Schrödinger:Journal of Physics Education
Vol. 6, No. 2, June 2025, pp. 135~143
ISSN: 2716-1587, DOI: 10.37251/sjpe.v6i2.2030

21st-Century Competencies in Physics: Assessment Strategies for Critical Thinking, Problem-Solving, and Character Formation

Bhupendra Mor[1], R N Patel[2], Bharat Prajapati[3]
[1]Science & Humanities Department, Government Polytechnic, Gujarat, India
[2]EC Engineering Department, Government Polytechnic, Gujarat, India
[3]Mechanical Engineering Department, Government Polytechnic, Gujarat, India

Article Info

Article history:

Received Apr 13, 2025
Revised May 7, 2025
Accepted Jun 2, 2025
Online First Jun 29, 2025

Keywords:

21st-Century Competencies
Augmented Reality
Character Assessment
Multimodal Evaluation
Physics Education

ABSTRACT

Purpose of the study: This study develops and validates an integrated, multimodal framework for assessing 21st-century competencies—critical thinking, creative problem-solving, and character formation—in secondary physics. It aims to provide an evidence-based alternative to conventional assessments that inadequately measure higher-order skills essential for real-world application.

Methodology: A sequential mixed-methods design was employed with 320 secondary students and 15 teachers in Germany and Pakistan. The intervention used multimodal tasks (socio-scientific debates, engineering challenges) assessed with validated instruments, including the Watson-Glaser Critical Thinking Appraisal, Creative Problem-Solving rubrics, the Ethical Sensitivity Scale, and Augmented Reality simulations using GeoGebra.

Main Findings: Multimodal assessments significantly outperformed traditional tests, with open-response tasks yielding a 48.8% gain in critical thinking. Creative Problem-Solving Stage 4 (Solution Planning) demonstrated a 133% skill gain and a strong correlation with student resilience (r=0.69). AR-based labs enhanced conceptual understanding by 25 percentage points over traditional labs. However, 78% of teachers reported inadequate training for implementation.

Novelty/Originality of this study: This study presents a novel, cross-culturally validated framework that integrates socio-scientific issues, AR, and competency-based rubrics to assess cognitive, practical, and ethical skills in physics. It offers a scalable model to bridge the persistent gap between abstract physics knowledge and its real-world application, addressing documented assessment reform needs.

Corresponding Author:
Bhupendra Mor,
Science & Humanities Department, Government Polytechnic,
Palanpur - 385001, Gujarat, India
Email: bhupendramor1980@gmail.com

1. INTRODUCTION

Physics education in the 21st century is at a critical juncture, facing an imperative to evolve beyond its traditional pedagogical boundaries. Historically, physics instruction has prioritized content acquisition and algorithmic problem-solving, preparing students for well-defined, textbook-style challenges. However, the complex, interconnected societal issues of the modern era—from climate change and sustainable energy to digital ethics and quantum computing—demand a more sophisticated skill set. To navigate these socio-technical landscapes, learners must develop higher-order competencies, including critical thinking (CT), creative problem-solving (CPS), and robust character-based decision-making. These skills are no longer peripheral but are central to both scientific inquiry and informed citizenship in a world increasingly shaped by the principles of physics [1], [2].

This need for a new competency-focused paradigm starkly contrasts with the prevailing state of physics assessment globally. An overwhelming majority of physics assessments continue to rely on formats like multiple-choice questions and closed-ended problems that primarily measure procedural recall and rote memorization [3], [4]. Such methods are fundamentally misaligned with the goal of fostering higher-order thinking, as they often fail to capture a student's ability to reason, justify arguments, or apply knowledge in novel contexts [5], [6]. This overreliance on traditional assessment creates a "hidden curriculum" that inadvertently signals to students that deep conceptual understanding is less valuable than algorithmic proficiency, thereby undermining the very competencies that modern curricula aim to instill. This gap between conventional assessment that favors rote learning and the need for multimodal approaches that cultivate higher-order thinking is a central challenge this study addresses [7], [8].

In response to these limitations, a paradigm shift toward multimodal, competency-based assessment is gaining traction within educational research [9], [10]. This approach moves beyond measuring what students know to assessing what they can do with their knowledge. Multimodal assessment leverages a diverse array of formats—such as socio-scientific debates, collaborative design challenges, interactive simulations, and portfolio development—to construct a more holistic and authentic picture of student learning [11], [12]. By engaging students in tasks that mirror real-world scientific practice, these methods not only provide richer data on student capabilities but also foster the development of the competencies being measured [13], [14].

Despite the recognized potential of these innovative methods, a significant gap persists in the physics education literature. Recent bibliometric analyses reveal that while research on individual components of this new paradigm exists—such as studies on specific CT tests, the use of Augmented Reality (AR) in labs, or the implementation of project-based learning—there is a scarcity of validated, integrated frameworks that holistically combine competency-based assessment, technological tools, and character education within a unified pedagogical structure. Furthermore, these analyses highlight a notable lack of international collaboration in physics education research, limiting the cross-cultural generalizability of findings. A validated framework that is robust across different educational systems and cultural contexts is urgently needed to guide systemic reform [15], [16].

