JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

This is an open access article under the Creative Commons Attribution 4.0 International License

CHEMISTRY TEACHERS' KNOWLEDGE IN ASSESSING SCIENTIFIC INQUIRY COMPETENCE

Ya Zhu,
Shaohui Chi,
Zuhao Wang

Abstract. *Assessing scientific inquiry competence (SIC) is essential for fostering students' inquiry skills, and the effectiveness of such assessments largely depends on teachers' assessment knowledge. Grounded in a four-dimensional framework of science teacher assessment knowledge, this study developed and validated a scenario-based diagnostic instrument incorporating authentic video clips of inquiry-focused teaching situations. The instrument was administered to 146 novice chemistry teachers in China to examine their knowledge of assessing SIC. Results showed that teachers demonstrated relatively strong knowledge of assessment content but had only moderate awareness of assessment purposes and limited knowledge to apply appropriate assessment strategies and interpret evidence. These findings highlight significant gaps in teachers' assessment knowledge of SIC, especially in integrating multiple dimensions of assessment in practice. The study underscores the need for targeted professional development to enhance teachers' assessment literacy and support effective inquiry-based science teaching, particularly in discipline-specific contexts such as chemistry.*

Keywords. *chemistry teachers, mixed-methods research, scientific inquiry competence, assessment literacy*

Ya Zhu, Shaohui Chi, Zuhao Wang
East China Normal University, P. R. China

Introduction

The generation of information and knowledge is increasingly intertwined with the growing complexity of social and environmental challenges in today's world. Responding to these challenges requires scientifically literate individuals with advanced inquiry skills. Scientific literacy and higher-order competencies have been shown to be not only critical for individuals to solve complex problems but also pivotal for promoting sustainable societal development (Effendi et al., 2021; Oliver & Oliver, 2018). Among these competencies, scientific inquiry competence (SIC) plays a foundational role. Accordingly, K–12 science education standards in many countries have highlighted the explicit development of SIC, often referred to as scientific inquiry, inquiry skills, or science practices. These abilities are widely regarded as core components of scientific literacy and have become central learning goals in recent science education reforms (Ministry of Education, 2017; NRC, 2012). To effectively foster SIC, valid and reliable assessments in inquiry-based teaching are essential—not only for tracking student progress but also for informing and improving teaching.

Despite its recognized importance, the assessment of SIC remains a persistent challenge in science education. Effective assessment involves eliciting and evaluating students' abilities to formulate research questions and hypotheses, design and conduct investigations, analyze and interpret data, and draw evidence-based conclusions (Vo & Simmie, 2024). The cognitive, procedural, and sociocultural complexity of SIC necessitates sophisticated, context-sensitive assessment practices, where conventional grading approaches often fall short. Existing assessment approach such as standardized tests, performance-based tasks, and curriculum-aligned assessments have limitations in capturing the process-oriented and context-sensitive nature of SIC. These challenges are particularly evident in inquiry-based teaching, where student thinking is often complex, non-linear, and highly contextualized.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Teachers play a pivotal role in overcoming the challenges of assessing scientific inquiry. They are tasked with designing authentic assessment tasks, interpreting nuanced student responses, and providing feedback that fosters scientific reasoning (De Jong et al., 2003; Chen et al., 2020; Petermann & Vorholzer, 2022, 2023). Drawing on the work of Magnusson (1999), Abell and Siegel (2011) have conceptualized science teacher assessment literacy as comprising the knowledge, values, and principles essential for designing and interpreting assessments—specifically, the knowledge of assessment purposes, content, strategies, and the interpretation and use of evidence. However, many teachers face difficulties in implementing effective inquiry-based assessments due to limited assessment literacy. While tools like lab notebooks and student inquiry reports can offer valuable insights into students' thinking, teachers often lack the expertise to interpret this evidence accurately and use it to inform teaching (Feldon, 2010).

Despite increasing attention to science teacher assessment literacy, several limitations have persisted in the existing literature. Most studies have focused on general science education, often overlooking the distinctive features of specific disciplines such as chemistry. Inquiry in chemistry frequently involves abstract, submicroscopic reasoning, multi-step experimental procedures, and complex data interpretation, which differ markedly from inquiry practices in other science domains (Cooper & Stowe, 2018). Furthermore, many assessment instruments lack authenticity, failing to reflect the nuanced, process-oriented, and context-sensitive nature of inquiry tasks and student interactions in inquiry-based science classrooms (Rodriguez et al., 2020). These limitations underscore the need for discipline-specific, contextually grounded instruments to effectively evaluate teachers' knowledge of assessing scientific inquiry competence.

In summary, assessing students' scientific inquiry competence present significant challenges, especially in chemistry, where inquiry processes involve abstract reasoning and complex experiments. Many teachers lack the assessment knowledge to effectively design and interpret such tasks (Ruiz-Primo & Furtak, 2007; Yan & Pastore, 2022). While teacher knowledge is central to high-quality science teaching, its development—especially in the domain of assessment—remains an ongoing challenge (van Driel, Hume, & Berry, 2023). Moreover, existing assessment tools are frequently too generic and insufficiently aligned with real classroom contexts. These limitations underscore the pressing need for subject-specific and contextually authentic assessment instruments tailored to the demands of chemistry education (Chen et al., 2019).

Literature Review

Assessment of Scientific Inquiry Competence: Approaches and Challenges

Scientific inquiry competence (SIC) has become a central focus in science education worldwide, prompting the development of diverse assessment approaches. Standardized tests, such as PISA, provide scalability through structured multiple-choice and open-ended items. However, these assessments often fail to capture the iterative, contextual, and process-oriented nature of authentic inquiry (Yamamoto et al., 2018). Despite these limitations, standardized formats remain widely used, primarily evaluating the end-products of inquiry rather than students' reasoning and decision-making processes.

Performance-based assessments offer deeper insights into students' inquiry practices by engaging them in authentic or simulated scientific tasks. Although these formats generate rich data, they are resource-intensive and often difficult to scale, which limits their practical implementation (Emden & Sumfleth, 2016). Meanwhile, curriculum-based assessments align closely with classroom instruction and are useful for providing targeted feedback to inform teaching and learning (Hintze et al., 2006).

The integration of digital technologies has opened new opportunities for assessing SIC. Platforms such as the Web-based Inquiry Science Environment (WISE) and Inq-ITS incorporate dynamic simulations, real-time feedback, and AI-based analytics to guide students through the phases of inquiry while capturing detailed behavioral data (Beller, 2013; Gobert et al., 2013; Sui et al., 2024). Virtual performance assessments enable students to conduct investigations in simulated environments, logging their reasoning and decisions in real time (Zachos et al., 2000). These tools exemplify a shift toward interactive, adaptive, and process-oriented assessment formats. While no single assessment format fully captures the complexity of scientific inquiry, emerging research has suggested that hybrid approaches—such as combining digital simulations with teacher-led formative assessments—can offer a more comprehensive evaluation of both procedural knowledge and epistemic reasoning (Furtak et al., 2014). This calls for an integrated assessment system that leverages the strengths of multiple formats depending on learning goals and classroom contexts.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

While these innovations mark significant progress, they also expose enduring challenges in the assessment of scientific inquiry competence. Many current assessment models emphasize cognitive or psychological dimensions of inquiry while underrepresenting emotional, ethical, and sociocultural aspects (Wu et al., 2014). Moreover, both hands-on and virtual inquiry tasks—despite their authenticity—are often difficult to standardize or scale due to infrastructure limitations and the need for specialized teacher training (Quellmalz et al., 2012; Kuo et al., 2015). The absence of a universally accepted framework for assessing SIC further complicates efforts to ensure coherence and comparability across contexts.

Emerging scholarship calls for multimethod and multidimensional approaches to SIC assessment (Kruit et al., 2018; Teig, 2024). However, the effectiveness of these approaches ultimately hinges on teachers' capacity to design, implement, and interpret assessments. As assessment tools become increasingly sophisticated, teachers are required to demonstrate advanced diagnostic and interpretive expertise. These growing demands underscore the critical need to understand and support science teachers' assessment knowledge—particularly within the context of inquiry-based teaching.

Science Teacher Assessment Literacy: Frameworks and Gaps

Science teacher assessment literacy (STAL) has been increasingly recognized as a pivotal factor influencing science teaching. Building on Magnusson's (1999) empirical work, Abell and Siegel (2011) conceptualized STAL as the professional knowledge, values, and principles that teachers employ in designing and interpreting assessments. Their framework delineates four core dimensions: assessment purposes, assessment content, assessment strategies, and assessment interpretation and action.

