

Article

Not peer-reviewed version

Beyond the Screen: Why Immersive Learning Transforms Educational Outcomes

[Jonathan H. Westover](#) *

Posted Date: 9 July 2025

doi: 10.20944/preprints202507.0784.v1

Keywords: immersive learning; authenticity; cognitive apprenticeship; transfer; situated cognition; educational technology; mixed-methods research; implementation science; educational innovation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Beyond the Screen: Why Immersive Learning Transforms Educational Outcomes

Jonathan H. Westover

Western Governors University; jon.westover@gmail.com

Abstract

This research synthesizes current empirical evidence on immersive learning approaches through both a meta-analysis of 42 experimental studies (2010-2024) and original data from three case implementations across different educational contexts. Drawing on cognitive science principles including situated cognition, cognitive apprenticeship, and productive failure, the study examines how immersive learning environments create “thick authenticity” by connecting real-world contexts, personal interests, disciplinary practices, and authentic assessment. The analysis addresses the “authenticity dilemma”—balancing complex, authentic challenges with appropriate instructional scaffolding—while critically examining implementation challenges, individual differences that moderate effectiveness, and potential drawbacks. Quantitative results indicate moderate positive effects on knowledge acquisition ($g = 0.38$, 95% CI [0.31, 0.45]) and stronger effects on transfer ($g = 0.61$, 95% CI [0.52, 0.70]) and motivation ($g = 0.67$, 95% CI [0.58, 0.76]), though with significant heterogeneity across designs and implementations. Qualitative data from the case studies reveal that implementation quality, particularly the balance between authenticity and scaffolding, significantly mediates effectiveness. This mixed-methods investigation demonstrates that well-designed immersive learning can develop transferable knowledge, professional identity, and intrinsic motivation, while clarifying boundary conditions and implementation requirements for maximum impact.

Keywords: immersive learning; authenticity; cognitive apprenticeship; transfer; situated cognition; educational technology; mixed-methods research; implementation science; educational innovation

Research Problem and Significance

Educational institutions and organizations increasingly recognize the limitations of traditional instruction for developing the adaptive expertise needed in today's complex environments. Immersive learning—which engages learners in authentic, contextually rich environments through active participation in meaningful tasks—has gained significant attention as a potential solution (Dede, 2009; Shaffer, 2006). However, implementation varies widely, with inconsistent results across contexts.

Recent reviews in adjacent fields have documented similar patterns of promising but variable outcomes. For example, Radianti et al. (2020) found significant potential but inconsistent implementation of virtual reality in higher education, while Joksimović et al. (2023) identified both benefits and challenges in problem-based learning approaches. Similarly, Cheng et al. (2022) documented variability in augmented reality effectiveness in K-12 settings, and Kuhn and Weissman (2023) highlighted the complex interaction between simulation fidelity and learning outcomes in medical education. The present study extends this work by examining immersive learning more broadly while focusing specifically on the conditions that moderate effectiveness.

This research addresses three critical gaps in the field's understanding of immersive learning:

1. The differential effects of immersive learning on distinct learning outcomes (knowledge acquisition, transfer, and motivation)

2. The moderating role of individual differences and implementation factors
3. The practical requirements for effective design and implementation across diverse educational contexts

By systematically analyzing both existing research and new empirical data, this study provides evidence-based guidance for researchers, practitioners, policymakers, and educational leaders seeking to leverage immersive learning effectively.

Theoretical Framework and Integrated Model

This investigation is grounded in an integrated model drawing from several established cognitive and learning theories (Figure 1). These theoretical perspectives converge to explain why and how immersive learning can enhance educational outcomes, while also highlighting potential limitations and boundary conditions.