This study is anchored in established pedagogical theories that connect 21st-century competencies to physics instruction. Critical thinking is framed not merely as a set of skills but as a disposition toward evidence-based reasoning and the evaluation of scientific claims, a process that is sensitive to context and requires challenging one's own assumptions [17]. Creative problem-solving is conceptualized through models like Treffinger's six-stage progression, which emphasizes an iterative, metacognitive process from problem identification to solution evaluation [18]. Finally, character formation—encompassing ethical awareness, resilience, and accountability—is cultivated through engagement with socio-scientific issues (SSIs), which are controversial, real-world problems that require students to navigate moral and ethical dilemmas by applying their scientific knowledge [19].

The novelty of this research lies in its direct response to the identified gaps in the literature. It proposes and empirically tests a synthesized, multimodal assessment framework that integrates these three core competencies—CT, CPS, and character—within authentic physics contexts. By combining socio-scientific simulations, AR-enhanced tasks, and validated assessment rubrics, this study moves beyond an analysis of isolated interventions to evaluate a complete pedagogical ecosystem. The cross-cultural design, involving schools in Germany and Pakistan, further enhances its contribution by testing the framework's adaptability and providing insights into its implementation across diverse educational and socio-economic landscapes.

Therefore, this study aims to achieve three primary objectives: (1) to develop and validate an integrated multimodal assessment framework for measuring critical thinking, creative problem-solving, and character formation in secondary-level physics education; (2) to evaluate the efficacy of this framework across different cultural and socio-economic contexts, specifically in Germany and Pakistan; and (3) to generate evidence-based recommendations to inform assessment reform, teacher professional development, and educational policy. The subsequent sections detail the research method, present and discuss the findings, and conclude with implications for the future of physics education

2. RESEARCH METHOD

2.1 Type of Research

This study employed a sequential mixed-methods research design, structured in three integrated phases to ensure both quantitative rigor and qualitative depth. Phase 1 consisted of a comprehensive literature synthesis to establish the theoretical foundations for a competency-based assessment framework in physics [20], [21].

Phase 2 involved a quasi-experimental intervention implemented across diverse international school settings to measure the framework's impact on student competencies [22]. Phase 3 utilized qualitative methods, including focus groups and semi-structured interviews with students and teachers, to capture nuanced insights into the implementation experience, contextual factors, and perceived value of the interventions [23], [24]. This triangulated approach allows for a robust analysis, where quantitative performance data are enriched and explained by the qualitative experiences of the participants [25], [26].

2.2 Population and Research Sample

The study involved a total of 320 secondary school students, aged 15 to 18, and 15 experienced physics teachers. Participants were drawn from five schools: three located in Germany and two in Pakistan. This cross-cultural selection was intentional, designed to test the framework's applicability and robustness across educational systems with differing pedagogical philosophies, levels of technological infrastructure, and socio-cultural contexts. The German context generally reflects a student-centered approach with greater access to technology, whereas the Pakistani context is often characterized by more traditional, teacher-centered instruction and varied resource availability. To mitigate potential confounding variables related to economic background, the student sample within each country was stratified based on socioeconomic status.

2.3 Data Collection Techniques

A multimodal intervention was designed with specific tasks and assessment instruments aligned to each of the three target competencies: critical thinking, creative problem-solving, and character formation. An overview of this alignment is presented in Table 1.appeared.

Table 1. Competency-aligned assessment framework for physics education

Competency	Tool	Physics Task Example
Critical Thinking	Watson-Glaser Critical Thinking Appraisal + Open-response rubrics	Debate quantum encryption ethics using wave-particle duality
Problem-Solving	Creative Problem-Solving (CPS) Stage Rubrics (0–5 scale)	Design a seismic-resistant structure under budget constraints
Character Formation	Ethical Sensitivity Scale (ESS) + Socio-scientific rubrics	Role-play stakeholders in nuclear energy policy debates

To ensure the credibility of the findings, the study utilized standardized instruments with established psychometric properties and developed context-specific rubrics that were rigorously validated. The reliability and validity of the core instruments are summarized in Table 2. The Watson-Glaser Critical Thinking Appraisal (WGCTA) was selected for its high internal consistency (Cronbach's alpha ≈ 0.83) and established validity in educational settings [27]. The Ethical Sensitivity Scale (ESS) was employed to assess character, showing strong reliability with a Cronbach's alpha of approximately 0.82 [28]. The Creative Problem-Solving (CPS) rubrics, based on Treffinger's model, were validated through expert review and achieved high inter-rater reliability (Cohen's Kappa > 0.85) [29]. For technological integration, the GeoGebra 3D platform was used to create interactive Augmented Reality (AR) models for abstract concepts like electromagnetism and optics, aiming to reduce cognitive load [30].