Assessment purposes encompass teachers' knowledge of the goals and functions of assessment, including diagnostic, formative, summative, and metacognitive roles. Assessment content highlights the necessity of aligning assessment tasks with instructional objectives, with authentic inquiry assessments reflecting the multi-phased nature of scientific investigations. Assessment strategies involve the specific tools and techniques—such as questioning, rubrics, and written feedback—that teachers use to elicit meaningful thinking, alongside proficiency in employing these methods effectively. Finally, assessment interpretation and action refer to teachers' ability to analyze student responses and leverage these insights to inform teaching decisions, deepen conceptual knowledge, and support students' learning.

Subsequent empirical studies have further elucidated how these dimensions manifest in classroom practice. For instance, Gottheiner and Siegel (2012) investigated science teachers' use of formative strategies like prediction and explanation within genetics instruction. More recently, Hung and Wu (2024) refined the STAL model by examining assessment practices specifically within inquiry-based teaching contexts. Their findings underscore the complex interplay among assessment purposes, student work (as content), strategies, and scoring literacy, emphasizing that effective assessment competence extends beyond technical instrument use to encompass reflective judgment and pedagogical adaptation.

Despite these advancements, current research remains predominantly qualitative and generalized, with limited attention to disciplinary specificity (Abell & Siegel, 2011; Klenowski, 2011). Chemistry presents distinctive challenges for inquiry assessment due to its reliance on abstract reasoning, multi-step experimental procedures, and complex data interpretation (Cooper & Stowe, 2018). Yet, systematic or empirical investigations into chemistry teachers' assessment knowledge are scarce (Rodriguez et al., 2020). Moreover, many existing assessment tools employed in STAL research lack authentic contextual grounding. They frequently feature decontextualized or hypothetical tasks that fail to reflect the ambiguity, openness, and complexity characteristic of real classroom inquiry (Pellegrino, DiBello, & Goldman, 2016). As a result, such tools offer limited insight into how teachers engage with assessment practices in authentic, content-rich settings like chemistry, where representational competence and experimental reasoning are central to inquiry-based teaching.

Addressing these gaps necessitates research that (1) empirically examines science teachers' assessment knowledge within discipline-specific contexts, and (2) develops authentic, scenario-based assessment instruments that reflect the real-world demands of inquiry-based teaching. Such efforts would facilitate more targeted professional development, enhance teachers' assessment competence, and ultimately contribute to improve student learning outcomes in science education.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Theoretical Framework

To address these limitations, this study proposed a theoretical framework to explore how chemistry teachers conceptualize and apply their knowledge in assessing students' scientific inquiry competence (SIC). Building on Abell and Siegel's (2011) model and incorporating insights from recent empirical studies (Gottheiner & Siegel, 2012; Hung & Wu, 2024), the framework integrated four interconnected domains of teacher knowledge: assessment content, purposes, strategies, and interpretation.

Central to this framework is the notion that teachers need to hold a clear and curriculum-aligned knowledge of what constitutes scientific inquiry competence. In line with national curriculum standards and cognitive models of scientific reasoning (Abd-El-Khalick et al., 2004; Talanquer et al., 2013; Wang et al., 2010; Grob et al., 2019), this content knowledge serves as the foundation upon which assessment activities are designed and interpreted. In chemistry teaching context, this means recognizing both inquiry as a means for knowledge construction and inquiry as a learning objective.

Equally important is teachers' knowledge of the assessment purposes in inquiry-based contexts. Effective assessment requires distinguishing between diagnostic, formative, summative, and metacognitive functions (Gipps, 2011) and using each purpose appropriately throughout the instructional process. For example, formative assessments embedded in inquiry tasks can guide student thinking in real time (Sasse et al., 2025), while summative assessments at the end of a unit measure learning outcomes (Heyde & Siebrits, 2019). Metacognitive assessments support students' ability to reflect on and regulate their learning during inquiry.

The third domain addresses teachers' knowledge of assessment strategies to elicit evidence of SIC. Effective practice involves selecting or designing tools—such as performance tasks, lab notebooks, observational rubrics, and peer/self-assessments—that capture students' reasoning and investigative processes. As Magnusson et al. (1999) emphasized, the success of these strategies depends on teachers' ability to adapt and integrate them into teaching.

The final domain involves interpreting assessment data and providing actionable feedback. Grounded in Pellegrino et al.'s (2001) "assessment triangle," this component stresses the importance of making sense of student responses and translating them into informed instructional actions. In inquiry-based contexts, feedback should be descriptive, targeted, and forward-looking, helping students recognize gaps in reasoning and refine their inquiry processes—especially in performance-based or open-ended tasks.

This framework conceptualizes science teacher assessment knowledge as an integrated system that supports valid, responsive, and context-sensitive evaluation of inquiry. While recent research acknowledges the complexity of assessing inquiry, it often overlooks teachers' central role in implementing assessments. This study addresses that gap by investigating chemistry teachers' assessment knowledge and laying the groundwork for targeted professional development and assessment tool design.

Grounded in the reviewed literature and informed by the proposed theoretical framework, this study seeks to deepen our understanding of chemistry teachers' knowledge of assessing students' scientific inquiry competence (SIC). The literature review highlighted persistent challenges in inquiry assessment and underscored the need for discipline-specific assessment tools. The theoretical framework provided a structured lens through which different dimensions of teachers' assessment knowledge could be conceptualized and examined.

Accordingly, this study addressed the following research questions:

1. What are the reliability and validity of the developed instrument in examining chemistry teachers' knowledge of assessing students' scientific inquiry competence?
2. How do novice chemistry teachers demonstrate their knowledge across the four dimensions of inquiry assessment: assessment content, assessment purposes, assessment strategies, and assessment interpretation and feedback?

Research Methodology

General Background

The Chemistry Curriculum Standards for upper-secondary schools in China emphasize scientific inquiry competence (SIC) as a core component of students' key competencies (Ministry of Education, 2017). In response, the need to effectively assess SIC has emerged as a pressing concern. However, research that specifically investigates teachers' knowledge for assessing SIC remains limited, particularly in chemistry. Addressing this gap, the present study developed and validated a diagnostic instrument grounded in authentic instructional scenarios, aiming to

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

evaluate chemistry teachers' knowledge across four key domains: assessment content, purposes, strategies, and interpretation.

Grounded in a constructivist perspective on assessment and informed by the science teacher assessment literacy (STAL) framework proposed by Abell and Siegel (2011), this study adopted a mixed-methods approach. The research was conducted from September 2023 to September 2024 and involved 146 novice chemistry teachers from 20 provinces across mainland China. Data were collected through a scenario-based paper test and semi-structured cognitive interviews (30 minutes). During the assessment, teachers watched six video segments of inquiry-based chemistry lessons (focused on vitamin C) and responded to corresponding open- and closed-ended items. The assessment took teachers approximately 45 minutes to complete, not including the time spent watching the video segments.

The combination of quantitative and qualitative data allowed for triangulation, providing a more comprehensive understanding of teachers' strengths, difficulties, and thought processes in assessing SIC. This instrument not only supports empirical insights into teachers' assessment literacy but also offers implications for professional development and instructional improvement in inquiry-based science education.

Participants

A convenience sampling strategy (Robinson, 2014) was used to administrated 146 novice chemistry teachers enrolled in a graduate program at East China Normal University in Shanghai. These participants, representing 20 provinces across mainland China, voluntarily took part in the study. As shown in Table 1, 81% were female, and most taught in urban high schools. Nearly all participants (99%) had experience with inquiry-based teaching. All held bachelor's degrees from Ministry-affiliated normal universities and were currently pursuing graduate studies. All had less than five years of teaching experience.

Ethical approval was obtained from the institutional review board at East China Normal University. Before the assessment, we provided all participating teachers with informed consent forms to ensure that their participation was voluntary, that they were fully aware of the purpose and procedures of the study, and that their data would be kept confidential and used solely for research purposes. We also gave small gifts to the participating teachers as a token of appreciation. Prior to the interviews, teachers were informed that the conversations would be audio-recorded.

Table 1
Demographic Characteristics of Participants at Baseline

Baseline characteristic	*n*	%
Gender		
Male	27	19
Female	119	81
School level		
Primary school	1	1
Lower secondary school	34	23
Upper secondary school	111	76
School location		
Urban area	136	93
Town area	10	7
Rural area	0	0
Experience in inquiry-based teaching		
Yes	145	99
No	1	1

Note. $N = 146$ (n for each condition). Participants were on average 1–5years teaching experience.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Instrument

Item development was guided by a comprehensive review of existing formative assessment instruments and empirical research in science education (e.g., Yan & Pastore, 2022). Items were systematically aligned with Abell and Siegel's (2011) science teacher assessment literacy framework to ensure thorough coverage of the four key knowledge domains.

The development process involved a panel of experts—including three chemistry education professors, two experienced teachers, and members of the research team—who evaluated content relevance, clarity, and alignment with curricular goals. The Delphi technique was employed for expert validation, involving multiple rounds of iterative review to refine items based on feedback concerning their relevance, clarity, and alignment with assessment objectives (Linstone & Turoff, 1975; McKenna, 1994). This process resulted in enhanced item precision and comprehensive coverage of the assessment dimensions.

