the optimization of scale items but also enhances the reliability and validity of the assessment instrument, making it particularly suitable for evaluating complex, hierarchical, and multidimensional constructs such as chemistry experiment teaching literacy.

3. Design of Rasch-Based Evaluation Scale for Chemistry Experimental Teaching Literacy among Pre-Service Teachers

3.1. Establishment of the Dimensional Framework

The construction of the evaluation dimensions was based on the task structure inherent in chemistry experimental teaching, using a logical decomposition of typical instructional steps as the starting point. This approach involved a detailed examination of the processes that teachers perform during experimental activities, including preparation, operation, explanation, and reflection. Textbook experiment procedures were reviewed, and classroom teaching videos were analyzed to extract observable behaviors from each segment of the instructional task chain. These behaviors were then organized into four hierarchical levels: knowledge understanding, skill execution, expression and organization, and thinking adjustment.

Based on this analysis, a preliminary dimensional framework was drafted and subsequently validated through interviews with eight experienced frontline teachers and teaching researchers. The interviews focused on whether the identified behaviors realistically occur in pre-service teachers' practice, whether any items overlapped across dimensions, and whether there were logical inconsistencies in classification. As shown in Figure 1, the hierarchical structure of the dimensions is as follows: the foundational level emphasizes mastery of experimental principles and accurate identification of phenomena; the basic behavioral level covers standardized procedure execution and laboratory safety management; the intermediate level addresses the logic of explanation, board organization, and clarity of communication; and the higher-order cognitive level reflects the ability to adjust teaching strategies, refine instruction, and foster student-centered learning. Ultimately, this process confirmed four dimensions and twelve initial items, forming the fundamental structure for evaluating experimental teaching literacy.

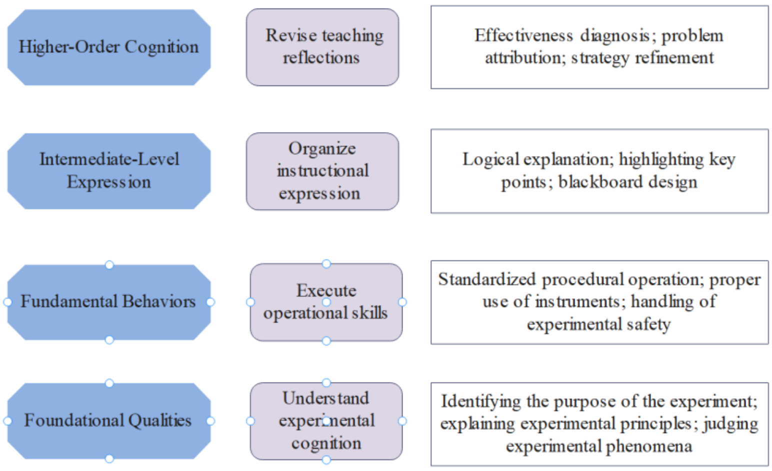

Figure 1. Multidimensional Structural Model of Pre-Service Teachers' Chemistry Experimental Teaching Literacy.

3.2. Item Generation and Compilation

Items were developed based on the instructional competency indicators defined in the dimensional framework, with particular attention to measurability, linguistic clarity, and operational relevance. The researcher collected 25 sets of chemistry experiment-teaching coursework from normal universities within the province. By analyzing the structure of experimental tasks, lesson segments, and patterns of teacher-student interaction, common classroom scenarios were identified, and typical linguistic expressions and observable behaviors were extracted.

When converting these behaviors into draft items, the principle of "one item, one meaning" was strictly applied to prevent overlap or cross-dimensional mixing. All items were phrased as descriptive statements, maintained within a length of 14 to 22 characters to facilitate subsequent coding and Rasch modeling. Wording was designed to be closely aligned with classroom practice while maintaining pedagogical logic, avoiding abstract terminology and subjective judgments. The content covered essential task behaviors, including experiment design explanations, standardized operational demonstrations, logical structuring of verbal explanations, error identification strategies, and reflective teaching adjustments. For each dimension, over twenty initial items were generated, and after removing duplicates and overly general statements, 58 items were retained for expert review. The item number was controlled to ensure comprehensive coverage of key behaviors while maintaining sample stability for model analysis and providing space for further refinement.

