International Multidisciplinary Research Journal
ISSN: 2424-7073
Available Online at https://imrjournal.info/
Volume V ; November 2026; Page No.13-26
DOI: https://doi.org/10.47722/imrj.2001.88

Research Article

EFFECTIVENESS OF AI-ASSISTED VISUALIZATION IN IMPROVING HIGH SCHOOL STUDENTS' UNDERSTANDING OF CHEMICAL BONDING AND MOLECULAR STRUCTURE: A SYSTEMATIC LITERATURE REVIEW

Aygerim Sadygulova[1&2]

[1]Nazarbayev Intellectual School of Natural Sciences and Mathematics, Medeu District, Almaty, Kazakhstan;
[2]Bolashak International Scholarship Program, George Washington University School of Business, Washington, D.C., United States

ARTICLE INFO

Article History:

Received 15.07.2026
Accepted 15.10.2026
Published 30.11.2026

Keywords:

artificial intelligence;
chemistry education;
molecular visualization;
chemical bonding;
submicroscopic representation;
Johnstone's triplet;
virtual reality;
intelligent tutoring systems;
systematic literature review;
high school chemistry

ABSTRACT

Chemical bonding and molecular structure are governed by submicroscopic phenomena that students cannot observe directly, making them a persistent source of misconception in high school chemistry. Artificial intelligence (AI) tools — including intelligent tutoring systems (ITS), generative AI and large language models (LLMs), and augmented and virtual reality (AR/VR) simulations — are increasingly proposed as scaffolds for this invisible content, yet evidence specific to secondary-level visualization remains scattered. This review synthesizes a PRISMA-guided search of five databases (Web of Science, Scopus, Google Scholar, ERIC, PubMed; January 2018–April 2025; 487 records screened, 30 studies retained), narrowed through a secondary relevance screen and supplementary 2024–2026 searching to 21 studies bearing directly on AI-assisted visualization of bonding and molecular structure. ITS produce short-term learning gains approaching human tutoring (effect sizes near 0.79 SD); AR/VR molecular visualization improves spatial and conceptual understanding of three-dimensional structure, including two recent high-school studies reporting large effects; and generative AI supports fluent, motivating explanations but remains unreliable for mechanistic reasoning (approximately 28% full accuracy). Because nearly all primary evidence relies on immediate post-testing, these gains are best read as short-term rather than confirmed durable conceptual change, mediated by instructional scaffolding rather than technological sophistication alone. Interpreted through Johnstone's chemistry triplet, the most effective tools explicitly bridge submicroscopic and symbolic representations rather than presenting visualization in isolation. The review identifies persistent underrepresentation of secondary-education contexts, limited concept-specific granularity, and weak integration of cognitive-load and mental-model theory into tool design, and proposes a research and practice agenda addressing these gaps

Copyright© The Author(s) 2025. Open Access under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits use, sharing, adaptation, distribution, and reproduction in any medium or format, giving proper credit to the original author, including a link to the license, and indicating if any changes were made. To view a copy of the licence, please visit https://creativecommons.org/licenses/by/4.0/

1. Introduction

Chemistry is widely regarded as one of the most cognitively demanding secondary-school subjects, owed less to content volume than to its distinctive representational structure. Johnstone (1982, 1991) formalized this as a "triplet": chemical knowledge must be constructed and translated across the macroscopic level of observable phenomena, the submicroscopic level of atoms, ions, and molecules, and the symbolic level of formulae and equations. The submicroscopic level is uniquely inaccessible to direct perception, approachable only through mental models and representational tools (Johnstone, 1991). Nakhleh (1992) showed that when this level is left implicit, students actively construct internally consistent, scientifically incorrect models — reasoning, for instance, that ionic bonds "break" on dissolution. Talanquer (2011) linked this to dual-process cognition: students default to fast, associative reasoning rather than the slower, model-based reasoning accurate submicroscopic thinking requires. Taber (2013) argued the triplet is a normative standard: genuine understanding of bonding requires an explanatory narrative connecting electron arrangement to both symbolic notation and macroscopic property — precisely the task for which interactive, explorable visualization, this review's focus, is well suited.