A pilot test involving 43 upper secondary chemistry teachers was conducted to optimize item quality, instrument structure, and administration procedures. Retrospective probing interviews assessed participants' interpretation and responses to test items. Based on feedback and data analysis, two items with low discriminatory power were removed. Subsequent improvements included adding subtitles and transcripts to the videos (see Appendix 2) to aid comprehension, revising terminology to align with national standards—specifically replacing "formative assessment" with "process-oriented assessment," consistent with the Chemistry Curriculum Standards (Ministry of Education, 2017), thereby emphasizing ongoing student learning processes.

The final instrument comprised 21 items—multiple-choice, fill-in-the-blank, and short-answer questions—with 4–6 items targeting each sub-dimension (see Appendix 1). To increase ecological validity, a video of an inquiry-based vitamin C lesson was segmented into six parts and served as stimulus material for 17 items (Q1–Q17), complemented by four scenario-based items (Q18–Q21). Utilizing a consistent topic helped reduce cognitive load and content variability, while minimizing participants' reliance on external resources such as internet searches or AI assistance.

Table 2
Basic Information of Video Segment Used for Items

No.	Video Content	Duration(minute)	Item
A	The effect of vitamin C on dough	2:11	Q1, Q2, Q6, Q8, Q16
B	Student's experiment process on vitamin C content in spinach and oranges	3:04	Q3, Q4, Q6, Q8, Q17
C	Discussion on the principles of "determining vitamin C content in fruits"	2:57	Q12, Q14 Q9–Q11
D	Discussion on experimental results of "Comparing vitamin C content in different fruit juices"	1:42	Q5, Q7,Q9–Q11
E	Experiment improvement of "comparing vitamin C content in different fruit juices"	2:18	Q8, Q9–Q11
F	Reflection on the "determination of vitamin C in fruits" experiment	2:43	Q14, Q16, Q17, Q9–Q11
	Scenario-based item		Q18–Q21

Scoring Criteria and Reliability Measures

To ensure consistency and objectivity in scoring, a standardized scoring framework was developed. A detailed scoring rubric was established to evaluate varying levels of teacher knowledge. The validation of the scoring rubric followed a rigorous and systematic process. Expert review was conducted by a panel consisting of chemistry education professors, two experienced secondary chemistry teachers, and members of the research team. This panel evaluated the appropriateness, clarity, and alignment of the scoring rubrics with the assessment objectives. The instrument employed a differentiated scoring scheme: 4 out of 21 items were rated on a 0–1 scale, 11 items on a 0–2 scale, while the remaining fill-in-the-blank and short-answer questions were scored based on the presence of key conceptual elements.

To ensure inter-rater reliability (Gwet, 2014), two researchers with expertise in science education were trained as independent raters. A calibration process was conducted, including guided scoring sessions and consensus-

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

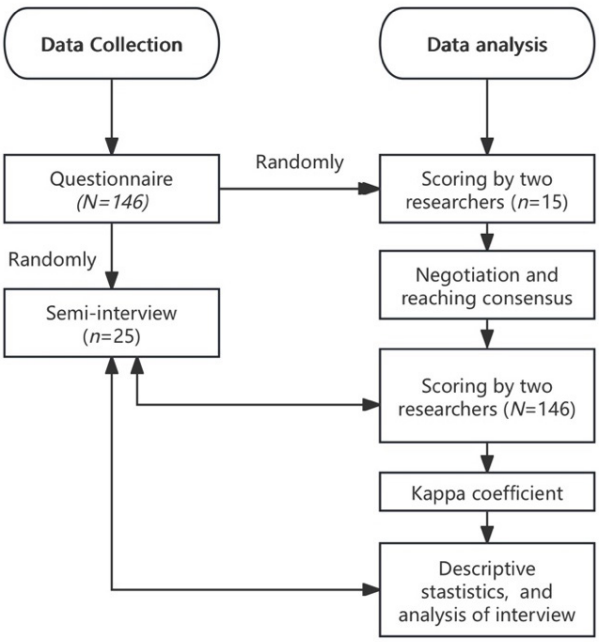
building discussions on ambiguous responses. Any scoring discrepancies were resolved through iterative dialogue. The resulting inter-rater agreement yielded a Cohen's kappa coefficient of .85 ($p < .01$), indicating a high level of reliability (Cohen, 1960).

Data Analysis

This study conducted a reliability analysis of the instrument using SPSS 27.0 to assess internal consistency, employing Cronbach's α coefficient. All knowledge dimensions exceeded $\alpha = .70$, confirming the instrument's reliability.

To quantify teachers' knowledge, z-scores (ranging from 0 to 1) were calculated for each sub-dimension. Participants were then classified into three groups based on percentile ranks: professional (top 25%), developing (middle 50%), and limited (bottom 25%). Descriptive statistics (means and standard deviations) were used to examine performance variability. Cognitive interviews with 25 novice teachers (each 30–40 minutes) provided qualitative insights into common conceptions and misconceptions across the knowledge domains. In addition to z-scores and percentile classification, mean and standard deviation were calculated to examine the central tendency and variability of scores within and across the four dimensions.

Figure 1
Data Collection and Analysis Procedures



Research Results

The results (see Figure 2) indicated variation in teachers' proficiency across the four assessment knowledge domains. The lowest mean score was observed in the assessment strategy dimension ($M = 0.467$), suggesting that teachers experienced challenges in applying a range of assessment strategies. This domain also showed the highest standard deviation ($SD = 0.214$), indicating substantial variability among participants, potentially influenced by factors such as teaching experience and professional development. In contrast, teachers demonstrated relatively stronger performance in the assessment knowledge ($M = 0.650$) and assessment content ($M = 0.566$) dimensions. Their performance in the assessment purpose and feedback domains was moderate, reflecting partial understanding and application of these areas.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Figure 2
Results of Chemistry Teachers' Knowledge of Assessing SIC

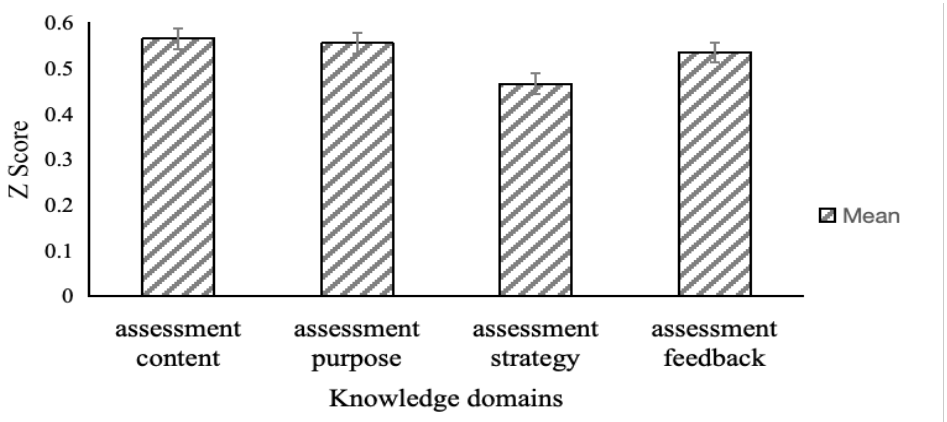
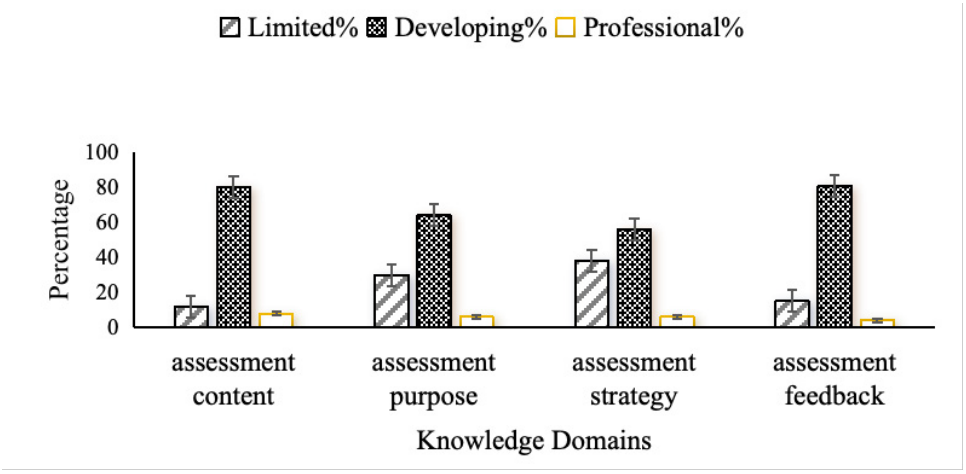


Figure 3 illustrates the distribution of proficiency levels across four domains of assessment knowledge: content, purpose, strategy, and feedback. Proficiency is categorized into three levels: Limited, Developing, and Professional. In the domain of assessment content, many participants (approximately 80%) were classified as Developing, with around 15% in the Limited category and fewer than 5% in the Professional category. A similar pattern was observed for assessment purpose, with about 65% categorized as Developing, 30% as Limited, and fewer than 5% as Professional. For assessment strategy, approximately 55% of teachers were at the Developing level, followed by 40% at Limited, and less than 5% at Professional. In the feedback domain, about 80% were identified as Developing, 15% as Limited, and fewer than 5% as Professional. Overall, the data indicate that most participants fell into the Developing category across all four domains, with relatively few achieving Professional proficiency.