3.3. Expert Review and Revision

The expert review process consisted of two structured rounds. Seven chemistry education assessment experts were invited to evaluate each item dimension by dimension. Using a combination of independent scoring and collective discussion, the experts assessed scientific accuracy, clarity, and alignment with the intended construct. Feedback was categorized into four types: wording modification, semantic splitting, dimensional reclassification, and item deletion or replacement. Items with rating discrepancies greater than two points were discussed collectively, and revisions were made after reaching consensus.

After the first round, 13 items were removed due to impracticality or inconsistency with classroom practice. The second round involved focus-group discussions to resolve issues of cross-dimensional overlap and ambiguous phrasing, resulting in refined item wording and improved classification logic. In the final stage, the remaining 45 items were organized by dimension and submitted to three independent reviewers for blind evaluation. Upon confirming that no significant ambiguities or redundancies remained, the finalized item set was prepared for modeling and analysis.

3.4. Rasch Modeling Analysis

Rasch modeling was employed to validate the measurement properties of the scale, focusing on the relationship between item difficulty and respondent ability. Analytical parameters were configured in the Winsteps software, assessment data were imported, the Likert rating matrix was set, and the model computation was initiated. Infit and Outfit statistics were examined item by item to identify abnormal fluctuations. Misfitting items were compared with original descriptions to determine whether issues arose from logical inconsistencies or semantic ambiguity. Items identified as problematic were either deleted, restructured, or reworded to maintain unidimensionality.

As illustrated in Figure 2, the modeling process comprises six key stages: dimension definition, item generation, expert revision, small-sample testing, modeling computation, and item refinement, ultimately producing a stable model output. The Wright Map, together with the standard error distribution curve, was used to examine measurement precision and ensure that the scale accurately assesses respondents across the full spectrum of ability levels.

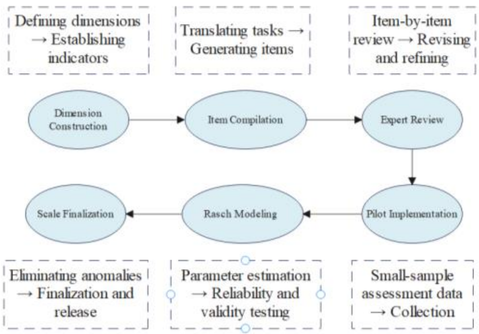

Figure 2. Multidimensional Structural Model of Chemistry Experiment Teaching Literacy for Pre-Service Teachers.

3.5. Finalization of the Scale

Following Rasch analysis and fit testing, items with poor model fit or low discrimination were removed. This iterative refinement resulted in a final set of 36 items, which constitute the formal version of the evaluation scale. As shown in Figure 3, the finalized scale provides a comprehensive and hierarchical assessment framework, with indicators and evaluation criteria systematically aligned with each dimension, enabling precise and multidimensional measurement of pre-service teachers' chemistry experimental teaching literacy.