These demands are further illuminated by Sweller's (1988; Sweller et al., 1998) cognitive load theory, distinguishing intrinsic load (content complexity), extraneous load (poorly designed materials), and germane load (integrating new knowledge). Modeling a covalent molecule such as water requires holding electron configuration, bond energetics, three-dimensional geometry, polarity, and symbolic convention in mind simultaneously — a substantial, multiplicatively interacting intrinsic load (Sweller et al., 1998) that instruction must manage by minimizing extraneous load while preserving capacity for germane processing (Paas et al., 2003). Wu and Shah (2004) showed this balance is easily lost: a poorly designed static two-dimensional depiction of a three-dimensional structure can increase rather than reduce load, a risk acute for students with weaker spatial ability. Stieff (2011) showed spatial reasoning's contribution is task-specific, mattering most for genuine spatial-transformation tasks, implying visualization tools should yield their largest gains where bonding and structure demand the most spatial reasoning. Gilbert (2005) framed external visualizations as scaffolds for internal mental models, while Chittleborough and Treagust (2007) found most high-school students could not move fluently between macroscopic observation and submicroscopic explanation.

Experimental and correlational evidence converges on dynamic, interactive representations as superior to static ones for closing this gap. Williamson and Abraham (1995) found computer animation of particulate-level phenomena produced more accurate mental models than static instruction; Ardac and Akaygun (2004) replicated this with multimedia coordinating all three representational levels, consistent with Mayer's (2009) Cognitive Theory of Multimedia Learning; and Kozma (2003) argued the benefit derives from the cognitive work of mapping between representations, not any single representation's fidelity. Molecular-modeling software confirms this directly: Dori and Barak (2001) found high-school students using computerized modeling outperformed conventionally taught peers on structure–property reasoning, and Barnea and Dori (1999) found the benefit was largest for students with the

lowest initial spatial ability — a compensatory, equity-relevant pattern recurring throughout this literature. Meta-analyses situate these findings more broadly: Merchant et al. (2014) reported a positive VR effect across K–12 and higher education (d = 0.51), concentrated in high-interactivity conditions; Hu et al. (2025) found simulations produced larger gains for conceptual understanding than recall, moderated by instructional integration. Radianti et al. (2020) and Makransky and Lilleholt (2018) add a caveat: immersion does not automatically translate into learning, since the relationship is mediated by the cognitive processing scaffolding supports.

Artificial intelligence adds a qualitatively new capability: dynamic adaptation to the individual learner. Zawacki-Richter et al. (2019) identified four functional AI-in-education categories but found the literature dominated by computer-science rather than learning-science perspectives, a theory–practice gap echoed by Chen et al. (2020): AI built without reference to cognitive load or mental-model theory risks superficial score gains rather than genuine cognitive restructuring. ITS are the most extensively evidenced category: VanLehn's (2011) meta-analysis found well-designed ITS achieved effect sizes near 0.79 SD relative to classroom instruction, comparable to human tutoring, through step-level feedback and adaptive sequencing — mechanisms well matched to the analogy-driven misconceptions documented in bonding (Nakhleh, 1992; Talanquer, 2011). Generative AI and LLMs are newer and less settled: Kasneci et al. (2023) identified genuine strengths (personalized explanation, Socratic dialogue) alongside a structural limitation central here — LLMs are text-based and cannot natively generate the dynamic visualizations this literature identifies as central to submicroscopic understanding — suggesting they are best positioned as conversational scaffolds alongside, not substitutes for, dedicated visualization engines.

Two recent reviews approach but do not occupy this study's intersection. Erümit and Sarıalioğlu (2025) reviewed AI in science and chemistry education broadly, finding accelerating publication since 2021 and generally positive effects, but a pronounced skew toward higher education and away from secondary schooling. Abdulrahman et al. (2024) reviewed submicroscopic-level activities without restricting attention to AI, finding the most effective shared features — dynamic representation, cross-level integration, explicit scaffolding — that align with what well-designed AI environments can provide. Three gaps follow: the evidentiary base remains concentrated in higher education; the literature lacks concept-specific granularity, rarely specifying which AI tool best supports which of covalent bonding, ionic electron transfer, VSEPR geometry, or bond energetics; and cognitive-load/mental-model theory, well articulated for chemistry learning generally (Sweller et al., 1998; Mayer, 2009), remains weakly integrated into AI-assisted visualization design (Zawacki-Richter et al., 2019; Chen et al., 2020). Addressing these gaps, this review's contribution is stated explicitly here: unlike Erümit and Sarıalioğlu (2025), whose scope spans AI in science and chemistry education broadly, and unlike Abdulrahman et al. (2024), whose submicroscopic-level focus is not restricted to AI, this review narrows specifically to AI-assisted visualization of bonding and molecular structure, maps tool categories explicitly onto Johnstone's (1982) triplet and Sweller's (1988) cognitive load theory rather than post-test outcomes alone, and incorporates secondary-education evidence via the supplementary 2026 search (Section 2.4).