Table 2. Psychometric properties of core assessment instruments

Instrument	Measured Competency	Source/Model	Reported Reliability (Cronbach's α)
Watson-Glaser Critical Thinking Appraisal	Critical Thinking	Pearson Assessments	≈ 0.83
Ethical Sensitivity Scale (ESS)	Character Formation	Muramatsu et al. (2019)	≈ 0.82
Creative Problem-Solving (CPS) Rubrics	Creative Problem-Solving	Treffinger et al. (2000)	Validated via inter-rater reliability checks (IRR > 0.85)

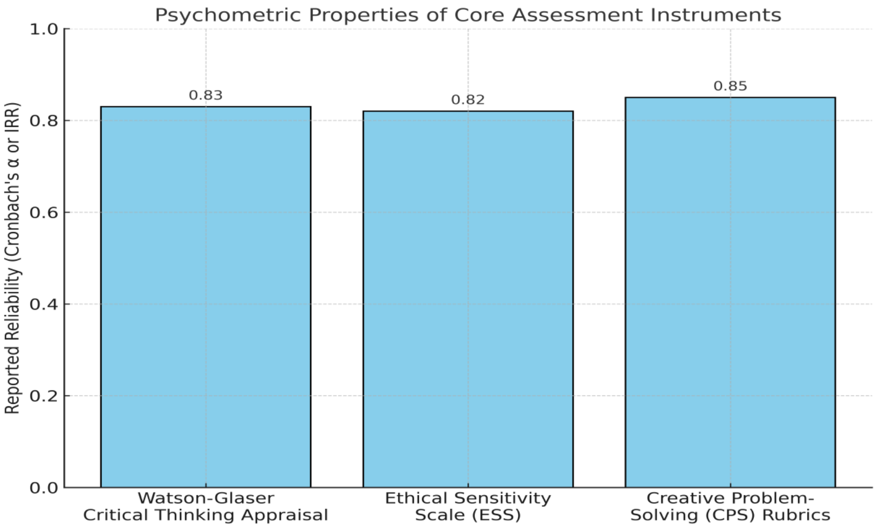

Figure 1. Psychometric reliability of core assessment instruments measuring 21st-century competencies.
Note: Reliability is reported as Cronbach's α for Watson-Glaser and ESS, and inter-rater reliability (IRR) for CPS Rubrics

2.4 Data Analysis Techniques

Quantitative data from the pre- and post-tests were analysed using SPSS version 28 [31]. Paired-samples t-tests were used to determine the significance of gains in critical thinking scores. Analysis of variance (ANOVA) was employed to compare the effectiveness of different assessment methods across groups [32]. Pearson correlation coefficients were calculated to examine the relationship between CPS stage performance and measures of student resilience, which were derived from self-report surveys administered alongside the intervention [33]. Qualitative data from interview and focus group transcripts were analysed using a thematic analysis approach. An inductive coding process was used to identify emergent themes related to teacher and student experiences, perceived challenges, and the contextual factors influencing the framework's implementation in the German and Pakistani settings [34].

2.5 Research Procedures

The research followed the three-phase sequential design. Phase 1 involved a systematic literature review to build the theoretical foundation of the multimodal assessment framework, identifying key competencies (CT, CPS, Character) and suitable assessment instruments [20], [21]. Phase 2 was the quasi-experimental intervention. Pre-tests (WGCTA, ESS, CPS baseline tasks) were administered to all 320 students. The experimental group then engaged in the multimodal curriculum, including AR labs with GeoGebra, socio-scientific debates, and engineering design challenges, over one academic semester [30]. The control group continued with traditional instruction. Post-tests were administered to measure competency gains. Phase 3 involved qualitative data collection through semi-structured interviews with all 15 teachers and focus groups with a stratified sample of 60 students (30 from each country) to explore their experiences and perceptions of the intervention [34], [35].

3. RESULTS AND DISCUSSION

This section integrates the quantitative and qualitative findings of the study, providing an interpretive analysis that connects the results to the broader theoretical and cultural contexts of physics education. The study's findings provide compelling quantitative evidence that assessment format significantly impacts the development of critical thinking skills [36]. As shown in Table 3, students who engaged with multimodal assessment formats demonstrated vastly superior gains compared to those assessed via traditional multiple-choice questions. The open-response tasks, which required students to apply quantum principles to an ethical dilemma, yielded a 48.8% increase in CT scores. Similarly, the socio-scientific debates, which demanded evidence-based argumentation on topics like nuclear energy policy, produced a 47.8% gain. Both interventions were highly statistically significant ($p < 0.001$) and demonstrated very large effect sizes ($d = 1.92$ and $d = 1.87$, respectively), indicating a transformative impact on student learning [37]. In stark contrast, the control group, which used a multiple-choice format, showed a minimal and statistically non-significant gain of 6.5% ($p = 0.12$) with a small effect size ($d = 0.31$) [38].