Primary Dimension	Secondary Indicators	Rating Criteria
Understand Experimental Cognition	Identify the Purpose of the Experiment	Able to deeply comprehend the multiple purposes of the experiment and communicate them effectively.
		Able to clearly understand and explain the main purpose of the experiment to students.
		Able to recognize the basic tasks that need to be completed in the experiment for the current lesson.
	Interpretation of Experimental Principles	Able to explain experimental principles and their connection to observed phenomena in a clear and accessible manner.
		Able to correctly and clearly articulate the fundamental principles of the experiment.
		Able to provide a general explanation of the theoretical basis of the experiment.
	Judge Experimental Phenomena	Able to keenly observe all phenomena and scientifically analyze any abnormalities.
		Able to accurately observe and record the main experimental phenomena.
		Able to record most of the obvious experimental phenomena.
Execute operational skills	Standardize Operational Procedures	Demonstrates proficient, highly standardized operation with precision and efficiency.
		Able to correctly complete operations in accordance with standardized procedures.
		Able to complete basic operations with generally correct steps.
	Standardized Use of Instruments	Able to use instruments accurately and proficiently, and handle common malfunctions
		Able to use common instruments independently and according to standards
		Able to operate instruments correctly under guidance or reminders
	Experimental Safety Management	Strong safety awareness, able to proactively prevent and properly handle accidents.
		Possesses good safety awareness and follows safety regulations during operations.
		Aware of basic safety rules and performs operations with caution.
Organizing Instructional Expression	Logical Explanation	Explanation is clear and well-structured, with rigorous logic that students can easily follow.
		Explanation is organized, with key steps presented logically.
		Able to explain in a basic sequence that students can understand with effort.
	Key Point Guidance	Guidance is skillful and stimulates students' deeper thinking.
		Able to pose questions at critical teaching points to effectively guide students' thinking.
		Able to highlight key content in teaching and provide prompts.
	Blackboard Design	Blackboard design is excellent, clearly structured, and effectively supports learning.
		Blackboard is organized, with key points emphasized and easy for students to take notes.
		Blackboard presents core knowledge content that students can understand.
Revise Teaching Reflections	Effect Diagnosis	Able to comprehensively evaluate teaching outcomes and accurately identify all key issues.
		Able to assess teaching effects based on classroom observation and student feedback, identifying core issues.
		Able to conduct basic evaluation of teaching outcomes and recognize the obvious issues.
	Problem Attribution	Able to deeply analyze root causes, considering multiple factors with clear logic.
		Able to accurately analyze the main causes of problems.
		Able to identify some direct causes of problems.
	Strategy Adjustment	Able to formulate specific, feasible, and multiple solutions and plans based on attribution.
		Able to propose clear and targeted improvement strategies according to the causes.
		Able to provide preliminary adjustment ideas based on identified problems.

Figure 3. Schematic Diagram of Indicators and Evaluation Criteria at All Levels of the Scale.

4. Empirical Analysis

4.1. Sample Selection and Data Collection

Pilot-test participants were recruited from two normal universities with well-established educational practicum systems, ensuring that students had prior exposure to experimental teaching activities and structured coursework. A total of 271 third- and fourth-year pre-service chemistry teachers participated in the study. To enhance representativeness and reduce sampling bias, a stratified sampling method was applied, taking into account region, gender, and engagement in coursework-based practical activities, with each stratum maintained at a balanced proportion. Students majoring in disciplines other than chemistry or those who had transferred for short-term programs were excluded to ensure homogeneity of the sample.

The evaluation scale was administered through two channels: an on-site, closed-book session conducted under standardized conditions, and a synchronized online survey via the Wenjuanxing platform. Participants were instructed to complete the scale within 40 minutes to maintain attention and minimize fatigue effects. After data collection, 27 questionnaires were discarded due to incompleteness or significant inconsistencies, resulting in 244 valid cases for analysis. Data cleaning and preparation were carried out using R software, including standardization of responses, treatment of missing values, and identification of duplicates, producing a complete and reliable data matrix suitable for subsequent Rasch modeling. This rigorous sampling and data collection approach ensured that the dataset was representative, comprehensive, and suitable for validating the measurement properties of the scale.

4.2. Model Fit and Data Verification

Initial model fitting involved examining Infit and Outfit mean-square residual statistics for each item. Most items fell within the acceptable range of 0.5 to 1.5, indicating that the majority of items functioned as intended. Extreme cases were trimmed to reduce potential distortion, and the re-estimated reliability coefficients were 0.89 for Person Reliability and 0.92 for Item Reliability, reflecting a high degree of stability in measurement.

During the analysis, a Wright Map was generated to visually depict the alignment between respondents' abilities and item difficulties. This allowed identification of clusters of items that were either too easy or too difficult relative to participant ability, as well as gaps in the scale coverage. Items with notable misfit or low discrimination were documented for subsequent refinement. The data verification process provided essential evidence for assessing the structural rationality, internal consistency, and measurement precision of the scale, ensuring that the instrument accurately captured the multidimensional construct of chemistry experimental teaching literacy.

4.3. Scale Refinement and Finalization

Based on model-fit outputs, items were further screened and refined. Ambiguously worded items were rephrased, and the updated items were recalibrated within the model to verify improvements in fit. The optimized structure balanced content coverage and model convergence, with each subscale containing 8 to 12 items. Retained items were carefully reviewed to ensure stable measurement precision across the full range of respondent abilities.