High-school students persistently fail to construct accurate mental representations of bond formation, three-dimensional molecular geometry, and the link between electron arrangement and macroscopic property — not a peripheral difficulty, since the submicroscopic level supplies the causal mechanism through which observation becomes explanation. Traditional approaches have not resolved this, and while AI-assisted visualization offers strong theoretical promise, evidence for its effectiveness in this specific context remains fragmented and loosely connected to the cognitive-science foundations that explain why the problem is hard. This review synthesizes the available evidence on AI-assisted visualization for high-school bonding and molecular-structure understanding, evaluates which tool categories are most effective for which conceptual challenge, and specifies the research gaps most in need of investigation, guided by four research questions framed using the Population–Intervention–Comparison–Outcome–Setting (PICOS) approach:

RQ1. To what extent does AI-assisted teaching improve high school students' conceptual understanding and visualization of submicroscopic chemical phenomena, compared with traditional or non-AI-based instruction?

RQ2. Which types of AI tools — intelligent tutoring systems, AI-enhanced simulations, AR/VR platforms, adaptive learning systems, and LLM assistants — are most effective in supporting visualization of specific submicroscopic concepts, including atomic structure, bonding, electron transfer, molecular geometry, and energy change?

RQ3. How do AI-assisted tools support students in bridging the submicroscopic and symbolic levels of chemical representation, as described by Johnstone's (1982) chemistry triplet?

RQ4. What advantages, limitations, and contextual factors are documented as influencing the effectiveness of AI-assisted teaching for submicroscopic visualization in high school chemistry?

2. Materials and Methods

2.1 Design and Reporting Standard

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). It synthesizes and re-scopes a systematic evidence corpus originally assembled for a broader question — AI applications in chemistry education generally — narrowing the analytic focus to studies bearing on AI-assisted visualization of submicroscopic bonding and molecular structure. This adaptation is made explicit below and revisited as a limitation in Section 4.3.

2.2 Search Strategy and Sources

Searches were conducted across five databases selected for interdisciplinary coverage of educational technology, chemistry, and computer science: Web of Science, Scopus, Google Scholar, ERIC (via EBSCOhost), and PubMed. Boolean strings combined two keyword clusters: ("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "generative AI" OR "ChatGPT" OR "large language model" OR "intelligent tutoring system" OR "virtual laboratory" OR "augmented reality" OR "learning analytics") AND ("chemistry education" OR "chemical education" OR "teaching chemistry" OR "learning chemistry" OR "STEM education"), limited to records published between January 2018 and April 2025.

2.3 Eligibility Criteria

Studies were included if they (a) addressed AI or machine-learning technologies in chemistry or closely related science education; (b) reported empirical findings, a theoretical framework, or a substantive analytical synthesis; (c) appeared in a peer-reviewed journal, conference proceeding, or scholarly book; and (d) were published in English within the 2018–2025 window. Studies were excluded if they addressed AI in chemical research without educational implications, were opinion pieces or short commentaries lacking empirical substance, were duplicates, were unavailable in full text, or were not published in English.

2.4 Screening, Selection, and Secondary Relevance Screen

The five-database search returned 487 records. After removal of 62 duplicates, 425 unique records underwent title/abstract screening, of which 298 were excluded as outside scope. The remaining 127 records underwent full-text assessment; applying the eligibility criteria above, together with supplementary hand-searching of reference lists, yielded 30 studies retained in the parent evidence corpus (Figure 1).

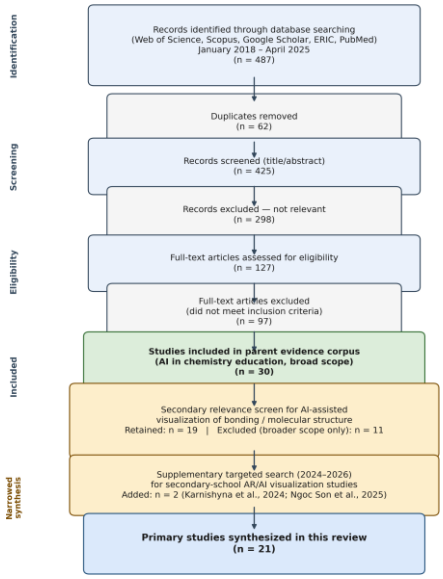


Figure 1. PRISMA-based study identification and screening flow, including the secondary relevance screen and supplementary search applied to narrow the parent evidence corpus to studies on AI-assisted visualization of chemical bonding and molecular structure.