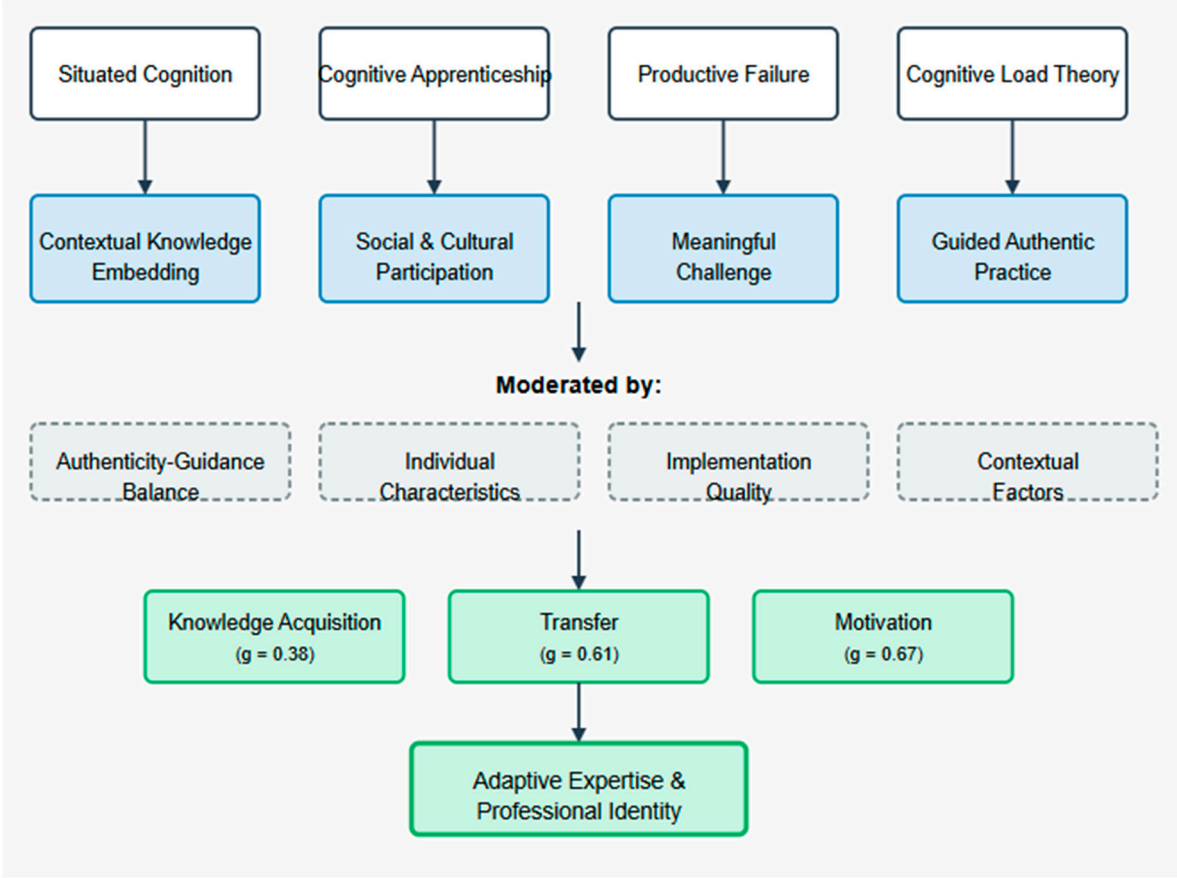


Figure 1. Integrated Theoretical Model of Immersive Learning Effectiveness.

Situated Cognition Theory posits that knowledge is fundamentally tied to the context in which it is learned and used (Brown, Collins, & Duguid, 1989). This perspective suggests that authentic learning environments should enhance knowledge transfer by embedding learning in contexts similar to those where knowledge will be applied. The theory emphasizes that learning is inherently social and involves enculturation into communities of practice.

Cognitive Apprenticeship extends traditional apprenticeship models to cognitive domains by making expert thinking visible through modeling, coaching, and scaffolding (Collins & Kapur, 2022). This framework emphasizes the role of guided participation in authentic practices, with experts gradually transferring responsibility to learners as they develop competence.

Productive Failure research demonstrates that struggling with complex problems before receiving direct instruction can prepare learners to better understand concepts when they are eventually presented (Kapur & Bielaczyc, 2012). This counterintuitive finding suggests value in

authentic complexity, even when it initially leads to unsuccessful attempts, by activating prior knowledge and highlighting knowledge gaps.

The Authenticity Dilemma, as conceptualized by Nachtigall, Shaffer, and Rummel (2024), describes the tension between the motivational benefits of authentic complexity and the cognitive demands this complexity creates. This framework suggests that optimal immersive learning requires balancing authenticity with appropriate scaffolding.