Figure 3
Distribution of Chemistry Teachers' Assessment Knowledge Across Four Domains by Proficiency Level (Limited, Developing, Professional)



Knowledge of Assessment Content

Most chemistry teachers recognized Item 1 (Video A: The effect of vitamin C on dough) and Item 3 (Video B: Student's experiment process on vitamin C content in spinach and oranges) as examples of scientific inquiry. Their responses frequently included terms such as "experiment," "control variable," and "controlled experiment," indicating a predominantly procedural interpretation. One participant outlined the inquiry sequence as identifying a problem,

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

proposing a hypothesis, designing and conducting experiments, making comparative observations, analyzing data, and drawing conclusions—affirming video A's alignment with the essential features of scientific inquiry. Teachers described scientific inquiry as a comprehensive process encompassing question formulation, controlled experimental design, systematic observation, and evidence-based conclusions drawing. For example, Teacher F023 stated that video A clearly articulated a research question, employed a three-group controlled experiment, and interpreted observations to reach a conclusion—demonstrating a complete inquiry process.

Despite acknowledging the structured inquiry process in Video A, several teachers expressed concerns about its scientific rigor. A primary critique was the absence of an explicitly stated hypothesis. As Teacher F024 noted, the lack of a clear hypothesis made the research question appear ambiguous. Additionally, teachers pointed to limitations in the experimental design, citing vague procedural goals and superficial conclusions. Teacher F125 remarked that both the design and the data analysis lacked sufficient rigor. One teacher observed that video A merely demonstrated a process without explaining the underlying scientific concepts, while others noted the absence of student engagement, suggesting it resembled a demonstration rather than an authentic inquiry activity.

Items 4 (identifying dimensions of scientific inquiry competence) and 6 (designing assessment dimensions for inquiry competence) received relatively high scores, suggesting that teachers placed strong emphasis on components such as question formulation, observational documentation, data analysis, and reporting standards. However, responses to Item 5 indicated that many teachers were not yet fully familiar with the specific proficiency levels and detailed criteria associated with scientific inquiry competence.

In summary, chemistry teachers performed well in knowledge of assessment content but struggled to accurately interpret the requirements of the curriculum standards for assessing scientific inquiry competence. While most teachers were able to identify key components of scientific inquiry—such as formulating research questions, developing hypotheses, designing experiments, analyzing data, and drawing conclusions—some mistakenly viewed inquiry as a rigid, linear process. This suggests that participants' knowledge of scientific inquiry competence was largely grounded in the procedural aspects of inquiry, while their awareness of the specific proficiency levels outlined in the curriculum standards remains limited.

Knowledge of Assessment Purpose

The primary goal of assessing scientific inquiry competence (SIC) is to evaluate students' inquiry-based thinking and support instructional decision-making. Assessment practices can be classified as diagnostic, formative, or summative, each differing in purpose, timing, and application. In this study, teachers were guided to align assessment methods with intended purposes using video-based stimuli and structured worksheets. However, low performance on Item 8 (assessment purposes) and Item 9 (selecting appropriate formative assessment methods) revealed notable gaps in teachers' knowledge of how to purposefully assess SIC.

First, some teachers demonstrated difficulty distinguishing between formative and summative assessments. While formative assessment is intended to provide process-oriented feedback and support the development of inquiry competence, summative assessment is designed to evaluate learning outcomes. For example, Teacher F045 misclassified laboratory worksheets and inquiry projects as summative assessments. Similarly, Teacher F023 perceived group discussions merely as instructional strategies rather than valid assessment tools.

Second, even when teachers were aware of various assessment methods, they faced difficulties in implementing them effectively within inquiry-based teaching due to contextual constraints. Some were uncertain about the appropriate contexts for using group discussions and oral reports. For example, Teacher F066 noted that time constraints made it challenging to determine the suitability of such formative approaches for assessing scientific inquiry competence. This uncertainty underscores the need for clearer guidance on how to integrate diverse assessment strategies into inquiry-based teaching.

Overall, the findings revealed substantial gaps in teachers' knowledge of assessment purposes and their ability to select appropriate methods within inquiry-based settings. These challenges constrained their proficiency in designing coherent assessment strategies, thereby diminishing the teaching value of assessment.

Knowledge of Assessment Strategies

The assessment strategies of scientific inquiry competence encompass formative and summative assessment, self-assessment and peer-assessment, performance-based assessment, and portfolio assessment. Items 12–15 addressed teachers' knowledge of these strategies, with Item 15 (inquiry worksheets) receiving the lowest aver-

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

age score (M = 0.479). An analysis of responses to the worksheet *Comparing Vitamin C Content in Different Fruits* showed that chemistry teachers primarily listed assessment elements, such as experimental design, procedures, objectives, data analysis, and conclusions, but failed to specify detailed evaluation criteria. For instance, while the scoring rubric emphasized variable identification and control, many teachers only mentioned experimental design without further elaboration.

Interview data revealed additional gaps. Some teachers relied on intuition when assessing student work (e.g., Teacher F056 stated, "I don't know how to design worksheets, so I assess based on feeling"), while others reported simply adapting existing worksheets from online sources (e.g., Teacher F067 noted, "I usually download and modify them"). In summary, teachers demonstrated limited knowledge of assessment instruments, which constrained their ability to develop essential competencies in designing and evaluating inquiry-based assessment tools.

Knowledge of Feedback and Interpretation

Feedback on the assessment of scientific inquiry competence constitutes a critical component in guiding students to reflect upon and enhance their inquiry skills. Items 16–21 revealed several noteworthy patterns. Teachers performed well on structured tasks, with relatively high average scores for Item 16 (refining experimental procedures; M= 0.808) and Item 17 (selecting the optimal experimental design; M = 0.829). However, significant challenges emerged in assessing students' ability to formulate inquiry questions, as reflected by the notably low average score for Item 18 (M = 0.308). This suggests that many teachers had difficulty distinguishing between varying levels of inquiry-based questions.

Scores for Items 19 through 21 were more evenly distributed, indicating a moderate level of knowledge in those areas. Many teachers exhibited limited familiarity with the assessment criteria, which impeded their ability to differentiate the complexity of student-generated questions. As one teacher (F023) noted, "The distinction between simple and inquiry-based questions is unclear." Additionally, some participants expressed difficulty in integrating subject-matter knowledge into the assessment process, particularly when evaluating students' formulation of questions and hypotheses. For example, Teacher F025 stated, "I rarely assess students explicitly based on different levels."

In summary, while teachers demonstrated competence in evaluating structured experimental tasks (e.g., refining procedures, selecting appropriate designs), they experienced considerable difficulties in assessing formulating inquiry questions. These challenges were primarily due to limited knowledge in question complexity and unfamiliarity with detailed assessment criteria.

Discussion

Procedural Understanding of Assessment Content

Many teachers equated scientific inquiry with the execution of experimental procedures. A process-oriented view emphasizes "doing" rather than "thinking" (McElhaney & Linn, 2011; Rönnebeck et al., 2016). This limited knowledge constrains teachers' ability to assess students' conceptual reasoning and epistemic engagement (Ruiz-Primo & Furtak, 2007). These tendencies align with longstanding patterns in science education, where procedural skills are prioritized at the expense of deeper conceptual understanding (Bartos & Lederman, 2014).

Several interconnected factors may contribute to this pattern. First, pre-service training programs often provide insufficient emphasis on inquiry-based pedagogical content knowledge (PCK), resulting in fragmented conceptions of the epistemic dimensions of inquiry (Dorji & Tshering, 2024; Suh & Park, 2017). Second, teaching and learning materials—such as chemistry textbooks—frequently portray inquiry as a fixed sequence of experimental steps, thereby reinforcing a narrow, procedural model (Davis et al., 2011). Third, the accountability-driven culture prevalent tends to place disproportionate value on outcomes and task completion, often at the expense of epistemic practices, a concern echoed in other educational contexts as well (Birenbaum, 2016). Collectively, these factors constrain teachers' holistic knowledge of scientific inquiry and contribute to the prevalence of superficial assessment practices in inquiry-based teaching settings.