Because the parent search targeted AI in chemistry education broadly rather than submicroscopic visualization specifically, a secondary relevance screen was applied to these 30 studies, assessing each against two focused criteria: (a) explicit engagement with molecular structure, chemical bonding, submicroscopic representation, or spatial/three-dimensional visualization, and (b) content bearing on RQ1–RQ4. Nineteen of the 30 studies met both criteria and constitute the primary evidence synthesized in Section 3; the remaining 11 — addressing general assessment accuracy, learning-analytics platforms, or teacher AI-readiness without a visualization or bonding focus — are drawn upon in Section 4 for context (accuracy limitations, ethics, implementation) but not treated as direct visualization evidence.

Because the parent search closed in April 2025, a supplementary targeted search was conducted in June–August 2026 for studies published after that date on AI-assisted or AR/VR-based visualization of molecular structure or bonding in secondary schools, using the terms ("augmented reality" OR "virtual reality" OR "generative AI") AND ("molecular structure" OR "chemical bonding" OR "molecular geometry") AND ("high school" OR "secondary school") across Google Scholar and publisher databases. This identified two studies meeting the same eligibility criteria (Karnishyna et al., 2024; Ngoc Son et al., 2025), bringing the primary evidence base to 21 studies. This supplementary step is a deviation from a single closed-date PRISMA search, reported transparently as such; it was undertaken specifically to address the secondary-education gap the parent review itself identified.

2.5 Data Extraction and Synthesis

For each of the 21 primary studies, the following were extracted: authorship and year, AI tool category, chemistry topic and representational level(s) addressed, educational level and geographic context, study design, and reported outcomes. Data were synthesized narratively by AI tool type (ITS; generative AI/LLMs; AR/VR immersive visualization; adaptive learning/ML-based assessment), consistent with the taxonomy in Section 1, and interpreted against the cognitive-load and chemistry-triplet frameworks introduced there. To differentiate stronger from weaker evidence, studies were tiered by design rigor and evidentiary directness (MMAT-informed; Pace et al., 2012): the quasi-experimental studies with explicit comparisons and the meta-analyses form the higher-certainty core of Section 3, while findings reaching this review only via secondary citation (†/‡, Table 1) are lower-certainty — short of a full MMAT/GRADE protocol (Section 4.3).

3. Results

3.1 Overview

Of the 21 primary studies, publication dates ranged from 1999 to 2025, with the two most recent (2024–2025) providing the reviews only dedicated high-school AR evidence. Consistent with Erümit and Sarıalioğlu (2025), secondary-education contexts remained a minority even after the supplementary search, underscoring RQ4's contextual-factors concern. Table 1 summarizes findings by AI tool category.

Table 1. Summary of findings by AI tool category (RQ1–RQ2)

AI Tool Category	Representative Studies	Concept(s) Targeted	Key Reported Effect
Intelligent tutoring systems (ITS)	VanLehn (2011); Walsh et al.‡	Equation balancing; stepwise bonding/stoichiometric reasoning	ITS ≈ 0.79 SD above classroom instruction; improved equation-balancing accuracy
AR/VR molecular visualization	Aw et al.†; Rodríguez et al.†; Karnishyna et al. (2024); Ngoc Son et al. (2025)	3D geometry; nucleophile/electrophile structure; acid–base/pH structure	45–95% reporting increased comprehension/confidence; large effects in high-school samples (partial η^2 = .231)

AI Tool Category	Representative Studies	Concept(s) Targeted	Key Reported Effect
VR immersive laboratories	Su & Cheng†; Dunnagan et al.†; Edwards et al.†; Ferrell et al.†	Lab procedure; molecular organic chemistry; spectroscopy	Motivational gains consistent (58–90% increases in interest/usefulness); performance gains inconsistent (no significant difference for IR spectroscopy)
Generative AI / LLMs	Yik & Dood (2024); Watts et al. (2023); Alasadi & Baiz (2024); Iyamuremye & Ndihokubwayo‡	Reaction mechanisms; electron movement; bonding/atomic-structure explanation; AI-generated imagery	High sophistication but limited accuracy (≈28% fully accurate mechanisms); +16.6% on bonding/atomic-structure assessment; emerging multimodal capability
Adaptive learning / ML assessment	Noyes et al. (2020); Connor et al.†	Intermolecular forces (London dispersion); spectral-interpretation reasoning	Good group-level agreement with human coders; eye-tracking reveals heuristic/misconception patterns

Note. † = as cited in Chiu (2021); ‡ = as cited in Iyamuremye et al. (2024). Both are secondary citations to studies retrieved through the parent review's search and reported there.