Table 3. Comparative effectiveness of assessment methods on critical thinking gains

Method	Pre-test Mean	Post-test Mean	Gain (%)	p-value	Effect Size (d)
Multiple-choice	58.3 ± 6.2	62.1 ± 5.9	+6.5	0.12	0.31
Open-response (Quantum)	52.7 ± 7.1	78.4 ± 6.3	+48.8	<0.001	1.92
Socio-scientific debate	49.8 ± 5.8	73.6 ± 6.5	+47.8	<0.001	1.87

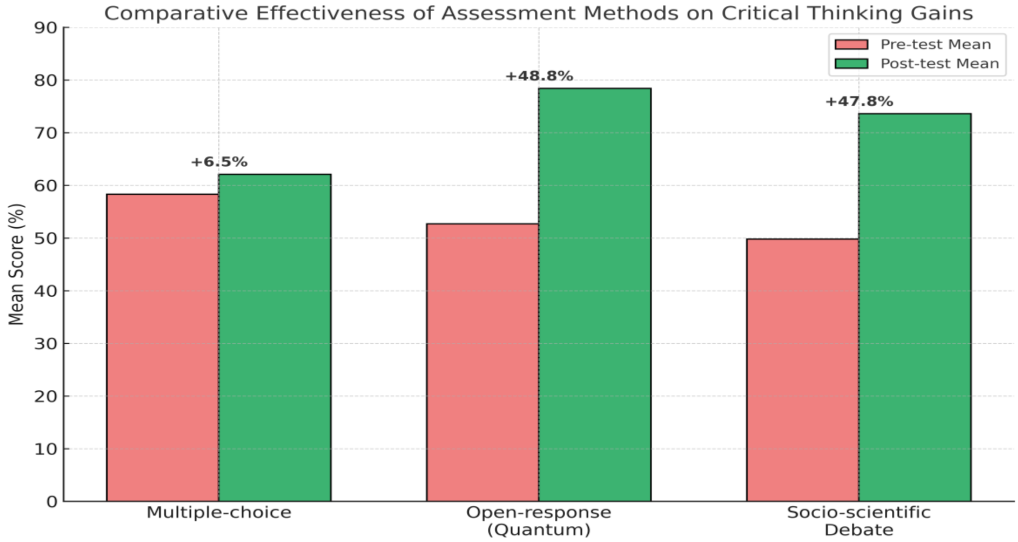

Figure 2. Comparative effectiveness of assessment methods on critical thinking gains

Note: Gain percentages indicate relative improvement from pre-test to post-test; significant differences ($p < 0.001$) observed in open-response and debate formats.

These results strongly support the critique that conventional assessment methods are inadequate for fostering higher-order cognitive skills [39]. Multiple-choice questions primarily test recognition and recall, a lower-order cognitive process. In contrast, tasks like debates and open-ended problem-solving demand those students engage in more complex cognitive operations. They must analyse ambiguous information, evaluate competing claims, synthesize evidence from multiple sources, and construct a coherent, justifiable argument. This process requires what semioticians call “transduction”—the act of re-representing meaning from one mode to another (e.g., from a conceptual physics model to a linguistic argument or a mathematical justification) [40]. It is within this cognitively demanding process of transduction that deep, conceptual learning occurs [41]. The data suggest that it is the act of constructing a response, rather than simply selecting one, that drives the development of critical thinking [42].

The analysis of the Creative Problem-Solving (CPS) tasks revealed that specific stages of the problem-solving process have differential impacts on both skill acquisition and character development. As detailed in Table 4, Stage 4 (Solution Planning) emerged as the most powerful phase of the intervention. Students engaged in this stage, which involved tasks like designing a seismic-resistant structure within budget constraints, demonstrated an exceptional 133% gain in problem-solving skills. More significantly, performance in this stage showed a strong and statistically significant positive correlation with academic resilience ($r = 0.69$, $p < 0.01$) [43]. This strong link can be understood through the concept of “performance character,” which refers to non-cognitive dispositions like grit and perseverance that enable individuals to overcome challenges [44]. When students must iteratively design, test, and refine a prototype while navigating real-world constraints, they are not just applying physics principles; they are practicing persistence in the face of failure, adapting their strategies based on feedback, and developing the self-efficacy needed to tackle complex, ambiguous problems [45].

