Journal of Educational Research and Pedagogical Innovation
Vol. 1 No. 1, June 2026, pp.35-46
https://https://www.jerpi.bin-foundation.org/index.php/jerpi
DOI: https://

Design and Evaluation of Virtual Laboratory-Based Chemistry Learning to Improve Scientific Literacy and Conceptual Understanding

Mahdian[1]*, Parham Saadi[1]
[1]Department of Chemistry Education, Lambung Mangkurat University, Banjarmasin, Indonesia
*Corresponding Author: mahdian_kimia@ulm.ac.id

ABSTRACT

Conventional undergraduate chemistry practicum continues to face unresolved challenges, including limited reagent availability, safety hazards in handling reactive alkali metals, and persistently low student scientific literacy and conceptual understanding. This study designed, implemented, and evaluated a multimedia virtual laboratory approach integrating Crocodile Chemistry (macroscopic and symbolic levels) with PhET Interactive Simulations: Neon Lights & Other Discharge Lamps (submicroscopic level) to address all three nodes of Johnstone's Chemistry Triplet in a Metal Chemistry Practicum course. A quasi-experimental one-group pretest-posttest design was employed with 32 undergraduate students. Instruments included a PISA 2022-based scientific literacy test (OECD, 2023), a revised Bloom's Taxonomy C1–C4 conceptual understanding test, a learning implementation observation sheet, and a student perception questionnaire. Validity was confirmed via Aiken's V (≥0.75) and reliability via Cronbach's alpha (≥0.70). Data were analyzed using normalized N-Gain and the Wilcoxon Signed-Rank Test ($\alpha = 0.05$). The results indicated improvements in scientific literacy (N-Gain = 0.54; r = 0.88) and conceptual understanding (N-Gain = 0.64; r = 0.87), both statistically significant ($p < 0.001$) with large effect sizes. Implementation fidelity reached 94.6%, and student perceptions were strongly positive (mean = 4.36/5). These findings suggest potential effectiveness of sequentially integrating Crocodile Chemistry and PhET, mediated by a Virtual Student Worksheet that explicitly scaffolds cross-representational connections, as an instructional approach for improving scientific literacy and conceptual understanding in university chemistry education. Future studies with control-group designs are needed to establish stronger causal claims.

Keywords: conceptual understanding; crocodile chemistry; PhET interactive simulations; scientific literacy; virtual laboratory

Article history

Received: 19 Mei 2026	*Revised:27 Juni 2026*	*Accepted:30 Juni 2026*	*Published:30 Juni 2026*
date	*date*	*Date*	*date*

Citation (APA Style): Mahdian, & Saadi, P. (2026). Design and evaluation of virtual laboratory-based chemistry learning to improve scientific literacy and conceptual understanding. *Journal of Educational Research and Pedagogical Innovation*, *1*(1), 35-46.

INTRODUCTION

Chemistry is an inherently multi-representational discipline that requires learners to navigate three qualitatively distinct yet interrelated levels of representation simultaneously: the macroscopic level, referring to directly observable phenomena; the submicroscopic level, encompassing particulate and atomic mechanisms; and the symbolic level, including chemical notation, equations, and graphs. Collectively, these three levels constitute

Johnstone's Chemistry Triplet, a framework formalized by Johnstone (1991, 1993) and widely recognized as the most fundamental epistemological challenge in chemistry education. When students fail to integrate all three levels coherently, they tend to revert to shallow, algorithm-dependent learning rather than developing transferable scientific knowledge (Permatasari et al., 2022; Stipek et al., 2024). This challenge is directly consequential for chemistry practicum quality, where students frequently complete laboratory procedures mechanically without a genuine understanding of the underlying mechanisms.

Among the most strategically important learning outcomes in chemistry education, scientific literacy and conceptual understanding occupy inseparable positions. Scientific literacy, as operationalized by the PISA 2022 framework (OECD, 2023), encompasses three interdependent competencies: explaining phenomena scientifically, evaluating and designing scientific enquiry, and interpreting data and scientific evidence. These competencies are unlikely to develop from passive exposure to lectures or textbooks alone; they require authentic, interactive, and inquiry-driven learning experiences (Afina et al., 2021; Ambarwati et al., 2023; Bybee, 2011). Conceptual understanding constitutes the cognitive architecture upon which meaningful scientific literacy is built: without mechanistic understanding of why phenomena occur, a student's capacity to evaluate and produce scientifically valid explanations will remain constrained (Ristanti & Sumarti, 2024; Bruce et al., 2022). Both outcomes must therefore be cultivated simultaneously through deliberate instructional design.