Feedback from participants regarding clarity and applicability was incorporated into minor wording adjustments, enhancing comprehensibility without altering the intended meaning of the items. The finalized scale demonstrated good stability, discriminative power, and practical applicability, suitable for administration across different grade levels of pre-service chemistry teachers. Subsequent small-scale repeated administrations confirmed the reliability and robustness of the instrument, providing confidence that the scale can serve as a consistent and effective tool for evaluating experimental teaching literacy. The final item set, together with comprehensive measurement guidelines, was archived for wider dissemination and future use in research and teacher training programs.

5. Conclusion

Systematic assessment of experimental teaching literacy among pre-service chemistry teachers is a crucial mechanism for improving teacher education quality and informing professional development. This study clarified the conceptual definition of experimental teaching literacy, constructed a multidimensional evaluation framework, and implemented a rigorous process of item generation, expert review, and Rasch-based modeling analysis. The resulting scale exhibits a stable structure, strong model fit, and high discriminative capacity. Through empirical administration and validation, the instrument has demonstrated its practical utility in capturing variations in pre-service teachers' competence across multiple dimensions.

In addition to providing an effective diagnostic tool for evaluating current abilities, the scale can inform targeted interventions and curriculum improvements. Future research may extend its application to cross-regional comparisons, longitudinal tracking of teacher development, and integration with teaching practicum assessments. Such extensions will enhance the understanding of developmental trajectories in experimental teaching competence, enabling teacher educators to design more effective programs that support the growth of highly competent chemistry educators and strengthen the overall quality of chemistry instruction.

References

1. J. D. Novak, "Application of advances in learning theory and philosophy of science to the improvement of chemistry teaching," *Journal of chemical Education*, vol. 61, no. 7, p. 607, 1984. doi: 10.1021/ed061p607
2. B. J. Reys, S. Dingman, N. Nevels, and D. Teuscher, "High School Mathematics: State-Level Curriculum Standards and Graduation Requirements," *Center for the Study of Mathematics Curriculum*, 2007.
3. G. Abyzbekova, Z. Zholdasbayeva, A. Tapalova, S. Yespenbetova, G. Balykbayeva, and K. Arynova, "The effectiveness of the competence approach in the training of chemistry teachers," *Journal of Chemical Education*, vol. 100, no. 9, pp. 3484-3493, 2023. doi: 10.1021/acs.jchemed.3c00496
4. S. W. Bae, J. H. Lee, and J. Park, "Development of a field-based chemistry experiment teaching model to strengthen pre-service teachers' competence for teaching chemistry experiments," *Asia-Pacific Science Education*, vol. 7, no. 2, pp. 522-548, 2021. doi: 10.1163/23641177-bja10037
5. N. Yang, T. Liu, and H. Wang, "A study of chemistry laboratory teaching in secondary schools under the guidance of core literacy," In *2024 9th International Conference on Social Sciences and Economic Development (ICSSED 2024)*, July, 2024, pp. 605-610. doi: 10.2991/978-94-6463-459-4_68
6. M. Bruehl, D. Pan, and I. J. Ferrer-Vinent, "Demystifying the chemistry literature: Building information literacy in first-year chemistry students through student-centered learning and experiment design," *Journal of Chemical Education*, vol. 92, no. 1, pp. 52-57, 2015. doi: 10.1021/ed500412z
7. J. E. Upahi, R. Gbadamosi, and V. E. Boniface, "Scientific Literacy Themes Coverage in the Nigerian Senior School Chemistry Curriculum," *Journal of Turkish Science Education*, vol. 14, no. 2, pp. 52-64, 2017.
8. C. Mönch, and S. Markic, "Exploring pre-service chemistry teachers' pedagogical scientific language knowledge," *Education Sciences*, vol. 12, no. 4, p. 244, 2022. doi: 10.3390/educsci12040244
9. E. J. Yezierski, and D. G. Herrington, "Improving practice with target inquiry: High school chemistry teacher professional development that works," *Chemistry Education Research and Practice*, vol. 12, no. 3, pp. 344-354, 2011. doi: 10.1039/c1rp90041b
10. C. Morais, G. G. Junior, and C. André, "Exploring a Contextualized STEM Integration in Chemistry Education Laboratory: Insights from Pre-Service Teacher Training," *European Journal of STEM Education*, vol. 10, no. 1, p. 18, 2025. doi: 10.20897/ejsteme/17213