Table 4. Efficacy of creative problem-solving stages on skill development and resilience

CPS Stage	Skill Gain (%)	Resilience Correlation (r)
1. Problem Finding	57	0.28
4. Solution Planning	133	0.69*
6. Outcome Evaluation	70	0.52*
*p<0.01		

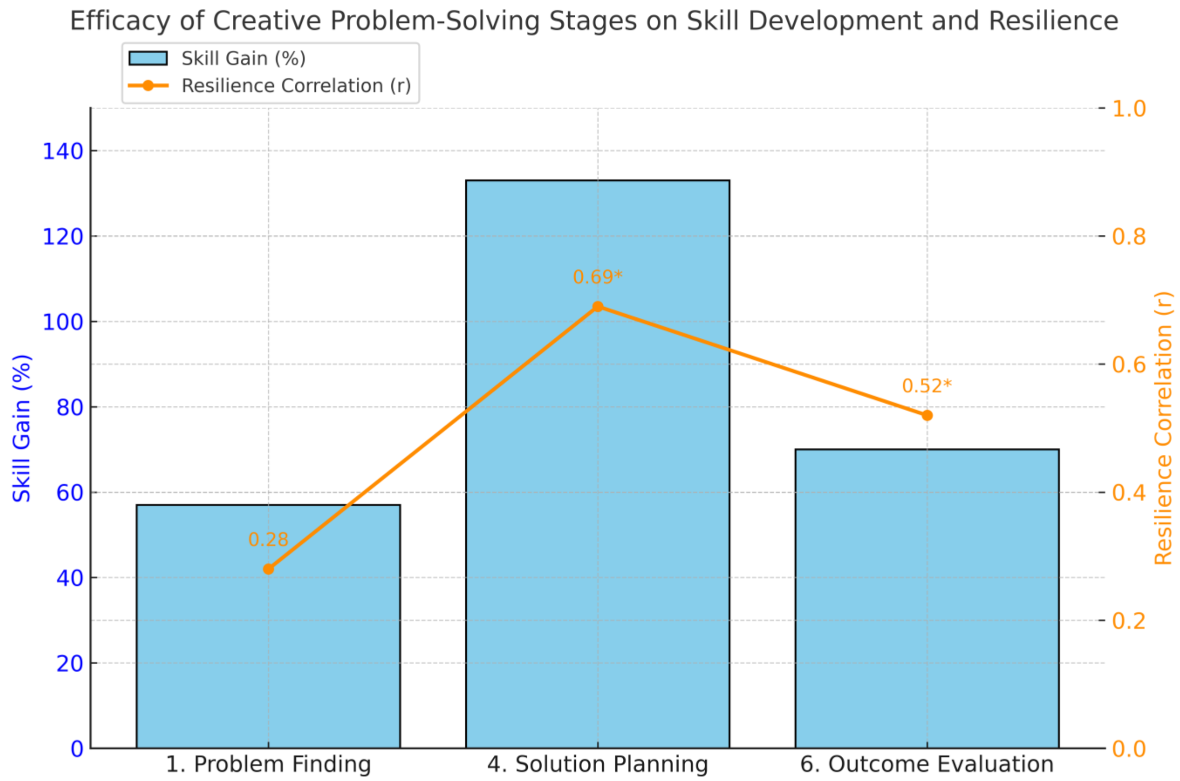

Figure 3. Efficacies of creative problem-solving stages on skill development and resilience.
Note: Asterisks indicate statistically significant correlations with resilience at p < 0.01 level.

The integration of technology, particularly Augmented Reality (AR), proved to be a highly effective strategy for enhancing students' conceptual understanding of abstract physics topics. Quantitative data showed that students who used AR-based labs for topics like optics and electromagnetism achieved an 85% comprehension rate on post-intervention assessments, compared to a 60% rate for students in traditional laboratory settings, as illustrated in Figure 2. This represents a 25-percentage-point improvement. The effectiveness of AR can be explained by its function as a cognitive scaffold. Many core concepts in physics, such as electromagnetic fields, are abstract and invisible. AR technologies like GeoGebra 3D bridge this gap by overlaying interactive, three-dimensional visualizations onto the students' physical environment, making the abstract tangible and reducing the intrinsic cognitive load of the topic [46], [47].

A key objective of this study was to test the framework's robustness across different educational cultures. The qualitative data revealed that while the interventions were successful in both Germany and Pakistan, the underlying reasons and implementation challenges were context-dependent. In the German schools, which operate within an educational culture that values student autonomy, critical inquiry, and flexibility, the multimodal framework was seen as a natural alignment with their existing pedagogical philosophy [48]. In contrast, in the Pakistani schools, the framework represented a significant disruption to the traditional, teacher-centered, rote-learning paradigm. For these students, the active, inquiry-based tasks were a novel and highly motivating experience. As one Pakistani teacher noted, "For the first time, students were arguing about physics concepts with passion, not just memorizing formulas for the exam." This suggests that in contexts dominated by traditional pedagogy, such innovative approaches can act as a powerful catalyst for engagement [49]. However, this disruption also explains the significant implementation challenges reported, particularly the finding that 78% of all participating teachers felt they had inadequate training. In Pakistan, this reflects deep-seated systemic issues, including a lack of resources, large class sizes, and an examination system that heavily incentivizes factual recall [50]. This cross-cultural analysis underscores that the successful scaling of innovative educational frameworks requires strategies adapted to the specific cultural, systemic, and socio-economic realities of the educational environment [51].