Among all topics in undergraduate Metal Chemistry Practicum, the identification of metal ions through flame tests is simultaneously the most conceptually rich and the most operationally challenging. At the macroscopic level, the experiment involves observing characteristic emission colors produced when alkali metal compounds are heated: red for Li^+ ($\lambda \approx 670$ nm), yellow for Na^+ ($\lambda \approx 589$ nm), violet for K^+ ($\lambda \approx 404$ nm), reddish-purple for Rb^+ ($\lambda \approx 421$ nm), and blue for Cs^+ ($\lambda \approx 455$ nm). Beneath these observable colors lies an entirely submicroscopic mechanism: atomic electrons absorb thermal energy, become excited to higher-energy orbitals, and then emit photons of precisely quantized energy when returning to the ground state — energy satisfying the equation $\Delta E = hf = hc/\lambda$. This atomic mechanism explains why each metal produces a distinct color, yet conventional practicum instruction rarely succeeds in conveying this mechanistic understanding. Preliminary classroom observation indicated that a substantial proportion of students enrolled in the Metal Chemistry Practicum were unable to explain the atomic mechanism underlying the flame colors they observed, even after completing a full conventional practicum session. These conceptual difficulties are further compounded by significant operational constraints, including the high cost of high-purity alkali metal reagents, safety hazards in handling compounds that react violently with water and air, and unfavorable student-to-equipment ratios that frequently lead to cross-contamination between samples (Kurniawati & Rahmawati, 2024; Manyilizu, 2023; Raziana et al., 2025; Felder & Brent, 2016).

Multimedia virtual laboratory technology has emerged as a technologically grounded and empirically supported response to the limitations of conventional practicum settings. A virtual laboratory is broadly defined as a computer-based instructional medium that presents a digital representation of a real laboratory environment, enabling students to explore chemical phenomena with precision, repeat procedures without material constraints, and visualize processes otherwise inaccessible to direct observation (Lestari, 2023; Potkonjak et al., 2016). The empirical basis for this approach is substantial: a meta-analysis by Li et al. (2024) covering 46 controlled studies reported significant positive effects on learning outcomes (Hedges' $g = 0.686$; 95% CI [0.414, 0.959]). A systematic review by Wörner et al. (2022), drawing on 64 studies, concluded that combining virtual and real experiments tends to produce deeper conceptual understanding than either approach used alone. In the chemistry domain, Fadilah and Sukmawati (2022) demonstrated that virtual laboratory activities structured with inquiry-based worksheets were associated with improved science process skills, while Pane et al. (2024) reported enhanced scientific literacy through a STEM-integrated virtual laboratory platform. However, the potential effectiveness of any virtual laboratory platform is constrained by the extent to which it facilitates all levels of Johnstone's Chemistry Triplet — and it is here that a critical gap emerges. Crocodile Chemistry accurately simulates the flame colors of alkali metals and provides quantitative emission wavelength data, making it a suitable medium for the macroscopic and symbolic levels. It does not, however, provide animation of electron movement between orbitals, representation of photon emission, or interactive energy-level diagrams. This absence of submicroscopic visualization constitutes a structural pedagogical gap that may prevent the formation of a complete mechanistic understanding when Crocodile

Chemistry is used as the sole instructional medium (Widarti et al., 2022; Serrano-Pérez et al., 2023; Rutten et al., 2012).

To address this submicroscopic gap, the present study integrates PhET Interactive Simulations — specifically Neon Lights & Other Discharge Lamps (University of Colorado Boulder) — as a complementary medium grounded in empirically validated cognitive and pedagogical principles, with the explicit goal of making the invisible visible (Wieman et al., 2008). This simulation enables students to observe in slow-motion the excitation of electrons to higher energy levels and the subsequent emission of colored photons as electrons return to the ground state — precisely the atomic processes underlying flame test phenomena yet inaccessible through conventional or Crocodile Chemistry-based instruction. Finkelstein et al. (2005) demonstrated that for topics requiring multi-level visualization, PhET simulations may match or exceed the effectiveness of physical laboratory equipment. The sequential integration of both platforms, mediated by a Virtual Student Worksheet (VSW) that explicitly guides students in connecting macroscopic Crocodile Chemistry data with submicroscopic PhET energy diagrams and corresponding symbolic equations, represents an application of the Multiple External Representations (MER) principle: the potential benefits of multiple representations are contingent on the quality of scaffolding that explicitly connects those representations (Treagust et al., 2017; Chang & Linn, 2013).

Despite the growing body of research on virtual laboratories, most existing studies examine single platforms, single affective outcomes, or single levels of Johnstone's Chemistry Triplet, without simultaneously targeting scientific literacy and conceptual understanding within an integrated multimedia design (Haryanti et al., 2023; Sreekanth et al., 2022). Furthermore, no prior study has explicitly designed and evaluated the integration of two complementary virtual platforms operating at different representational levels within the specific context of undergraduate Metal Chemistry Practicum. The present study was designed to address this gap. Specifically, the study pursued four objectives: (1) to design a multimedia virtual laboratory learning scenario integrating Crocodile Chemistry and PhET for the topic of alkali metal ion identification through flame tests; (2) to implement the designed scenario in the Metal Chemistry Practicum course; (3) to evaluate its association with improvements in students' scientific literacy based on the PISA 2022 competency framework (OECD, 2023); and (4) to evaluate its association with improvements in students' conceptual understanding based on the revised Bloom's Taxonomy. The findings are intended to contribute a replicable, evidence-informed instructional model for technology-mediated chemistry practicum in Indonesian higher education.