Misconceptions About the Assessment Purpose

This study identified widespread misconceptions among teachers regarding the purpose of assessing scientific inquiry competence (SIC), particularly in distinguishing formative from summative assessments. Many teachers

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

placed greater emphasis on final products rather than formative feedback, indicating a limited knowledge of assessment as an integral instructional tool.

This confusion likely originates from teacher education programs, where formative assessment is often presented in generalized terms and not explicitly framed as an embedded feature of inquiry-based pedagogy (Black, 2015). Consequently, pre-service training may not sufficiently prepare teachers to incorporate formative assessment effectively within authentic inquiry contexts.

As a result, a persistent misalignment exists between teachers' assessment knowledge and the goals of inquiry-based curricula. Teachers frequently encounter difficulties in evaluating both procedural and conceptual dimensions of inquiry, particularly in chemistry, where assessment tasks require nuanced judgments of students' reasoning processes. This challenge reflects a broader issue in science education: the complexity of operationalizing inquiry outcomes in ways that are both measurable and pedagogically meaningful (Shavelson et al., 2008). Consequently, many teachers tend to rely predominantly on summative assessments, which inadequately capture students' abilities to formulate questions, design investigations, or engage in epistemic reasoning.

Addressing these misconceptions necessitates sustained professional development efforts that explicitly integrate formative assessment within discipline-specific inquiry teaching practices, thereby enhancing the alignment between assessment and instructional goals.

Fragmented Assessment Strategies and the Need for Integrated Approaches

The study further revealed that most teachers lack coherent and integrated assessment strategies for evaluating scientific inquiry competence. Although authentic scientific inquiry inherently requires the integration of cognitive reasoning and procedural skills, teachers predominantly rely on fragmented assessment tools—such as isolated worksheets or disconnected lab reports, that fail to provide a holistic assessment of scientific inquiry competence.

This fragmented approach reflects broader limitations in teachers' assessment literacy. Designing effective performance-based assessments requires skills in rubric development and evidence interpretation, which many teachers have not acquired (Reynolds & Park, 2021). Moreover, within the Chinese educational context, these challenges are exacerbated by a dominant exam-oriented culture that heavily emphasizes content-driven summative testing (Yan et al., 2021). This systemic preference discourages teachers from adopting more complex, process-oriented assessment approaches essential for authentic inquiry evaluation.

Consequently, without adequate institutional support and sustained, targeted professional development, teachers are likely to continue employing reductive and fragmented strategies. Such practices inadequately capture the multidimensional nature of inquiry learning, thereby limiting the effectiveness of assessment in fostering students' scientific inquiry competencies.

Challenges in Providing Diagnostic Feedback and Interpreting Inquiry Performance

This study revealed a significant gap in chemistry teachers' knowledge regarding the provision of diagnostic feedback and the interpretation of students' inquiry performance. While participants could identify fundamental inquiry components, such as hypotheses and research questions, they often struggled to evaluate their quality, relevance, and scientific rigor, particularly within the chemistry discipline. For example, differentiating between investigative and descriptive questions or assessing the robustness of a hypothesis requires a combination of pedagogical content knowledge (PCK) and disciplinary expertise—competencies that many teachers lacked.

These findings corroborate prior research emphasizing the critical role of PCK in inquiry-based assessment (Reynolds & Park, 2021). Consistent with Black and Wiliam's (2018) assertion, effective formative assessment depends on clear rubrics and criteria—elements that were notably absent in the observed assessment practices. Furthermore, teachers' reliance on fragmented assessment tools, such as isolated lab reports or worksheets, constrained their ability to integrate multiple evidence sources necessary for comprehensive judgment (Özer, 2023).

Collectively, these results highlight the urgent need for sustained, subject-specific professional development focused on assessment design, interpretation, and feedback delivery. Enhancing teachers' assessment literacy—especially in formative assessment strategies and epistemic reasoning—is essential to improving the effectiveness and authenticity of inquiry-based science teaching.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Conclusions and Implications

This study investigated chemistry teachers' knowledge of assessing scientific inquiry competence (SIC) through authentic instructional scenarios presented via video segments. By integrating both video-based and scenario-based assessment items, the study introduced an innovative approach for examining teachers' assessment knowledge within the context of inquiry-based instruction. The findings contributed to the growing body of literature advocating for more integrated and context-sensitive evaluations of teacher knowledge. Notably, the developed instrument demonstrated effectiveness in capturing teachers' knowledge of SIC assessment and showed promise as a professional development resource for both pre-service and in-service chemistry teachers.

Despite these contributions, the study uncovered several challenges. A key issue was the misalignment between assessment content, purpose, strategies, and feedback and the overarching goals of inquiry-based curricula. Many teachers adopted a predominantly "process-centered" view of inquiry, focusing on procedural elements such as experimental design and data collection, while underestimating the potential of formative assessment in fostering students' inquiry competence. These findings highlighted the importance of strengthening teachers' knowledge of assessment purposes—particularly the distinction and integration of formative and summative assessments—as a means of improving the timing, design, and contextual application of assessment strategies. Moreover, teachers' difficulties in clarifying assessment objectives, selecting appropriate methods, and distinguishing levels of inquiry competence underscored the complexity of developing robust assessment literacy in inquiry-based teaching contexts.

The study highlighted the need for teacher education programs to provide more robust support in cultivating assessment literacy, especially with respect to scientific inquiry tasks. Professional development also needed to prioritize the development of competencies in designing high-quality assessment instruments and delivering accurate, meaningful feedback. Additionally, teacher development initiatives needed to focus on the construction and use of instruments—such as rubrics—that supported the precise and consistent evaluation of students' inquiry processes.

Despite the valuable insights provided by this study, several limitations should be acknowledged. First, the participant sample was limited to chemistry teachers, which may restrict the generalizability of the findings to other subject areas. Teachers in disciplines such as biology or physics may encounter different challenges when assessing scientific inquiry competencies. Second, the study focused exclusively on teachers with one to three years of teaching experience. As a result, the perspectives and practices of more experienced educators—who may have developed different approaches or encountered additional challenges—were not represented. Third, although the use of video-based and scenario-based assessment tasks offered a practical and context-rich way to explore teachers' assessment knowledge, these tools may not fully capture the dynamic and complex nature of real-time classroom assessment. Factors such as classroom interactions, spontaneous feedback, and time constraints are difficult to simulate in such formats. Therefore, while the study yielded meaningful findings, caution should be taken when extending its implications beyond the immediate context.

To address the limitations identified in this study, future research should include a more diverse sample of teachers, encompassing a broader range of teaching experience, subject disciplines, and geographical contexts. Expanding the sample would enhance the generalizability of findings and provide more comprehensive insights into the assessment of scientific inquiry competence across educational settings. Longitudinal studies are also recommended to examine how teachers' assessment knowledge develops over time and to assess the sustained impact of professional development initiatives. Additionally, future research could explore the interplay between teachers' subject-matter knowledge and their assessment practices to better understand how disciplinary knowledge informs assessment decisions. Employing participatory research approaches may further uncover both the challenges teachers face and the support structures that facilitate effective inquiry assessment. Finally, integrating classroom observations, analyses of student learning outcomes, and longitudinal designs would offer a more holistic view of how teachers' inquiry assessment knowledge evolves and translates into classroom practice.

This study provided a nuanced understanding of chemistry teachers' assessment knowledge within the context of scientific inquiry, thereby contributing to a more comprehensive conceptualization of assessment literacy, especially in relation to inquiry-oriented instructional tasks. The findings revealed persistent challenges in teachers' assessment practices, particularly due to the lack of sustained professional development focused on formative assessment in inquiry-based teaching settings. Addressing the disconnect between educational policy intentions and actual classroom implementation required equipping teachers with appropriate assessment tools, effective strategies, and continuous support.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

From a practical standpoint, the study highlights several priority areas for professional development. These include the construction and application of rubrics specifically designed for assessing inquiry processes, the formative use of group discussions and peer feedback, the interpretation of open-ended student responses, and the alignment of assessment instruments with curricular standards. Strengthening teacher knowledge in these areas is essential to promoting more effective and coherent inquiry-based science education.

Acknowledgment

The authors sincerely appreciate the anonymous reviewers for their valuable feedback and insightful suggestions. Thanks to 146 chemistry teachers who generously participated in the assessment and 43 chemistry teachers who took part in the pilot study.

Declaration of Interest

The authors declare no competing interest.