3.2 Intelligent Tutoring Systems

ITS are the most rigorously evidenced category: VanLehn's (2011) meta-analysis found effect sizes near 0.79 SD above classroom instruction, comparable to human tutoring, driven by step-level feedback and adaptive sequencing, while cognitive-modeling ITS improved equation-balancing performance specifically (Walsh et al., as cited in Iyamuremye et al., 2024). These mechanisms map onto the systematic, analogy-driven misconceptions documented in bonding instruction (Nakhleh, 1992; Talanquer, 2011): an ITS that diagnoses a student's model of ionic bonding as treating charge as an inherent atomic property, rather than a consequence of electron transfer, addresses the misconception at its cognitive source.

3.3 AR/VR Molecular Visualization

AR/VR provides the most direct evidence for three-dimensional visualization. In the parent corpus, NuPOV (an AR app for spatial molecular interaction) increased self-reported confidence for 45% of users (Aw et al.†), and MoleculARweb was rated supportive of understanding by 82% of teachers and 83% of students (Rodríguez et al.†; see Table 1). Two studies from the supplementary search substantially strengthen the high-school-specific evidence this literature has lacked: Karnishyna et al. (2024) found 49 secondary students using an AR molecular-geometry module significantly outperformed traditionally taught controls on content and spatial-ability measures, and Ngoc Son et al. (2025), in a quasi-experimental study of 191 Grade 8 students, found AR-integrated acid–base–pH–oxide–salt instruction produced significantly greater improvement (partial η^2 = .231), with three-dimensional visualization

identified as central to the gain alongside infrastructure barriers. Together these findings show AR/VR is not merely engaging but measurably effective at the secondary level specifically — directly answering RQ1–RQ2 for this category.

3.4 VR Immersive Laboratories and Adaptive/ML-Based Assessment

Immersive VR laboratories, distinct from AR structural tools, show a consistent pattern: positive on motivation, inconsistent on performance. Su and Cheng† reported higher post-test scores for a VR laboratory simulation; Edwards et al.† and Ferrell et al.† documented motivational gains (58.8% and 89.7% increases, respectively), while Dunnagan et al.† found no significant performance difference versus traditional infrared-spectroscopy labs — mirroring Makransky and Lilleholt's (2018) finding that immersion affects learning only through the cognitive processing it supports. Machine-learning tools extend this evidence base into assessment: Noyes et al. (2020) automated analysis of London dispersion force explanations with good group-level agreement with human coders, and eye-tracking (Connor et al.†) identified specific misconception heuristics during spectral interpretation — relevant to assessing genuine understanding rather than surface performance.

3.5 Generative AI and Large Language Models

Generative AI evidence is the least favorable category on accuracy, though promising on engagement and, increasingly, visual capability. Yik and Dood (2024) found ChatGPT-3.5 produced fully accurate explanations for only 28% of organic-reaction-mechanism prompts despite high explanation sophistication, and prompt cueing raised sophistication but not accuracy. Watts et al. (2023) found chatbot mechanistic responses seldom discussed electron movement, the submicroscopic feature bonding instruction most needs. That prompt cueing raised sophistication without raising accuracy (Yik & Dood, 2024) points to a genuine model limitation, not merely a prompt-design problem: as a text-based system lacking a grounded chemical representation (Kasneci et al., 2023), ChatGPT produces fluent language without the underlying electron-level model, and multi-step mechanism reasoning is itself high in intrinsic cognitive load (Sweller, 1988) — arguing for supervised classroom use in which students critique AI explanations against the electron-movement standard (Section 4.2). Against this, Iyamuremye and Ndihokubwayo‡ reported ChatGPT use increased performance by 16.6% on bonding/atomic-structure assessments, and Alasadi and Baiz (2024) demonstrated an emerging multimodal capability to generate and interpret chemical imagery — beginning to address the limitation Kasneci et al. (2023) identified, that text-only LLMs cannot natively produce visualizations. Tinto et al. (2026) similarly find generative AI maturing toward more critical, visualization-supportive use, while stressing that error management and AI literacy remain necessary.

4. Discussion

4.1 Synthesis Against the Research Questions

RQ1 (overall effectiveness). Because nearly all primary evidence relies on immediate post-testing (Section 4.3), "effectiveness" here denotes short-term learning gains, not confirmed durable conceptual change. On that basis, the evidence supports a qualified positive answer: AI-assisted tools outperform non-AI instruction most consistently where the tool

provides adaptive, diagnostic feedback (ITS: ≈0.79 SD; VanLehn, 2011) or dynamic three-dimensional visualization matched to spatial-reasoning demand (AR/VR: large effects in the two high-school studies; Karnishyna et al., 2024; Ngoc Son et al., 2025). Effectiveness is markedly less consistent where the technology is engaging but not diagnostically adaptive (VR laboratories without explicit misconception targeting) or where fluency is mistaken for accuracy (generative AI mechanistic explanation) — extending to the bonding/structure domain the general finding from Merchant et al. (2014) and Hu et al. (2025) that instructional integration, not visual sophistication alone, predicts learning gains.