Table 1. Comparison of conventional laboratory, crocodile chemistry, and PhET discharge lamps

Aspect	Conventional Laboratory	Crocodile Chemistry	PhET Discharge Lamps
Safety	High risk: reactive alkali metals, corrosive chemicals, open flame	Safe; no real chemicals required	Safe; browser-based, no installation required
Cost	High: reagents, equipment, maintenance	Licensed software and computers	Free and open-access (phet.colorado.edu)
Johnstone's Level	Macroscopic (direct observation only)	Macroscopic + Symbolic	Submicroscopic (real-time electron animation)
Electron visualization	Not observable	Not available	Real-time animation of excitation and photon emission; slow-motion mode available
Interactivity	Limited to prescribed procedures	Repeatable procedure simulation	Full user control: voltage, atom type, and speed
Evidence base	Standard laboratory protocols	Commercial chemistry simulation software	Hundreds of indexed publications (Wieman et al., 2008)

Synthesized from de Jong et al. (2013), Potkonjak et al. (2016), Wieman et al. (2008), and Li et al. (2024)

METHOD

Research Design

This study employed a quasi-experimental one-group pretest-posttest design (Creswell, 2014) with supplementary qualitative data. The quantitative component constituted the primary design because the study examined changes in students' scientific literacy and conceptual understanding before and after the implementation of the multimedia virtual laboratory. It should be acknowledged that the absence of a control group is a recognized limitation of this design: causal inferences cannot be drawn, and observed improvements may be attributable to factors beyond the intervention alone. This design was adopted on three contextual grounds: ethical (withholding the intervention from enrolled students was inappropriate given the practicum's curricular status), practical (only one intact class was available during the study period), and analytical (normalized N-Gain analysis provides a standardized measure of improvement relative to maximum possible gain within a single group). Qualitative data from open-ended questionnaire responses and structured observation notes were analyzed inductively to explain and contextualize the quantitative findings through triangulation; therefore, the qualitative evidence functioned as supportive explanatory data rather than as a separate sequential explanatory mixed-methods design. The study was conducted in accordance with the ethical standards of Universitas Lambung Mangkurat, and institutional approval was obtained prior to data collection.

Participants

The research participants were 32 undergraduate students enrolled in the Metal Chemistry Practicum course at the Department of Chemistry Education, Faculty of Teacher Training and Education, Universitas Lambung Mangkurat, during the second semester (even semester) of the 2025/2026 academic year. Purposive sampling was employed because the study targeted a specific, naturally formed instructional group (an intact class) rather than a random cross-section of the student population; this approach is appropriate when the research context requires studying a particular educational setting under real conditions (Creswell, 2014). All students enrolled in the course during the study period met the inclusion criteria and were invited to participate. Participants were included based on two criteria: (a) active enrollment in the Metal Chemistry Practicum course during the study period; and (b) prior completion of General Chemistry I and II as prerequisite courses. Participants were informed that their involvement was entirely voluntary, that their decision to participate or withdraw would have no effect on their academic standing, and that all data would be anonymized, stored securely, and used solely for research purposes. No participants withdrew during the study, yielding a complete dataset for all 32 individuals.

Multi-media Virtual Laboratory Learning Scenario

The learning scenario was designed in eight sessions using a structured inquiry approach (Lazonder & Harmsen, 2016) and was implemented across two face-to-face class meetings of 3 × 50 minutes each. Structured inquiry was considered more appropriate than guided inquiry in this study because the problem context, procedural sequence, simulation tools, data tables, and representational-linking prompts were provided in advance through the Virtual Student Worksheet (VSW), while students were required to analyze evidence, compare representations, and construct conceptual explanations. In this study, the term "session" refers to a distinct learning phase or sub-activity within a class meeting, not a separate instructional period; each meeting therefore comprised four sequential sessions totaling 150 minutes of instructional time. The overarching learning objectives for the scenario were: (a) to correctly identify alkali metal ions based on characteristic flame emission colors; (b) to explain the submicroscopic mechanism of electron excitation and photon emission underlying the flame test; and (c) to express the relationship between electron energy transitions and photon energy using the symbolic equation $\Delta E = hf = hc/\lambda$.

The VSW contained concrete structured tasks organized into four task blocks. First, in the macroscopic observation task, students recorded flame colors and wavelength values from Crocodile Chemistry for LiCl, NaCl, KCl, RbCl, and CsCl. Second, in the symbolic-processing task, students calculated photon energy using $\Delta E = hc/\lambda$ and arranged the metals according to wavelength and energy differences. Third, in the submicroscopic interpretation task, students observed PhET Discharge Lamps in slow-motion mode and identified electron excitation, electron

relaxation, photon emission, and the energy-level transition involved in each simulated event. Fourth, in the cross-representational mapping task, students completed a matching matrix that linked Crocodile Chemistry flame-color data with PhET energy-level diagrams and then wrote short explanations connecting color, wavelength, energy gap, and electron transition.

To make the Crocodile Chemistry–PhET connection explicit, the VSW instructed students to begin with the macroscopic flame color and wavelength recorded in Crocodile Chemistry, convert the wavelength into relative photon energy using $\Delta E = hc/\lambda$, locate the corresponding energy-gap pattern in the PhET diagram, and explain why different alkali metal ions emit different colors. Students were also required to answer comparison prompts, for example why shorter wavelengths correspond to larger photon energies and how the observed flame color reflects electron relaxation from an excited state to the ground state.