References

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., Niaz, M., Treagust, D., & Tuan, H. L. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*(3), 397–419. https://doi.org/10.1002/sce.10118

Abell, S. K., Smith, D. C., & Volkmann, M. J. (2006). Inquiry in science teacher education. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education* (pp. 173–199). Springer. https://doi.org/10.1007/978-1-4020-5814-1_9

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Springer. https://doi.org/10.1007/978-90-481-3927-9_12

Bartos, S. A., & Lederman, N. G. (2014). Teachers' knowledge structures for nature of science and scientific inquiry: Conceptions and classroom practice. *Journal of Research in Science Teaching, 51*(9), 1150–1184. https://doi.org/10.1002/tea.21168

Beller, M. (2013). Technologies in large-scale assessments: New directions, challenges, and opportunities. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 25–45). Springer. https://doi.org/10.1007/978-94-007-4629-9_3

Birenbaum, M. (2016). Assessment culture versus testing culture: The impact on assessment for learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 275–292). Springer. https://doi.org/10.1007/978-3-319-39211-0_16

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161–177. https://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Chen, Y. C., Pan, Y. T., Hong, Z. R., Weng, X. F., & Lin, H. S. (2019). Exploring the pedagogical features of integrating essential competencies of scientific inquiry in classroom teaching. *Research in Science & Technological Education*, *38*(2), 185–207. https://doi.org/10.1080/02635143.2019.1601075

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cooper, M. M., & Stowe, R. L. (2018). Chemistry education research—From personal empiricism to evidence, theory, and informed practice. *Chemical Reviews, 118*(12), 6053–6087. https://doi.org/10.1021/acs.chemrev.8b00020

Davis, E. A., Beyer, C., Forbes, C. T., & Stevens, S. (2011). Understanding pedagogical design capacity through teachers' narratives. *Teaching and Teacher Education, 27*(4), 797–810. https://doi.org/10.1016/j.tate.2011.01.005

De Jong, O., Veal, W. R., & Van Driel, J. H. (2003). Exploring chemistry teachers' knowledge base. In J. K. Gilbert, O. De Jong, R. Justi, D. F. Treagust, & J. H. Van Driel (Eds.), *Chemical education: Towards research-based practice* (pp. 369–390). Springer. https://doi.org/10.1007/0-306-47977-X_16

Dorji, K., & Tshering, P. (2024). Examining Bhutanese science teachers' epistemic views of scientific inquiry. *Science & Education*. https://doi.org/10.1007/s11191-024-00572-6

Effendi, D. N., Irwandani, Anggraini, W., Jatmiko, A., Rahmayanti, H., Ichsan, I. Z., & Rahman, M. M. (2021). Bibliometric analysis of scientific literacy using VOS viewer: Analysis of science education. *Journal of Physics: Conference Series, 1796*(1), Article 012096. https://doi.org/10.1088/1742-6596/1796/1/012096

Emden, M., & Sumfleth, E. (2016). Assessing students' experimentation processes in guided inquiry. *International Journal of Science and Mathematics Education, 14*(1), 29–54. https://doi.org/10.1007/s10763-014-9564-7

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Feldon, D. F., Timmerman, B. C., Stowe, K. A., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching, 47*(10), 1165–1185. https://doi.org/10.1002/tea.20382

Furtak, E. M., Morrison, D. E. B., & Kroog, H. (2014). Investigating the link between learning progressions and classroom assessment. *Science Education, 98*(4), 640–673. https://doi.org/10.1002/sce.21122

Gipps, C. (2011). *Beyond testing (classic edition): Towards a theory of educational assessment* (1st ed.). Routledge. https://doi.org/10.4324/9780203182437

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences, 22*(4), 521–563. https://doi.org/10.1080/10508406.2013.837391

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education, 23*(5), 531–557. https://doi.org/10.1007/s10972-012-9278-z

Grob, R., Holmeier, M., & Labudde, P. (2019). Analysing formal formative assessment activities in the context of inquiry at primary and upper secondary school in Switzerland. *International Journal of Science Education, 43*(3), 407–427. https://doi.org/10.1080/09500693.2019.1663453

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics.

Heyde, V. V. D., & Siebrits, A. (2019). Higher-order e-assessment for physics in the digital age using Sakai. *Physics Teacher, 57*(1), 32–34. https://doi.org/10.1119/1.5084925

Hintze, J. M., Christ, T. J., & Methe, S. A. (2006). Curriculum-based assessment. *Psychology in the Schools, 43*(1), 45–56. https://doi.org/10.1002/pits.20128

Hung, C.-S., & Wu, H.-K. (2024). High school science teachers' assessment literacy for inquiry-based science instruction. *International Journal of Science Education, 46*(7), 621–642. https://doi.org/10.1080/09500693.2023.2251657

Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation*, *37*(1), 78–83. https://doi.org/10.1016/j.stueduc.2011.03.003

Kruit, P. M., Jansen, O. R., Ed, van de B., & Schuitema, J. A. (2018). Assessing students' ability in performing scientific inquiry: Instruments for measuring science skills in primary education. *Research in Science & Technological Education, 36*(4), 413–439. https://doi.org/10.1080/02635143.2017.1421530

Kuo, C.-Y., Hsin-Kai, W., Tsung-Hau, J., & Hsu, Y.-S. (2015). Development and validation of a multimedia-based assessment of scientific inquiry abilities. *International Journal of Science Education, 37*(14), 2326–2357. https://doi.org/10.1080/09500693.2015.1078521

Linstone, H. A. (1985). The Delphi technique. In V. T. Covello, J. L. Mumpower, P. J. M. Stallen, & V. R. R. Uppuluri (Eds.), *Environmental impact assessment, technology assessment, and risk analysis* (pp. 241–268). Springer. https://doi.org/10.1007/978-3-642-70634-9_22

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 95–132). Springer. https://doi.org/10.1007/0-306-47217-1_4

McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching, 48*(7), 745–770. https://doi.org/10.1002/tea.20423

McKenna, H. P. (1994). The Delphi technique: A worthwhile research approach for nursing? *Journal of Advanced Nursing, 19*(6), 1221–1225. https://doi.org/10.1111/j.1365-2648.1994.tb01207.x

Ministry of Education of the People's Republic of China. (2017). *Chemistry curriculum standards for senior high school*. People's Education Press.

National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Oliver, C. D., & Oliver, F. A. (2018). *Global resources and the environment*. Cambridge University Press. https://doi.org/10.1017/9781316779484

Özer, F., & Sarıbaş, D. (2023). Exploring pre-service science teachers' understanding of scientific inquiry and scientific practices through a laboratory course. *Science & Education, 32*(3), 787–820. https://doi.org/10.1007/s11191-022-00325-3

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). The nature of assessment and reasoning from evidence. In J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Knowing what students know: The science and design of educational assessment* (pp. 37–54). National Academies Press.

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59-81.

Petermann, V., & Vorholzer, A. (2022). Relationship between beliefs of teachers about and their use of explicit instruction when fostering students' scientific inquiry competencies. *Education Sciences, 12*(9), Article 593. https://doi.org/10.3390/educsci12090593

Petermann, V., & Vorholzer, A. (2023). Teachers' use of explicit instruction when planning lessons to foster students' scientific inquiry competencies. In G. S. Carvalho, A. S. Afonso, & Z. Anastácio (Eds.), *Fostering scientific citizenship in an uncertain world* (Vol. 13, pp. 219–234). Springer. https://doi.org/10.1007/978-3-031-32225-9_14

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*(3), 363–393. https://doi.org/10.1002/tea.21005

Reynolds, W. M., & Park, S. (2021). Examining the relationship between the Educative Teacher Performance Assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *58*(5), 721–748. https://doi.org/10.1002/tea.21676

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology*, *11*(1), 25–41. https://doi.org/10.1080/14780887.2013.801543

Rodriguez, J. M. G., Hunter, K. H., Scharlott, L. J., & Becker, N. M. (2020). A review of research on process oriented guided inquiry learning: Implications for research and practice. *Journal of Chemical Education, 97*(10), 3506–3520. https://doi.org/10.1021/acs.jchemed.0c00649

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground – A literature review of empirical research on scientific inquiry activities. *Studies in Science Education, 52*(2), 161–197. https://doi.org/10.1080/03057267.2016.1206351

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84. https://doi.org/10.1002/tea.20163

Sasse, H., Weber, A. M., Reuter, T., & Leuchter, M. (2025). Teacher guidance and on-the-fly scaffolding in primary school students' inquiry learning. *Science Education, 109*(2), 579–604. https://doi.org/10.1002/sce.21921

Shavelson, R. J., Yin, Y., Furtak, E. M., Ruiz-Primo, M. A., Ayala, C. C., Young, D. B., Tomita, M. K., Brandon, P. R., & Pottenger, F. (2008). On the role and impact of formative assessment on science inquiry teaching and learning. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing science learning: Perspectives from research and practice* (pp. 21–36). National Science Teachers Association.