RQ2 (which tools for which concepts). Table 1's four-category synthesis supplies the concept-specific granularity Erümit and Sarıalioğlu (2025) found the broader literature to lack. AR/VR shows its clearest advantage for spatial-transformation tasks — three-dimensional geometry, nucleophile/electrophile structure — consistent with Stieff's (2011) task-specific account. ITS shows its advantage for stepwise, rule-governed symbolic tasks such as equation balancing, where diagnostic feedback on procedural error is most actionable. Generative AI's comparative advantage lies in personalized explanation and, increasingly, multimodal image generation (Alasadi & Baiz, 2024), but not yet reliable mechanistic accuracy — favoring concept-specific tool selection over a single "best" category.

RQ3 (bridging submicroscopic and symbolic levels). Read through Johnstone's (1982) triplet, the strongest tools explicitly connect representational levels rather than presenting one in isolation. AR/VR platforms that let students manipulate a three-dimensional structure while viewing its symbolic formula (as in MoleculARweb; Rodríguez et al., as cited in Chiu, 2021) exemplify this bridging directly. The LLM evidence illustrates the opposite risk: Watts et al. (2023) found chatbot responses used surface mechanistic language while largely omitting electron movement — the submicroscopic mechanism that gives a symbolic arrow-pushing diagram its explanatory force. Taber's (2013) normative reading of the triplet is instructive: a fluent symbolic-level explanation without submicroscopic grounding does not thereby support genuine understanding, however sophisticated it appears. AI-assisted visualization tools should therefore be evaluated on whether they prompt students to map between levels, not on structural accuracy or fluency alone.

RQ4 (advantages, limitations, and contextual factors). Advantages include personalized pacing and feedback (Iyamuremye et al., 2024), a compensatory effect for students with lower baseline spatial ability (Barnea & Dori, 1999), and consistent motivational gains (Ferrell et al.; Edwards et al., as cited in Chiu, 2021). Limitations include generative AI's accuracy ceiling (Yik & Dood, 2024), immersion's dependence on scaffolding (Makransky & Lilleholt, 2018; Radianti et al., 2020), and — from the parent corpus's broader-scope studies — concerns about data privacy, algorithmic bias, the digital divide, and low teacher preparedness. Contextual moderators include instructional integration (Hu et al., 2025), infrastructure availability (Ngoc Son et al., 2025), and explicit level-bridging design (RQ3).

4.2 Theoretical and Practical Implications, and Originality

Theoretically, this review's central contribution is showing that cognitive-load and chemistry-triplet frameworks — well established for chemistry learning generally but, as Zawacki-Richter et al. (2019) and Chen et al. (2020) documented, only weakly integrated into AI-in-education research — provide genuine explanatory purchase on why some AI-assisted

visualization tools succeed and others do not. The pattern is not "more adaptive AI is better" but "AI that reduces extraneous load, supports germane processing, and scaffolds explicit cross-level mapping is better." Practically, educators addressing three-dimensional structural concepts are best served by AR/VR platforms with demonstrated secondary-level evidence (Karnishyna et al., 2024; Ngoc Son et al., 2025); those addressing stepwise symbolic tasks by ITS; and generative AI is best used as a supplementary explanatory aid under critical supervision, given its accuracy limitations (Yik & Dood, 2024) — for instance, having students evaluate and critique AI-generated explanations rather than accept them uncritically (cf. Guo & Lee, 2023, as cited in the parent corpus). The infrastructure/training barriers Ngoc Son et al. (2025) identified, and low baseline AI familiarity among chemistry teachers (Berber et al., 2025, as cited in the parent corpus), indicate technology provision alone is insufficient without professional development; implementation cost also differs by category, from low-hardware smartphone-based AR to costlier VR laboratories, with generative AI requiring the greatest ongoing AI-literacy training investment (Section 3.5).

As stated in the Introduction, this review's originality rests on three points: the concept-specific narrowing of the AI-in-chemistry-education literature to submicroscopic visualization, absent from Erümit and Sarıalioğlu (2025); the explicit mapping of AI tool categories onto Johnstone's (1982) triplet and Sweller's (1988) cognitive load theory, rather than post-test score alone; and the incorporation of two 2024–2025 high-school AR studies that supply the field's most current secondary-education evidence, with the scope-narrowing methodology transparently flagged (Section 2.4).