The lecturer's role remained facilitative rather than transmissive: introducing each session with a brief orientation (≤5 minutes), ensuring that students followed the structured VSW sequence, circulating to monitor and scaffold student progress, and leading a whole-class synthesis discussion at the end of each meeting. The VSW was validated as part of the learning scenario and learning media package, separately from the learning implementation observation sheet. Its validation focused on task clarity, alignment with Johnstone's macroscopic–submicroscopic–symbolic representation levels, accuracy of chemical content, readability of instructions, and suitability for undergraduate practicum learning.

Meeting 1 — Crocodile Chemistry (Macroscopic + Symbolic Level):

- Session 1: Interface orientation and simulation calibration (objectives: familiarize with software controls; verify simulation accuracy against known emission wavelengths)
- Session 2: Flame test simulation for LiCl and NaCl — identification of Li^+ and Na^+; recording colors and wavelengths in VSW
- Session 3: Flame test simulation for KCl, RbCl, and CsCl — identification of K^+, Rb^+, and Cs^+; comparing emission spectra
- Session 4: Identification of unknown sample solutions using accumulated reference data; completing VSW macroscopic summary table

Meeting 2 — PhET Discharge Lamps (Submicroscopic Level + Three-Level Synthesis):

- Session 5: PhET orientation — observing single-electron transitions in slow-motion; identifying excitation and photon emission events
- Session 6: Multi-atom exploration — comparing energy-level diagrams across alkali metal atoms; connecting photon color to energy gap magnitude
- Session 7: Explicit cross-media connection via VSW — mapping Crocodile Chemistry macroscopic flame color data onto PhET submicroscopic energy diagrams for each alkali metal
- Session 8: Symbolic confirmation — deriving and verifying the equation $\Delta E = hf = hc/\lambda$ using combined data from both platforms; completing VSW symbolic section

Research Instruments

Four instruments were employed. Content validity was established by three expert validators (Validator 1: inorganic chemistry content; Validator 2: educational evaluation and test construction; Validator 3: learning media and technology) using Aiken's V coefficients, with $V \geq 0.75$ as the acceptance threshold (Aiken, 1985). Internal consistency reliability was assessed using Cronbach's alpha ($\alpha \geq 0.70$). For the observation sheet, inter-rater reliability was assessed using Cohen's kappa (κ) based on independent ratings by two trained observers during a pilot session; $\kappa \geq 0.75$ was set as the minimum acceptable level of agreement (Landis & Koch, 1977).

In addition to the four research instruments, the VSW was reviewed by the same expert validators as an instructional scaffolding document. Because the VSW was not used to score student outcomes or implementation fidelity, it was not included as a separate measurement instrument in Table 2; instead, its validity evidence was reported as part of the learning scenario and learning media validation.

The scientific literacy test comprised 14 structured essay items (five for explaining phenomena, four for designing enquiry, five for interpreting data and evidence), each scored on a rubric with four criteria: scientific accuracy (0–5 points), completeness of explanation (0–5 points), use of appropriate terminology (0–3 points), and logical coherence (0–2 points). Total raw scores were converted to a 0–100 scale using the formula: converted score = (raw score / maximum possible raw score) × 100. The conceptual understanding test comprised 15 multiple-choice items (eight at C1–C2 level; seven at C3–C4 level), each scored dichotomously (1 = correct; 0 = incorrect), with total scores converted to a 0–100 scale. The student perception questionnaire comprised 20 items rated on a five-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree), covering ease of use, visualization clarity, cross-level connection quality, motivation, and independent learning capacity for both platforms. The learning implementation observation sheet comprised 15 behavioral indicators, each rated on a four-point scale (1–4), yielding a maximum score of 60; implementation fidelity was expressed as a percentage of the maximum, calculated as (total score obtained / 60) × 100, averaged across the two independent observer ratings for each session.

Table 2. Research instrument blueprint

Variable	Indicator	Instrument Type	Items
Scientific Literacy	Explaining phenomena scientifically (PISA C1)	Essay Test	5
	Evaluating and designing scientific enquiry (PISA C2)	Essay Test	4
	Interpreting data and scientific evidence (PISA C3)	Essay Test	5
Conceptual Understanding	C1–C2: Recall and Comprehension	Multiple Choice	8
	C3–C4: Application and Analysis	Multiple Choice	7
Student Perception	Ease of use, motivation, cross-level connection (Crocodile + PhET)	Likert Scale	20
Learning Implementation	Fidelity of the eight-session scenario	Observation Sheet	15

Table 3. Content validity results – Aiken's V coefficients

Instrument	Validation Aspect	No. of Items	Aiken's V Range	Mean V	Decision
Scientific Literacy Test	Content relevance, language clarity, cognitive level alignment	14	0.79–0.93	0.87	Valid
Conceptual Understanding Test	Content relevance, construct alignment, distractor quality	15	0.78–0.91	0.85	Valid
Perception Questionnaire	Statement clarity, construct relevance, scale appropriateness	20	0.80–0.92	0.86	Valid
Observation Sheet	Indicator clarity, behavioral observability, completeness	15	0.78–0.90	0.84	Valid