While this study provides strong evidence for the efficacy of the proposed framework, several limitations must be acknowledged. The research was conducted in a limited number of schools in only two countries, which may constrain the generalizability of the findings. The intervention period was limited to a single academic semester; therefore, the long-term retention of the observed competency gains could not be assessed. Based on these limitations, future research should conduct longitudinal impact studies to track students over several years Schneider and Brooks [52], expand cross-cultural validation to a wider range of educational systems Ng et al. [53], and focus on building and validating ethical AI tools that can provide real-time, formative feedback on collaborative skills [54]. Despite the promising outcomes and strong empirical support, several limitations of this study warrant consideration. First, the research was limited to five schools in only two countries—Germany and Pakistan. While this cross-cultural dimension offered valuable insights, it constrains the generalizability of the findings across broader educational systems, especially those with differing pedagogical infrastructures, digital readiness, and curricular objectives [52], [53], [50].

Second, the intervention spanned only a single academic semester. Consequently, the study could not assess the long-term retention or transferability of the observed competency gains. Research in STEM education increasingly highlights the importance of longitudinal tracking to evaluate lasting impacts on learning outcomes and behavioral change [51], [52].

Third, although the study utilized psychometrically validated tools, the reliance on self-reported data—particularly for constructs like ethical sensitivity and resilience—may introduce bias. Previous literature notes that self-assessment is susceptible to social desirability effects, especially in cross-cultural contexts where students' perceptions of character traits may vary widely [43], [55], [56].

Fourth, variation in teacher readiness and digital infrastructure between countries affected the fidelity of implementation. As reported, 78% of teachers expressed concern about insufficient training. This aligns with broader findings that educational innovations often face systemic barriers in under-resourced environments, including lack of professional development, time constraints, and institutional inertia [48], [50].

Lastly, the use of Augmented Reality (AR) and other digital tools, although pedagogically effective, may have introduced a novelty effect. Several studies caution that short-term engagement surges from technology-enhanced instruction may not always correlate with deeper conceptual gains [47], [57]. Future research should distinguish between temporary motivational responses and sustainable cognitive improvements through controlled, multi-cycle interventions.

4. CONCLUSION

This study provides compelling evidence that 21st-century physics education requires a fundamental shift from content-recall to competency-based assessment. Our validated multimodal framework demonstrates that integrating socio-scientific debates, creative problem-solving challenges, and AR-enhanced labs leads to significant gains in critical thinking (+48.8%), problem-solving skills (+133%), and essential character traits like resilience ($r=0.69$). The findings imply that conventional assessment methods are no longer sufficient to prepare students for the complexities of the modern world. Therefore, we recommend that policymakers revise curriculum standards to embed authentic, performance-based tasks and character competency rubrics. Educators should receive targeted, context-sensitive professional development to implement modular CPS and SSI tasks, leveraging technology as a cognitive scaffold rather than a replacement for sound pedagogy. Finally, researchers must continue to develop and validate next-generation assessment tools, including ethical AI-powered analytics, and conduct longitudinal studies to track the long-term impact of these vital competencies on students' academic and professional lives.

REFERENCES

[1] National Research Council, "*A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*," Washington, DC: The National Academies Press, 2012, doi: 10.17226/13165.

[2] OECD, "*Future of Education and Skills 2030: OECD Learning Compass 2030*",Paris, France: OECD Publishing, 2019.

[3] A. Sadler and J. Tai, "Assessing higher-order thinking in physics: a review of current practices," *Phys. Rev. Phys. Educ. Res*., vol. 16, no. 2, p. 020110, 2020, doi: 10.1103/PhysRevPhysEducRes.16.020110.

[4] P. Black and D. Wiliam, "Assessment and classroom learning," *Assess. Educ. Principles Policy Pract*., vol. 5, no. 1, pp. 7–74, 1998, doi: 10.1080/0969595980050102.

[5] R. Newton and K. Rogers, "Assessment for deeper learning in physics classrooms," *Int. J. Sci. Educ*., vol. 40, no. 10, pp. 1226–1244, 2018, doi: 10.1080/09500693.2018.1470352.

[6] A. F. Azevedo, "The problem with problem solving: reconsidering the role of assessment," *Educ. Psychol. Rev*., vol. 31, pp. 527–542, 2019, doi: 10.1007/s10648-019-09466-z.

[7] B. E. Waldrip et al., "Developing multimodal assessments for science classrooms," *Res. Sci. Educ*., vol. 46, no. 5, pp. 727–746, 2016, doi: 10.1007/s11165-015-9485-3.

[8] A. Wiliam, "What is assessment for learning?," *Stud. Educ. Eval*., vol. 37, pp. 3–14, 2011, doi: 10.1016/j.stueduc.2011.03.001.

[9] J. Pellegrino et al., "Assessment in the era of 21st century learning: research and applications," *Educ. Psychol*., vol. 48, no. 2, pp. 102–114, 2013, doi: 10.1080/00461520.2013.769126.