4.3 Research Gaps and Limitations

Consistent with the broader literature (Erümit & Sarıalioğlu, 2025; Iyamuremye et al., 2024), secondary-education evidence remains a minority even after the supplementary search, and nearly all identified high-school AR/VR studies are quasi-experimental with immediate post-tests and no delayed follow-up, leaving open whether gains reflect durable restructuring or short-term improvement. No study directly compared two or more AI tool categories on the same concept, and studies measuring cognitive load or mental-model quality, rather than post-test accuracy alone, remain rare.

These gaps sit alongside limitations of the review itself. It re-scopes an existing 30-study corpus assembled for a broader parent question rather than an independent, prospectively registered search using bonding/visualization-specific terms from the outset; studies using alternative terminology (e.g., "particulate model," "3D structure") may not have been captured, and the review was not registered (e.g., with PROSPERO); future work on this question should register a protocol prospectively and use bonding-specific search terms from the outset. The supplementary 2026 search was targeted rather than systematic across all five databases. No formal risk-of-bias or GRADE assessment was conducted, given design heterogeneity, which also precluded meta-analytic pooling; effect sizes in Table 1 are drawn directly from the original studies and are not directly comparable across them. Finally, most primary evidence is quasi-experimental with immediate post-testing, limiting causal and longitudinal inference.

5. Conclusion

This systematic review synthesized evidence on AI-assisted visualization for high school students' understanding of chemical bonding and molecular structure, drawing on 21 studies from a PRISMA-guided search and secondary relevance screen, including two 2024–2025 studies providing new high-school-specific AR evidence. Intelligent tutoring systems and AR/VR molecular visualization produce the most consistent and best-theorized short-term learning gains; generative AI offers valuable but currently unreliable explanatory support best used under critical supervision; and engagement/motivational benefits are considerably more consistent across tool categories than deep, durable conceptual gains, which the largely immediate-post-test evidence base cannot yet confirm. Interpreted through Johnstone's chemistry triplet and Sweller's cognitive load theory, the most effective tools explicitly scaffold movement between submicroscopic and symbolic representations while managing, rather than adding to, the substantial intrinsic load bonding and molecular structure impose. The central practical implication is that concept-specific tool selection — AR/VR for spatial tasks, ITS for stepwise symbolic tasks, generative AI for supervised explanatory support — is likely more productive than treating "AI-assisted teaching" as one undifferentiated intervention. The central research implication is that future primary studies should be explicitly designed around cognitive-load and mental-model measures, delayed retention testing, and direct concept-specific comparisons between tool categories, building on the secondary-education evidence base that recent studies (Karnishyna et al., 2024; Ngoc Son et al., 2025) have only just begun to establish.

Acknowledgements

This work was completed during the author's academic internship under the Bolashak International Scholarship Program at the George Washington University School of Business, Washington, D.C., United States. The author received no specific grant funding for this research from funding agencies in the public, commercial, or not-for-profit sectors.

Artificial Intelligence Tools Disclosure

An AI-based large language model assistant (Claude, Anthropic) was used during preparation of this review to support literature search and organization, synthesis of extracted study data, and text editing for clarity and formatting. All AI-assisted output was critically reviewed and verified by the author, who takes full responsibility for the manuscript's accuracy and conclusions.

References

Abdulrahman, M., Bhutta, S. M., Alrammal, M., & Poduska, K. (2024). A review of research on learning activities addressing the submicroscopic level in chemistry. *Journal of Chemical Education, 101*(4), 1275–1290. https://doi.org/10.1021/acs.jchemed.4c00156

Alasadi, E. A., & Baiz, C. R. (2024). Multimodal generative artificial intelligence tackles visual problems in chemistry. *Journal of Chemical Education, 101*(7), 2716–2729. https://doi.org/10.1021/acs.jchemed.4c00138

Ardac, D., & Akaygun, S. (2004). Effectiveness of multimedia-based instruction that emphasizes molecular representations on students' understanding of chemical change. *Journal of Research in Science Teaching, 41*(4), 317–337. https://doi.org/10.1002/tea.20005

Barnea, N., & Dori, Y. J. (1999). High-school and undergraduate chemistry students' performance and gender differences in a computerized molecular modeling learning environment. *Journal of Science Education and Technology, 8*(4), 257–271. https://doi.org/10.1023/A:1009436509753

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chittleborough, G. D., & Treagust, D. F. (2007). The modelling ability of non-major chemistry students and their understanding of the sub-microscopic level. *Chemistry Education Research and Practice, 8*(3), 274–292. https://doi.org/10.1039/B6RP90035F

Chiu, W.-K. (2021). Pedagogy of emerging technologies in chemical education during the era of digitalization and artificial intelligence: A systematic review. *Education Sciences, 11*(11), 709. https://doi.org/10.3390/educsci11110709

Dori, Y. J., & Barak, M. (2001). Virtual and physical molecular modeling: Fostering model perception and spatial understanding. *Educational Technology & Society, 4*(1), 61–74.