Data Analysis

Improvements in scientific literacy and conceptual understanding were quantified using Hake's (1998) normalized gain formula:

$$g = \frac{S_{posttest} - S_{pretest}}{S_{max} - S_{pretest}}$$

Where:
g = N-Gain value
S_{post} = post-test score
S_{pre} = pre-test score
S_{max} = ideal maximum score

Table 4. Interpretation criteria for n-gain

N-Gain Value (g)	Category
$g > 0.70$	High
$0.30 \leq g \leq 0.70$	Moderate
$g < 0.30$	Low

(Hake, 1998)

The Shapiro-Wilk test indicated that the gain score distributions for both variables violated the normality assumption ($p < 0.05$); accordingly, the Wilcoxon Signed-Rank Test (non-parametric; $\alpha = 0.05$) was used for inferential analysis. Effect size was calculated as $r = |Z|/\sqrt{n}$ using the 32 paired observations, where $r > 0.50$ is classified as a large effect according to Cohen's (1988) conventions. Student perception data were analyzed descriptively using means and percentage distributions by category. Qualitative data from observation notes and open-ended responses were analyzed using inductive thematic analysis.

RESULTS AND DISCUSSION

Instrument Validity and Reliability

Prior to implementation, the eight-session multimedia virtual laboratory scenario (Objective 1) was reviewed by the three expert validators alongside the research instruments. Validators confirmed that the learning objectives, session sequencing, VSW design, and cross-platform connection tasks were aligned with the three levels of Johnstone's Chemistry Triplet and appropriate for the undergraduate level. Minor revisions to session instructions and VSW wording were incorporated before implementation. As reported in Table 3, all instrument items obtained Aiken's V coefficients ranging from 0.78 to 0.93, with instrument-level means between 0.84 and 0.87. These values indicate high inter-rater agreement across all validation aspects and confirm the content validity of all four instruments. Cronbach's alpha reliability coefficients were: scientific literacy test $\alpha = 0.81$, conceptual understanding test $\alpha = 0.78$, and student perception questionnaire $\alpha = 0.84$. For the observation sheet, inter-rater agreement between the two trained observers yielded a Cohen's kappa of $\kappa = 0.82$, indicating strong agreement (Landis & Koch, 1977). All values exceeded their respective minimum thresholds, confirming that all four instruments were psychometrically sound for use in data collection.

Learning Implementation

The eight-session scenario was implemented with an average fidelity rate of 94.6%, exceeding the pre-established minimum criterion of 85%, a threshold consistent with standards applied in prior virtual laboratory implementation research (Potkonjak et al., 2016; Lazonder & Harmsen, 2016). Meeting 1 (Crocodile Chemistry) achieved a fidelity rate of 93.0%, while Meeting 2 (PhET + Synthesis) reached 96.1%. The higher rate in Meeting 2 is consistent with the increased student engagement observed when students encountered PhET's electron excitation animations for the first time, suggesting that the novelty and visual accessibility of the submicroscopic visualization may have facilitated greater on-task involvement.

Two primary barriers were identified through structured observation notes. The first was an initial adaptation period to the PhET interface, requiring approximately 12 minutes at the start of Session 5, which was resolved through a brief instructor-led demonstration. The second barrier was the difficulty experienced by a subset of students in Session 7 when connecting flame color data from Crocodile Chemistry with the energy-level diagrams

in PhET — the most cognitively demanding transition in the scenario. This finding is consistent with Al-nakhle's (2022) conclusion that transitions across representational media require explicit scaffolding rather than self-directed discovery, and it underscores the importance of the VSW as a structured bridge between the two platforms.

The most theoretically significant qualitative finding was the emergence of spontaneous cross-level conceptual questions in Sessions 5 and 6 that had been entirely absent during the Crocodile Chemistry sessions. A representative example, raised independently by multiple students, was: "Why does sodium have a smaller electron energy jump than potassium, yet its flame is far brighter and more dominant?" This question demonstrates that students were actively integrating their submicroscopic observations from PhET with their macroscopic observations from Crocodile Chemistry — a form of spontaneous cross-representational reasoning consistent with what Wieman et al. (2008) and Bruce et al. (2022) associate with effective multimedia-mediated science learning.

Scientific Literacy and Conceptual Understanding

Table 5. Summary of Wilcoxon signed-rank test results for scientific literacy and conceptual understanding

Variable	Pre-test M ± SD	Post-test M ± SD	N-Gain	Category	Z	p	r	Interpretation
Scientific Literacy	48.6 ± 8.2	76.4 ± 6.9	0.54	Moderate	−4.96	< .001	0.88	Large effect
Explaining phenomena scientifically	46.2 ± 9.1	74.8 ± 7.3	0.53	Moderate	−4.74	< .001	0.84	Large effect
Evaluating and designing enquiry	50.4 ± 7.8	78.6 ± 6.2	0.57	Moderate	−4.82	< .001	0.85	Large effect
Interpreting data and evidence	49.2 ± 8.5	75.9 ± 7.1	0.53	Moderate	−4.60	< .001	0.81	Large effect
Conceptual Understanding	45.3 ± 9.4	80.2 ± 7.8	0.64	Moderate	−4.93	< .001	0.87	Large effect
Recall & Comprehension (C1–C2)	52.1 ± 8.7	83.4 ± 6.4	0.65	Moderate	−4.77	< .001	0.84	Large effect
Application & Analysis (C3–C4)	38.5 ± 10.2	77.0 ± 8.6	0.63	Moderate	−4.71	< .001	0.83	Large effect