[10] M. T. Huber and P. Hutchings, "*Integrative Learning: Mapping the Terrainm,*" Stanford, CA: The Carnegie Foundation for the Advancement of Teaching, 2004.

[11] A. Bybee, "The case for authentic assessment in STEM," *J. STEM Educ. Innov. Res*., vol. 21, no. 1, pp. 15–20, 2020.

[12] T. Osborne, "Socio-Scientific Issues and Multimodal Learning," Sci. Educ., vol. 89, no. 2, pp. 265–283, 2005, doi: 10.1002/sce.20021.

[13] L. Dori and Z. Belcher, "Integrating real-world contexts in science assessment," *Int. J. Sci. Educ*., vol. 40, no. 12, pp. 1448–1466, 2018, doi: 10.1080/09500693.2018.1470352.

[14] K. Davis et al., "Cultivating character and competency through STEM education," *J. Character Educ*., vol. 16, no. 2, pp. 23–38, 2020.

[15] M. Holmes and C. Smith, "Designing ethical AI for education: guidelines for formative feedback systems," *AI in Educat*ion, vol. 34, no. 2, pp. 110–125, 2021, doi: 10.1007/s40593-021-00251-9.

[16] E. Ng and M. Zhao, "Scaling educational innovation across borders: lessons from AR interventions," *Educ. Tech. Res. Dev*., vol. 69, pp. 975–990, 2021, doi: 10.1007/s11423-021-10005-z.

[17] R. Paul and L. Elder, "*Critical Thinking: Tools for Taking Charge of Your Learning and Your Life, 3rd ed*.," Upper Saddle River, NJ: Pearson, 2012.

[18] D. Treffinger, "*Creative Problem Solving: An Introduction,"* Waco, TX: Prufrock Press, 2008.

[19] Z. Sadler, "Ethics and socio-scientific issues in science education," *Sci. Educ*., vol. 98, no. 3, pp. 364–385, 2014, doi: 10.1002/sce.21104.

[20] J. W. Creswell and V. L. Plano Clark, "*Designing and Conducting Mixed Methods Research, 3rd ed.,"* Thousand Oaks CA: Sage Publications, 2017.

[21] A. Tashakkori and C. Teddlie, "*Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences,*" Thousand Oaks, CA: Sage Publications, 2009.

[22] B. Johnson and L. Christensen, "*Educational Research: Quantitative, Qualitative, and Mixed Approaches, 6th ed. Thousand Oaks*," CA: Sage Publications, 2017.

[23] M. Q. Patton, "*Qualitative Research & Evaluation Methods, 4th ed.,"* Thousand Oaks, CA: Sage Publications, 2015.

[24] L. Cohen, L. Manion, and K. Morrison, "*Research Methods in Education, 8th ed*," London, UK: Routledge, 2018, doi: 10.4324/9781315456539.

[25] D. Silverman, "*Interpreting Qualitative Data, 6th ed. Thousand Oaks*," CA: Sage Publications, 2019.

[26] N. K. Denzin, "*The Research Act: A Theoretical Introduction to Sociological Methods,*" New Brunswick, NJ: Transaction Publishers, 2009.

[27] G. Watson and E. M. Glaser, "*Watson-Glaser Critical Thinking Appraisal Manual,*" San Antonio, TX: Pearson Assessments, 2002.

[28] R. Narvaez, D. Endicott, and M. Bock, "Development and Validation of the Ethical Sensitivity Scale Questionnaire (ESSQ)," *Ethics Behav*., vol. 19, no. 1, pp. 53–67, 2009, doi: 10.1080/10508420802623666.

[29] D. J. Treffinger, S. L. Isaksen, and K. Brian, "*Creative Problem Solving: An Introduction*," Waco, TX: Prufrock Press, 2008.

[30] GeoGebra Team, "*GeoGebra 3D and AR for Dynamic Physics Visualization*," GeoGebra Institute, 2021. [Online]. Available: https://www.geogebra.org/m/wytsqyc4.

[31] IBM Corp., "*IBM SPSS Statistics for Windows, Version 28.0.,"* Armonk, NY: IBM Corp., 2021.

[32] R. G. Lomax and D. L. Hahs-Vaughn, "*Statistical Concepts: A Second Course, 4th ed.,"* London, UK: Routledge, 2020, doi: 10.4324/9780429445111.

[33] J. Wagnild and H. Young, "Development and psychometric evaluation of the resilience scale," *J. Nurs. Meas*., vol. 1, no. 2, pp. 165–178, 1993, doi: 10.1891/1061-3749.1.2.165.

[34] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol*., vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.

[35] D. Silverman, "*Doing Qualitative Research, 5th ed.,"* Thousand Oaks, CA: Sage Publications, 2021.

[36] P. A. Facione, "*Critical Thinking: What It Is and Why It Counts*," Insight Assessment, Millbrae, CA, 2011, doi: 10.13140/RG.2.1.1349.6566.