Erümit, A. K., & Sarıalioğlu, R. Ö. (2025). Artificial intelligence in science and chemistry education: A systematic review. *Discover Education, 4*, Article 178. https://doi.org/10.1007/s44217-025-00622-3

Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 9–27). Springer. https://doi.org/10.1007/1-4020-3613-2_2

Hu, B., Jia, W., Mi, S., & Bi, H. (2025). The impact of computer simulation on students' conceptual understanding in K-12 science education: A meta-analysis. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-025-10189-3

Iyamuremye, A., Niyonzima, F. N., Mukiza, J., Twagilimana, I., Nyirahabimana, P., Nsengimana, T., Habiyaremye, J. D., Habimana, O., & Nsabayezu, E. (2024). Utilization of artificial intelligence and machine learning in chemistry education: A critical review. *Discover Education, 3*, 95. https://doi.org/10.1007/s44217-024-00171-1

Johnstone, A. H. (1982). Macro- and micro-chemistry. *School Science Review, 64*(227), 377–379.

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

Karnishyna, D. A., Selivanova, T. V., Nechypurenko, P. P., Starova, T. V., & Semerikov, S. O. (2024). Enhancing high school students' understanding of molecular geometry with augmented reality. *Science Education Quarterly, 1*(2), 25–40. https://doi.org/10.55056/seq.818

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kozma, R. B. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205–226. https://doi.org/10.1016/S0959-4752(02)00021-X

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development, 66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education, 69*(3), 191–196. https://doi.org/10.1021/ed069p191

Ngoc Son, P., Dang, T. T. A., Hoai, V. T. T., Thai, H. M., Chu, T., & Nguyen, M. H. (2025). Augmented reality to enhance chemistry learning outcomes in Vietnamese lower secondary schools: A quasi-experimental study on acid–base–pH–oxide–salt topics. *European Journal of Educational Research, 14*(4), 1259–1275. https://doi.org/10.12973/eu-jer.14.4.1259

Noyes, K., McKay, R. L., Neumann, M., Haudek, K. C., & Cooper, M. M. (2020). Developing computer resources to automate analysis of students' explanations of London dispersion forces. *Journal of Chemical Education, 97*(11), 3923–3936. https://doi.org/10.1021/acs.jchemed.0c00445

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies, 49*(1), 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Stieff, M. (2011). When is a molecule three dimensional? A task-specific role for spatial reasoning in applying theoretical chemistry to stereochemistry. *Science Education, 95*(2), 310–336. https://doi.org/10.1002/sce.20422

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice, 14*(2), 156–168. https://doi.org/10.1039/C3RP00012E

Talanquer, V. (2011). Macro, submicro, and symbolic: The many faces of the chemistry 'triplet'. *International Journal of Science Education, 33*(2), 179–195. https://doi.org/10.1080/09500690903386435

Tinto, K., Gibson van Mil, L., & Scott, F. J. (2026). Towards effective and ethical GenAI in chemistry education. *Nature Reviews Chemistry, 10*(3), 156–157. https://doi.org/10.1038/s41570-026-00799-7

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Watts, F. M., Dood, A. J., Shultz, G. V., & Rodriguez, J.-M. G. (2023). Comparing student and generative artificial intelligence chatbot responses to organic chemistry writing-to-learn assignments. *Journal of Chemical Education, 100*(10), 3806–3817. https://doi.org/10.1021/acs.jchemed.3c00664

Williamson, V. M., & Abraham, M. R. (1995). The effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching, 32*(5), 521–534. https://doi.org/10.1002/tea.3660320508

Wu, H. K., & Shah, P. (2004). Leveraging visiospatial thinking in science education. *Science Education, 88*(3), 465–492. https://doi.org/10.1002/sce.10126

Yik, B. J., & Dood, A. J. (2024). ChatGPT convincingly explains organic chemistry reaction mechanisms slightly inaccurately with high levels of explanation sophistication. *Journal of Chemical Education, 101*, 1836–1846. https://doi.org/10.1021/acs.jchemed.4c00235

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

How to cite this article:
Sadygulova A. (2026) Effectiveness of Ai-Assisted Visualization in Improving High School Students' Understanding of Chemical Bonding and Molecular Structure: A Systematic Literature Review, *International Multidisciplinary Research Journal*, Volume V; November 2026; Page 13-26. DOI: https://doi.org/10.47722/imrj.2001.88