Note: n = 32 paired observations. Effect size r was calculated as $|Z|/\sqrt{n}$ using the 32 paired cases (Field, 2009). M = Mean; SD = Standard Deviation. Sub-indicator labels correspond to the three PISA 2022 competencies as defined in Table 2. Effect size r > 0.50 = large effect (Cohen, 1988). N-Gain categories per Table 4. All p-values are two-tailed

Table 5 reveals consistent improvements across all measured sub-indicators. The Wilcoxon Signed-Rank Test results confirmed statistical significance for both outcome variables ($p < .001$), with large effect sizes after recalculation using n = 32 paired observations (r = 0.88 for scientific literacy; r = 0.87 for conceptual understanding), both exceeding Cohen's (1988) large-effect threshold (r > 0.50). It is important to note, however, that these results indicate improvements after the intervention in a single group and do not permit causal claims; the observed gains may reflect factors beyond the instructional design itself, including the passage of time, increased familiarity with the subject matter, or practice effects from the pre-test.

The pattern in which conceptual understanding (g = 0.64) improved more than scientific literacy (g = 0.54) is theoretically interpretable. Conceptual understanding in chemistry is directly contingent on representational completeness: when students were guided by the VSW in Session 7 to connect macroscopic flame colors from Crocodile Chemistry with submicroscopic energy diagrams from PhET, and then express that relationship symbolically in Session 8, an integrated conceptual schema appeared to form. This process corresponds to the three-

level integration that Johnstone's Chemistry Triplet identifies as fundamental to deep chemical understanding. The pattern is consistent with Widarti et al. (2022), who reported higher N-Gain values for conceptual understanding (g = 0.67) among students using multiple representation-based virtual laboratories compared with conventional instruction (g = 0.45), and with Raziana et al. (2025), who found that interactive molecular visualization platforms were associated with stronger conceptual knowledge gains than text-based simulation alternatives.

The relatively lower N-Gain for scientific literacy (g = 0.54) is explicable through sub-indicator analysis. The "interpreting data and evidence" sub-indicator, alongside "explaining phenomena scientifically," recorded the lowest gains among all sub-indicators (both g = 0.53), likely because authentic data interpretation competency requires sustained exposure to varied and open-ended data contexts that a two-session, structured scenario cannot fully provide. Bazie et al. (2024) reached a parallel conclusion: virtual laboratories tend to be more effective at building procedural and conceptual knowledge than at developing open-ended interpretive competencies. The moderate rather than high N-Gain categories most likely reflect the brevity of the intervention — two class meetings totaling six hours of instruction — rather than a mathematical ceiling effect. Theoretically, with pre-test means of 48.6 and 45.3, the normalized gain formula permits values up to g = 1.0, meaning high-category gains were mathematically attainable; the moderate outcomes instead suggest that more sustained instructional exposure would be needed to achieve representational integration at a deeper level.

Student Perception

Table 6. Student perceptions of the multimedia virtual laboratory package

Evaluation Aspect	Mean Score (1–5)	Category
Ease of use — Crocodile Chemistry interface	4.2	Good
Clarity of flame color visualization (Crocodile Chemistry)	4.5	Very Good
Clarity of electron excitation animation (PhET)	4.6	Very Good
Ease of use — PhET Discharge Lamps	4.3	Very Good
Macro–submicro–symbolic connection (integrated)	4.4	Very Good
Increase in learning motivation	4.1	Good
Capacity for independent learning	4.4	Very Good
Overall mean	**4.36**	**Very Good**

Student perceptions of the multimedia package were strongly positive (overall mean = 4.36/5). Notably, the highest-rated dimension was the clarity of PhET's electron excitation animation (4.6) — the component that directly addressed the submicroscopic representational gap absent from prior instructional media. The cross-level connection dimension (4.4) also received a high rating, suggesting that the VSW-guided integration between platforms was perceived as meaningful and effective by students. These findings are consistent with Tsirulnikov et al. (2023), who reported the highest student satisfaction on dimensions related to visualization of non-observable processes, and with Pane et al. (2024), who documented motivational gains through STEM-integrated interactive virtual laboratory media.

Inductive thematic analysis of 32 open-ended responses yielded three consistent themes. The first, irreplaceable visibility, captured students' recognition that PhET provided access to atomic-level processes that no prior instructional medium had made visible. The second, facilitated connection, reflected how the VSW in Session 7 enabled students to construct meaning across both platforms, transforming two separate learning experiences into a coherent understanding. The third, longing for sensory experience, indicated that some students wished to confirm virtual observations through hands-on physical experimentation — a finding that aligns with Wörner et al. (2022) and Serrano-Pérez et al. (2023) in suggesting that multimedia virtual laboratories function most powerfully as complements to, rather than substitutes for, conventional practicum.