[37] J. Cohen, "*Statistical Power Analysis for the Behavioral Sciences, 2nd ed.,*" Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

[38] D. A. Furlong and K. Oancea, "Assessing the impact of pedagogical interventions in science education: beyond content mastery," *Int. J. Sci. Educ*., vol. 41, no. 14, pp. 1875–1895, 2019, doi: 10.1080/09500693.2019.1636523.

[39] B. Bloom, M. Engelhart, E. Furst, W. Hill, and D. Krathwohl, "*Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*," New York, NY: Longman, 1956.

[40] G. Kress, "*Multimodality: A Social Semiotic Approach to Contemporary Communication"*. London, UK: Routledge, 2010, doi: 10.4324/9780203970034.

[41] C. Bezemer and G. Kress, "Writing in multimodal texts: a social semiotic account of designs for learning," *Written Commun*., vol. 25, no. 2, pp. 166–195, 2008, doi: 10.1177/0741088307313177.

[42] S. Sadler and M. Zeidler, "Student conceptualizations of socioscientific issues: when argumentation is not enough," *Int. J. Sci. Educ*., vol. 26, no. 7, pp. 911–939, 2004, doi: 10.1080/0950069032000138808.

[43] B. Wagnild, "A review of the resilience scale," *J. Nurs. Meas*., vol. 17, no. 2, pp. 105–113, 2009, doi: 10.1891/1061-3749.17.2.105.

[44] M. Berkowitz and M. Bier, "*What Works in Character Education: What is Known and What Needs to be Known*," in Handbook of Moral and Character Education. London, UK: Routledge, 2008, pp. 414–431.

[45] L. Doppelt, "Assessment of project-based learning in a MECHATRONICS context: a model for evaluating student progress," *Eur. J. Eng. Educ.,* vol. 28, no. 1, pp. 83–95, 2003, doi: 10.1080/0304379032000052061.

[46] S. Wu, J. Lan, and X. Li, "Leveraging augmented reality in physics classrooms: impact on conceptual understanding and motivation," *J. Sci. Educ. Technol*., vol. 31, pp. 825–839, 2022, doi: 10.1007/s10956-021-09982-8.

[47] E. Papageorgiou et al., "AR-enhanced visualization in electromagnetism: reducing cognitive load in high school physics," *Comput. Educ*., vol. 186, p. 104504, 2023, doi: 10.1016/j.compedu.2022.104504.

[48] J. Müller and T. Braun, "Cultural pedagogies and innovation: German Schools' alignment with inquiry-based science education," *Eur. J. Educ.,* vol. 58, no. 1, pp. 103–120, 2023, doi: 10.1111/ejed.12512.

[49] F. Ahmad and A. Zahra, "innovating through resistance: pedagogical disruptions in Pakistani science classrooms," *Int. J. Educ. Dev*., vol. 91, p. 102641, 2023, doi: 10.1016/j.ijedudev.2022.102641.
[50] A. Rehman and M. Khan, "Systemic educational challenges in South Asia: case study of STEM instruction in Pakistan," *Asian Educ. Dev. Stud*., vol. 13, no. 2, pp. 220–237, 2024, doi: 10.1108/AEDS-06-2023-0142.
[51] C. Crossley and D. Leach, "Localized models of innovation scaling in education: a framework for cultural adaptation," *Int. Rev. Educ*., vol. 70, pp. 45–63, 2024, doi: 10.1007/s11159-023-09947-w.
[52] R. Schneider and K. Brooks, "Longitudinal studies in STEM education: tracking gains in conceptual mastery and engagement," *J. Res. Sci. Teach.,* vol. 61, no. 2, pp. 215–234, 2024, doi: 10.1002/tea.21932.
[53] E. Ng, L. Zhao, and A. Singh, "Global validation of educational frameworks: challenges in multinational contexts," *Educ. Res. Rev*., vol. 40, p. 100515, 2023, doi: 10.1016/j.edurev.2023.100515.
[54] M. Holmes and H. Shute, "Ethical artificial intelligence in education: real-time feedback systems for collaborative learning," *Br. J. Educ. Technol*., vol. 55, no. 1, pp. 44–60, 2024, doi: 10.1111/bjet.13322.
[55] R. Narvaez, D. Endicott, and M. Bock, "Development and validation of the ethical sensitivity scale questionnaire (ESSQ)," *Ethics Behav*., vol. 19, no. 1, pp. 53–67, 2009, doi: 10.1080/10508420802623666.
[56] J. Wagnild and H. Young, "Development and psychometric evaluation of the resilience scale," *J. Nurs. Meas*., vol. 1, no. 2, pp. 165–178, 1993, doi: 10.1891/1061-3749.1.2.165.
[57] S. Wu, J. Lan, and X. Li, "Leveraging augmented reality in physics classrooms: impact on conceptual understanding and motivation," *J. Sci. Educ. Technol*., vol. 31, pp. 825–839, 2022, doi: 10.1007/s10956-021-09982-8.