Suh, J. K., & Park, S. (2017). Exploring the relationship between pedagogical content knowledge (PCK) and sustainability of an innovative science teaching approach. *Teaching and Teacher Education, 64*, 246–259. https://doi.org/10.1016/j.tate.2017.01.021

Sui, C. J., Hsiao, S. Y., Yeh, S. C., Zhao, P., Chang, C. Y., & Lin, J. (2024). Do they have inquiry skill profiles? Exploring high school students' scientific inquiry in an animation-based activity. *Science Education, 108*(2), 467–494. https://doi.org/10.1002/sce.21844

Talanquer, V., Tomanek, D., & Novodvorsky, I. (2013). Assessing students' understanding of inquiry: What do prospective science teachers notice? *Journal of Research in Science Teaching, 50*(2), 189–208. https://doi.org/10.1002/tea.21005

Teig, N. (2024). Uncovering student strategies for solving scientific inquiry tasks: Insights from student process data in PISA. *Research in Science Education, 54*(2), 205–224. https://doi.org/10.1007/s11165-023-10134-5

Van Driel, J. H., Hume, A., & Berry, A. (2023). Research on science teacher knowledge and its development. In N. G. Lederman, D. L. Zeidler, & J. S. Lederman (Eds.), *Handbook of research on science education* (pp. 1123–1161). Routledge. https://doi.org/10.4324/9780367855758

Vo, D. V., & Simmie, G. M. (2024). Assessing scientific inquiry: A systematic literature review of tasks, tools, and techniques. *International Journal of Science and Mathematics Education*, *23*(4), 871–906. https://doi.org/10.1007/s10763-024-10498-8

Wang, J.-R., Kao, H.-L., & Lin, S.-W. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education, 26*(3), 522–529. https://doi.org/10.1016/j.tate.2009.06.014

Wu, P. H., Wu, H. K., & Hsu, Y. S. (2014). Establishing the criterion-related, construct, and content validities of a simulation-based assessment of inquiry abilities. *International Journal of Science Education*, *36*(10), 1630–1650. https://doi.org/10.1080/09500693.2013.871660

Yamamoto, K., Shin, H. J., & Khorramdel, L. (2018). Multistage adaptive testing design in international large-scale assessments. *Educational Measurement: Issues and Practice, 37*(4), 16–27. https://doi.org/10.1111/emip.12226

Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation, 74*, Article 101183. https://doi.org/10.1016/j.stueduc.2022.101183

Yan, Z., Li, Z., Pastore, E., Yu, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice, 28*(3), 228–260. https://doi.org/10.1080/0969594X.2021.1884042

Yang, Y., Zhu, G., Sun, D., & Chan, C. K. K. (2022). Collaborative analytics-supported reflective assessment for scaffolding pre-service teachers' collaborative inquiry and knowledge building. *International Journal of Computer-Supported Collaborative Learning, 17*(2), 249–292. https://doi.org/10.1007/s11412-022-09372-y

Zachos, P., Hick, T. L., Doane, W. E. J., & Sargent, C. (2000). Setting theoretical and empirical foundations for assessing scientific inquiry and discovery in educational programs. *Journal of Research in Science Teaching, 37*(9), 938–962. https://doi.org/10.1002/1098-2736(200011)37:9<938::AID-TEA5>3.0.CO;2-S

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Appendix 1

Instrument

Dear Teachers,
We sincerely invite you to participate in a study on the understanding of scientific inquiry assessment concepts. After viewing relevant videos or reading textual materials, we kindly ask you to complete an evaluation. The evaluation questionnaire consists of two parts: the first part collects basic information and contains 13 multiple-choice questions; the second part assesses the concepts related to scientific inquiry evaluation for chemistry teachers, comprising 21 questions that include multiple-choice, fill-in-the-blank, and short answer types. The estimated time to complete the evaluation is approximately 40 minutes. All data will be strictly confidential and used solely for academic research, processed in a completely anonymous manner. Your participation is crucial to the success of this study, and we extend our sincerest gratitude to you!

Part One: Basic Information

Gender
A. Male
B. Female
Type of School
A. Urban (including central urban areas and urban-rural fringe)
B. Town (including central town areas and special regions)
C. Rural (including rural central areas and villages)
Teaching Level
A. Primary School
B. Middle School
C. High School
In the past year, how many scientific inquiry classes have you designed and conducted?
A. 0 classes (never)
B. 1-2 classes (rarely)
C. 3-5 classes (sometimes)
D. 6-10 classes (occasionally)
E. More than 10 classes (frequently)

Part Two: Assessment of Chemistry Teachers' Understanding of Scientific Inquiry Evaluation Concepts

(I) Understanding the Content of Scientific Inquiry Assessment
Is "The effect of vitamin C on dough" (Video A) considered scientific inquiry? Please explain your reasoning.
What aspects of scientific inquiry skills does "The effect of vitamin C on dough" (Video A) demonstrate?
Is "Exploring the vitamin C content in spinach and oranges" (Video B) considered scientific inquiry? Please explain your reasoning.
What aspects of scientific inquiry skills does "Exploring the vitamin C content in spinach and oranges" (Video B) demonstrate?
What is the purpose of the discussion on the experimental results of "How to compare vitamin C content in different fruit juices" (Video D), and which content requirements of the chemistry curriculum standards does it correspond to? *(Students compared the vitamin C content in oranges and lemons through semi-quantitative analysis, illustrated by the following teacher-student dialogue):*
Teacher: The experiment is over; I would like each group to report their findings.
Student A: 19 drops for lemon, 22 drops for orange.
Student B: 22 drops for lemon, 23 drops for orange juice.
Student C: 25 drops for both lemon and orange juice.
Teacher: What conclusions can we draw from these three sets of experiments?
Student: Lemons have slightly higher vitamin C content than oranges.
Teacher: This aligns with our initial hypothesis; does anyone have further questions?
(II) Clarifying the Purpose of Scientific Inquiry Assessment

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

What types of assessment are present in the discussion of the experimental results from "How to compare vitamin C content in different fruit juices" (Video D)?
A. Summative Assessment
B. Formative Assessment
C. Diagnostic Assessment
D. All above
What is the purpose of the evaluation tool design in the scientific inquiry activity "How to compare vitamin C content in different fruit juices"? *(Short answer)*
Which evaluation methods are included in "The effect of vitamin C on dough" and "Exploring the vitamin C content in spinach and oranges" (Videos A and B) and to which type of assessment do they belong?
(1) Posing questions
(2) Answering questions
(3) Observing data
(4) Group discussions
(5) Experimental work sheets
(6) Experiment reports
(7) Oral reports
(8) Written tests
(9) Inquiry projects
(10) Pre-class tests.
Formative Assessment: _____;
Diagnostic Assessment: _____;
Summative Assessment: _____.
(III) Designing Scientific Inquiry Assessment Strategies
In the principal discussion of "Determining the vitamin C content in fruits" (Video C), what evaluation strategies did the instructor integrate?
A. Real-time formative assessment (interactive Q&A)
B. Process observation assessment (experimental records)
C. Collaborative efficiency assessment (group discussion)
D. Visualizing thought assessment (work sheet)
E. Structured expression assessment (experiment report)
F. Academic expression assessment (oral report)
G. Transfer application assessment (written test)
H. Innovative competency assessment (inquiry project)
The inquiry work sheet is an important assessment tool. What key content should be included in the work sheet for "Determining the vitamin C content in fruits"? *(Multiple choices)*
A. Description of the inquiry question
B. Task-oriented design
C. Explanation of experimental principles
D. List of experimental materials
E. Steps of experimental procedures
F. Data recording system
G. Evaluation and reflection
If you were the instructor in the principal discussion of "Determining the vitamin C content in fruits" (Video C), which evaluation strategies would you choose?
A. Group cooperative discussions (peer assessment)
B. Self-study through literature research (self-assessment)
C. Teacher-student Q&A (teacher assessment)
D. None of the above
From which aspects would you evaluate the work sheet for "Comparing vitamin C content in different fruits"? *(Refer to the case)*
(IV) Assessing and Providing Feedback on Scientific Inquiry Results
What aspects of the experimental process from "The effect of vitamin C on dough" (Video A) need improvement?