Implications for Theory and Practice

Three interconnected theoretical arguments emerge from the pattern of findings. The first concerns representational completeness. The results are consistent with the central proposition of Johnstone's Chemistry Triplet: improvements in conceptual understanding appear to be associated with the completeness of representational coverage in the instructional design. Crocodile Chemistry alone addresses only the macroscopic and symbolic levels; PhET alone lacks the authentic procedural context of a flame test experiment. Their sequential integration, mediated by the VSW, appears to produce a more pedagogically complete instructional package, as suggested by the differential N-Gain pattern (conceptual understanding > scientific literacy). Permatasari et al. (2022), in a systematic literature review, similarly concluded that multi-representational chemistry instruction tends to produce superior conceptual outcomes compared with single-representation approaches, though they also cautioned that instructional design quality is a moderating factor.

The second argument concerns the role of scaffolded cross-representational connection. The MER framework (Treagust et al., 2017; Chang & Linn, 2013) predicts that the potential benefits of multiple representations are not automatic but depend on the quality of scaffolding connecting those representations. The qualitative observation that cross-level student questions emerged only after Session 7 — the explicit VSW cross-media connection activity — is consistent with this prediction and suggests that the VSW played a pivotal role in enabling representational integration. This is aligned with Bruce et al. (2022) and Permatasari et al. (2022), who argued that multi-level representations are associated with deeper understanding only when students are actively guided to construct connections across levels.

The third argument concerns evidence-based technology selection. The perception data suggest that PhET Discharge Lamps was valued by students primarily for its ability to make the submicroscopic level visible, not for its technological features per se. This is consistent with Wieman et al. (2008) and Finkelstein et al. (2005), who emphasized that the effectiveness of PhET derives from its research-based pedagogical design rather than its technical sophistication. The implication for instructional technology selection is that platform choices should be guided by analysis of which representational level requires technological mediation, rather than by technological novelty or institutional convenience.

CONCLUSION

This study provides preliminary evidence suggesting that the sequential integration of Crocodile Chemistry and PhET Interactive Simulations — Neon Lights & Other Discharge Lamps, mediated by a Virtual Student Worksheet that explicitly scaffolds cross-representational connections across the macroscopic, submicroscopic, and symbolic levels, may represent a potentially effective instructional approach for undergraduate Metal Chemistry Practicum. The implementation achieved a fidelity rate of 94.6%, and the results indicated statistically significant improvements with large effect sizes in scientific literacy and conceptual understanding following the intervention. Student perceptions were strongly positive. These findings are consistent with the proposition that representational completeness in multimedia virtual laboratory design — and the quality of scaffolding connecting the three levels of Johnstone's Chemistry Triplet — are important determinants of instructional quality. However, given the one-group design, causal claims cannot be made; the findings should be interpreted as indicating potential effectiveness and informing the design of future studies with stronger experimental controls.

REFERENCES

Afina, D. R., Hayati, M. N., & Fatkhurrohman, M. A. (2021). Profil capaian kompetensi literasi sains siswa SMP Negeri Kota Tegal menggunakan PISA. PSEJ (Pancasakti Science Education Journal), 6(1), 10–21. https://doi.org/10.24905/psej.v6i1.111

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Al-nakhle, H. (2022). The effectiveness of scenario-based virtual laboratory simulations to improve learning outcomes and scientific report writing skills. PLOS ONE, 17(11), e0277359. https://doi.org/10.1371/journal.pone.0277359

Ambarwati, A. H., Aini, H., Putri, N. S., & Fadillah, N. K. (2023). Analisis literasi kimia: Pentingnya pemahaman konsep kimia di sekolah menengah. Jurnal Arjuna: Publikasi Ilmu Pendidikan, Bahasa dan Matematika, 2(1), 165–174. https://doi.org/10.61132/arjuna.v2i1.474

Bazie, H., Lemma, B., Workneh, A., & Estifanos, A. (2024). The effect of virtual laboratories on the academic achievement of undergraduate chemistry students: Quasi-experimental study. JMIR Formative Research, 8, e64476. https://doi.org/10.2196/64476

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bruce, M. R. M., Bruce, A. E., & Walter, J. (2022). Creating representation in support of chemical reasoning to connect macroscopic and submicroscopic domains of knowledge. Journal of Chemical Education, 99(4), 1734–1746. https://doi.org/10.1021/acs.jchemed.1c00292

Bybee, R. W. (2011). Scientific and engineering practices in K-12 classrooms. The Science Teacher, 78(9), 34–40. https://doi.org/10.2505/4/tst11_078_09_34

Chang, H. Y., & Linn, M. C. (2013). Scaffolding learning from molecular visualizations. Journal of Research in Science Teaching, 50(7), 858–886. https://doi.org/10.1002/tea.21089

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. Science, 340(6130), 305–308. https://doi.org/10.1126/science.1230579

Fadilah, N., & Sukmawati, W. (2022). Efektivitas penggunaan laboratorium virtual terhadap keterampilan proses sains dan hasil belajar kognitif mahasiswa. Jurnal Inovasi Pendidikan Kimia, 16(1), 23–31. https://doi.org/10.15294/jipk.v16i1.30906

Felder, R. M., & Brent, R. (2016). Teaching and learning STEM: A practical guide. Jossey-Bass.

Field, A. P. (2009). Discovering statistics using SPSS (3rd ed.). SAGE Publications.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. Physical Review Special Topics — Physics Education Research, 1(1), 010103. https://doi.org/10.1103/PhysRevSTPER.1.010103

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64–74. https://doi.org/10.1119/1.18809

Haryanti, I., Kurniawati, Y., & Lubis, F. H. (2023). Penerapan virtual laboratory dengan menggunakan model problem based learning (PBL) untuk mendukung proses pembelajaran kimia. Journal of Chemistry Education and Integration, 2(1), 42–49. https://doi.org/10.24014/jcei.v2i1.21771

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. Journal of Computer Assisted Learning, 7(2), 75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x

Johnstone, A. H. (1993). The development of chemistry teaching: A changing response to a changing demand. Journal of Chemical Education, 70(9), 701–705. https://doi.org/10.1021/ed070p701

Kurniawati, Y., & Rahmawati, S. (2024). Identifikasi kesulitan siswa dalam pelaksanaan praktikum kimia madrasah aliyah swasta di kota Pekanbaru. UNESA Journal of Chemical Education, 13(3), 251–257. https://doi.org/10.26740/ujced.v13n3.p251-257

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. Review of Educational Research, 86(3), 681–718. https://doi.org/10.3102/0034654315627366

Lestari, L. (2023). Review: Laboratorium virtual untuk pembelajaran kimia di era digital. Jambura Journal of Educational Chemistry, 5(1), 1–10. https://doi.org/10.34312/jjec.v5i1.15008

Li, J., Liang, W., & Lavi, R. (2024). Effectiveness of virtual laboratory in engineering education: A meta-analysis. PLOS ONE, 19(12), e0316269. https://doi.org/10.1371/journal.pone.0316269

Manyilizu, M. C. (2023). Effectiveness of virtual laboratory vs. paper-based experiences to the hands-on chemistry practical in Tanzanian secondary schools. Education and Information Technologies, 28(5), 4831–4848. https://doi.org/10.1007/s10639-022-11327-7

OECD. (2023). PISA 2022 Assessment and Analytical Framework. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Pane, E. P., Lubis, N. F., & Simarmata, G. (2024). Application of interactive virtual lab media based on a STEM approach in improving students' scientific literacy and learning motivation. Jurnal Penelitian Pendidikan IPA, 10(12), 1–8. https://doi.org/10.29303/jppipa.v10i12.8589

Permatasari, M. B., Rahayu, S., & Dasna, I. W. (2022). Chemistry learning using multiple representations: A systematic literature review. Journal of Science Learning, 5(3), 430–440. https://doi.org/10.17509/jsl.v5i3.41474

Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., & Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. Computers & Education, 95, 309–327. https://doi.org/10.1016/j.compedu.2016.02.002

Raziana, E., Yamtinah, S., & Djono. (2025). Chemvilative: Enhancing conceptual understanding in chemistry through an android-based virtual laboratory. Jurnal Penelitian Pendidikan IPA, 11(3), 678–686. https://doi.org/10.29303/jppipa.v11i3.9885

Ristanti, S. D., & Sumarti, S. S. (2024). Analisis pemahaman konsep dan kesulitan siswa kelas XI pada materi hidrolisis garam menggunakan tes TTMC dan TwTMC dengan model problem-based learning. Jurnal Inovasi Pendidikan Kimia, 18(1), 23–31. https://doi.org/10.15294/jipk.v18i1.46418

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. Computers & Education, 58(1), 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Serrano-Pérez, J. J., González-García, L., Flacco, N., Taberner-Cortés, A., García-Arnandis, I., Pérez-López, G., Pellín-Carcelén, A., & Romá-Mateo, C. (2023). Traditional vs. virtual laboratories in health sciences education. Journal of Biological Education, 57(1), 36–50. https://doi.org/10.1080/00219266.2021.1877776

Sreekanth, N. S., Varghese, N., & Babu, N. S. C. (2022). Virtual chemistry lab to virtual reality chemistry lab. Resonance, 27(8), 1371–1385. https://doi.org/10.1007/s12045-022-1432-0

Stipek, J., Rusek, M., & Vojir, K. (2024). Between symbols and particles: Investigating the complexity of learning chemical equations. Education Sciences, 14(6), 570. https://doi.org/10.3390/educsci14060570

Treagust, D. F., Duit, R., & Fischer, H. E. (Eds.). (2017). Multiple representations in physics education. Springer. https://doi.org/10.1007/978-3-319-58914-5

Tsirulnikov, D., Suart, C., Abdullah, R., Vulcu, F., & Mullarkey, C. E. (2023). Game on: Immersive virtual laboratory simulation improves student learning outcomes and motivation. FEBS Open Bio, 13(3), 478–490. https://doi.org/10.1002/2211-5463.13567

Widarti, H. R., Anggraini, D., & Mulyani, S. (2022). Development of a multiple representation-based virtual laboratory to improve students' conceptual understanding and motivation in learning chemistry. Journal of Physics: Conference Series, 2215(1), 012117. https://doi.org/10.1088/1742-6596/2215/1/012117

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. Science, 322(5902), 682–683. https://doi.org/10.1126/science.1161948

Wörner, S., Kuhn, J., & Scheiter, K. (2022). The best of two worlds: A systematic review on combining real and virtual experiments in science education. Review of Educational Research, 92(6), 911–952. https://doi.org/10.3102/00346543221079417