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

A. The quality and moisture content of the dough were not controlled (specific ratios and control methods should be clarified).
B. The elasticity was only compared between experimental groups at 0 ppm and 100 ppm, without comparing with a control group at 30 ppm.
C. The results were singular, focusing only on the effects of VC on dough elasticity and appearance.
D. The experimental design is thorough and does not require improvement.
What aspects of the experimental process for "Exploring the vitamin C content in spinach and oranges" (Video B) need improvement?
A. Replacing the rubber bulb pipette with a graduated pipette allows for accurate quantification by students.
B. Using a juicer instead of a mortar and pestle would yield a larger volume of juice.
C. Suggest establishing a blank control group.
D. The existing method is scientifically valid and does not require improvement.
During a hot summer, when opening a bottle of cola, one can observe many bubbles escaping. Teacher Z leads students to propose scientific inquiry questions based on this phenomenon, integrating their existing knowledge (such as gas solubility and the properties of carbon dioxide) and life experiences. The following are questions posed by four students:
Student A: What gas is present in the bubbles that escape from the cola?
Student B: How do temperature and pressure affect the solubility of gases in water?
Student C: Why are people more inclined to drink carbonated beverages in summer?
Student D: Is there a difference in carbon dioxide content among different brands of soda? How does this difference affect the rate of bubble formation?
Please assess the proposed questions based on the levels of inquiry abilities as per the "Core Competency of Scientific Inquiry and Innovation Awareness" in the General High School Chemistry Curriculum Standards (Levels 1 to 4), indicating the appropriate ability levels for each question along with justification.
Student A Level: _____; Reason: _____.
Student B Level: _____; Reason: _____.
Student C Level: _____; Reason: _____.
Student D Level: _____; Reason: _____.
Level 1: Able to pose simple questions.
Level 2: Able to pose questions based on curriculum content.
Level 3: Possesses a strong issue awareness and can propose inquiry-based questions in discussions with peers.
Level 4: Able to propose comprehensive inquiry topics based on literature and practical needs.
When an experimental group investigates the reaction "$Fe + 2Ag^+ \rightarrow Fe^{2+} + 2Ag$", they accidentally detect the presence of Fe^{3+} in the solution. The investigative process is as follows: excess iron powder is added to a 0.05 mol/L silver nitrate solution acidified with dilute nitric acid (pH ≈ 2), mixed, and allowed to stand. A black solid precipitates at the bottom of the beaker, and the solution gradually turns yellow. Concerning the "yellow solution" phenomenon, Student A believes that Fe^{3+} is present, while Student B questions how Fe^{3+} could be stable in excess iron powder. To verify their viewpoint, the upper liquid is sampled and KSCN solution is added, resulting in a red color (confirming the presence of Fe^{3+}). Meanwhile, a white precipitate is formed (it is known that Ag^+ reacts with SCN^- to form white precipitate AgSCN), and the amount of precipitate varies over the sampling time. The students propose the following hypotheses regarding the cause of Fe^{3+} production:
Student A: An oxidized layer (such as Fe_3O_4) exists on the surface of the iron powder, which reacts with the acidic solution to produce Fe^{3+};
Student B: The Fe^{2+} in the solution is oxidized by O_2 in the air ($4Fe^{2+} + O_2 + 4H^+ \rightarrow 4Fe^{3+} + 2H_2O$);
Student C: Under acidic conditions, NO_3^- (from the acidification) possesses strong oxidizing properties, oxidizing Fe^{2+} to Fe^{3+};
Student D: Residual Ag^+ (due to Ag^+ reacting with SCN^- to form AgSCN precipitate) may react with Fe^{2+} to produce Fe^{3+}.
Please evaluate the corresponding levels of ability in "Formulating Hypotheses and Assumptions" as per the levels outlined in the General High School Chemistry Curriculum Standards (Levels 1 to 4), providing justification for each.
Student A Level: _____; Reason: _____.
Student B Level: _____; Reason: _____.
Student C Level: _____; Reason: _____.
Student D Level: _____; Reason: _____.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

In the Curriculum Standards Level 4: Responses must be articulated using observable and measurable terminology; must be precisely articulated (not vague or broad); must identify the relevant variables for the research problem; the variables in the hypothesis must be operational.

An industrial wastewater containing hexavalent chromium (presented as $K_2Cr_2O_7$) can be reduced to Cr^{3+} using oxalic acid ($H_2C_2O_4$) under light conditions. Research has shown that small amounts of iron alum [$Al_2Fe(SO_4)_4{\cdot}24H_2O$] catalyze this reaction. To investigate the catalyst components in iron alum, the research team hypothesized that Fe^2 is the key component of the catalytic reaction. Student A's experimental design is illustrated in Table 1 (with other conditions such as temperature and light intensity kept constant). Please assess the level of "Experimental Design" ability according to the requirements stated in the General High School Chemistry Curriculum Standards (Levels 1 to 4).

The "Experimental Design" ability reaches Level _____; Reason: _____.

(Note: Level 1, Design a simple experimental plan and complete the experimental operations; Level 2, Design an experimental plan based on the hypothesis and collaborate with classmates to complete the operations; Level 3, Propose an experimental plan based on the hypothesis and independently complete the experiment; Level 4, Propose multiple inquiry plans based on the hypothesis, evaluate, and optimize the plans.)

A student notices a layer of green patina on the copper faucet used at home and learns through research that the primary component of this patina is copper carbonate [$Cu_2(OH)_2CO_3$]. In addition to oxygen and water, carbon dioxide is also necessary for copper to corrode. Please design an experiment to investigate the conditions for copper corrosion. Student B's experimental design can be seen in Figure 2 (distilled water is used, and it should be boiled prior to adding to the test tube). Please assess the level of "Experimental Design" ability according to the requirements of the General High School Chemistry Curriculum Standards (Levels 1 to 4).

The "Experimental Design" ability reaches Level _____; Reason: _____.

Appendix 2

Video A Transcript

Video A: Duration: 2 minutes and 11 seconds
Theme: The Effects of vitamin C on dough elasticity
Role: Narration

Vitamin C, also known as ascorbic acid, acts as a reducing agent. When added to flour, it is oxidized to dehydroascorbic acid, thereby transforming into an oxidizing agent. So, what effect does vitamin C have on dough? Let's conduct an experiment.

First, we will accurately measure 0.100 g of vitamin C using a precise electronic scale and mix it with flour in a kneading machine. The control group will have 30 ppm of vitamin C added, while experimental group (a) will have 100 ppm of vitamin C, and experimental group (b) will not contain any vitamin C.

Next, we will compare the elasticity of the dough after kneading. The elasticity of the dough in experimental group (a) is higher than that of experimental group (b). Specifically, the dough in group (a) rebounds from 20 cm to 15 cm, while the dough in group (b) rebounds from 20 cm to 17 cm, indicating that adding vitamin C improves the rebound elasticity of the dough.

Next, we will divide the dough into equal portions and compare the elasticity after resting. The diameter of the dough from group (a) is 12.4 cm, while the diameter from group (b) is 11.9 cm. Upon shaping, the elasticity of the dough in group (b) is soft and relaxed, while the control group exhibits some tension and tightness; in contrast, the dough in group (a) has burst open.

After fermentation, the heights of the dough are as follows: the height of the dough in experimental group (a) is 15.5 cm, group (b) is 14.9 cm, and the control group is 15.1 cm. What differences can we observe in the final products? In group (a), the bread cracks and tears at the sides, exhibiting a rough crust. In group (b), the bread surface shows similar cracks but has numerous small air holes. The crust of the control group bread has a rugged beauty with visible cracks, and its surface is smooth.

The bread from experimental group (a) is derived from a highly elastic dough that proudly maintains its shape. In contrast, the bread from group (b) arises from a less elastic dough that freely spreads within the mold, filling the corners and resulting in sharp edges. The control group dough has moderate elasticity, allowing it to maintain a balanced stretch and contraction.

https://doi.org/10.33225/jbse/25.24.774

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Upon slicing the bread, the air pockets in group (a) are irregular, with dark spots of sediment, while the air pockets in group (b) are round and lack clear direction. The control group's bread features vertical air pockets with a defined structure and a glossy finish.

In summary, vitamin C enhances the elasticity of dough, and this effect is similarly applicable to other oxidizing agents.

Received: April 14, 2025 Revised: June 02, 2025 Accepted: July 23, 2025

Cite as: Zhu, Y., Chi, S., & Wang, Z. (2025). Chemistry teachers' knowledge in assessing scientific inquiry competence. *Journal of Baltic Science Education*, *24*(4), 774–793. https://doi.org/10.33225/jbse/25.24.774

Ya Zhu PhD Candidate, School of Chemistry and Molecular Engineering, East China Normal University; Lecturer, Xinjiang Normal University, No. 3663, Zhongshan North Road, East China Normal University, Shanghai, P. R. China.
E-mail: zhuya0624@hotmail.com
ORCID: https://orcid.org/0000-0001-7565-0073

Shaohui Chi *(Corresponding author)* Professor, School of Chemistry and Molecular Engineering, East China Normal University, No. 3663, Zhongshan North Road, Shanghai, P. R. China.
E-mail: charlcy@163.com
ORCID: https://orcid.org/0000-0002-1261-7951

Zuhao Wang Professor, School of Chemistry and Molecular Engineering, East China Normal University, No. 3663, Zhongshan North Road, Shanghai, P. R. China.
E-mail: wangzuhao@126.com
ORCID: https://orcid.org/0000-0002-2967-8232

https://doi.org/10.33225/jbse/25.24.774