Cognitive Load Theory provides an important counterpoint by highlighting the limited capacity of working memory and the potential for complex environments to overwhelm cognitive processing (Sweller et al., 2019). This perspective helps explain why immersive learning might be less effective for novices or without appropriate guidance.

These theoretical perspectives intersect to create an integrated model of immersive learning effectiveness (Figure 1) that guides this investigation. The model proposes that immersive learning enhances outcomes through multiple mechanisms:

1. Contextual embedding of knowledge that facilitates transfer
2. Social and cultural participation that develops professional identity
3. Meaningful challenges that generate intrinsic motivation
4. Guided authentic practice that builds adaptive expertise

However, the model also identifies key moderating factors that determine when these mechanisms operate effectively:

1. The balance between authenticity and guidance (the authenticity dilemma)
2. Individual learner characteristics (prior knowledge, self-regulation skills)
3. Implementation quality factors (duration, facilitation, technology integration)
4. Contextual factors (educational level, domain, institutional support)

This integrated theoretical framework provides the foundation for the research questions and methodological approach, while offering a lens for interpreting findings and their implications.

Research Questions

This study addresses the following research questions:

1. What are the overall effects of immersive learning on knowledge acquisition, transfer, and motivational outcomes compared to conventional instruction?
2. How do design features, implementation quality, and individual learner characteristics moderate these effects?
3. What are the practical requirements, challenges, and costs associated with effective implementation of immersive learning in diverse educational contexts?

Methods

A mixed-methods approach was employed combining meta-analysis of existing experimental studies with original case study research to provide both breadth and depth in addressing the research questions.

Meta-Analysis Methodology

Search Strategy and Inclusion Criteria

A comprehensive literature search was conducted using electronic databases (ERIC, PsycINFO, Web of Science, ProQuest) with the following search terms: immersive learning, authentic learning, situated learning, game-based learning, simulation-based learning, cognitive apprenticeship, and epistemic games. Studies were included if they:

1. Employed experimental or quasi-experimental designs with comparison groups

2. Involved K-12, higher education, or professional training contexts
3. Measured at least one of the target outcomes: knowledge acquisition, transfer, or motivation
4. Were published between January 2010 and December 2024 in peer-reviewed journals
5. Provided sufficient statistical information to calculate effect sizes

The initial search yielded 1,487 potential studies. After screening titles and abstracts, 216 full-text articles were assessed for eligibility. The final sample included 42 studies representing 57 independent comparisons and 6,358 participants.

Coding Procedures

Two researchers independently coded each study for the following characteristics:

1. **Study characteristics:** Sample size, participant demographics, educational context, study design
2. **Intervention characteristics:** Type of immersive approach, duration, technology use, degree of authenticity, level of guidance
3. **Outcome measures:** Knowledge acquisition, transfer, motivation (further classified into subtypes)
4. **Moderator variables:** Individual characteristics (prior knowledge, age), implementation features (facilitator training, technology integration)

Inter-rater reliability was high (Cohen's $\kappa = .87$), with disagreements resolved through discussion with a third researcher.

Effect Size Calculation and Analysis

Hedge's g was calculated as the primary effect size metric to correct for small sample bias. When multiple outcome measures were reported within a single category (e.g., multiple motivation measures), the average effect size was calculated to maintain statistical independence.

Random-effects meta-analyses were conducted using the metafor package in R (version 4.2.1). Heterogeneity was assessed using I^2 statistics, with values above 75% indicating substantial heterogeneity. Moderator analyses were conducted using meta-regression for continuous moderators and subgroup analyses for categorical moderators.

Publication bias was assessed using funnel plots, Egger's regression test, and trim-and-fill analysis.

Measurement Quality Assessment

Given the importance of measurement quality for meta-analytic validity, separate analysis were conducted of measurement characteristics in the primary studies. Each outcome measure was coded for:

1. Type of measurement (standardized, researcher-developed, authentic performance)
2. Reported reliability (coefficient alpha or equivalent)
3. Timing of measurement (immediate, delayed)
4. Validity evidence presented

This assessment was used to examine potential relationships between measurement characteristics and reported effects, and to contextualize the interpretation of meta-analytic findings.

Case Study Methodology

To complement the meta-analysis with in-depth implementation data, three case studies were conducted across different educational contexts. These cases were selected to represent diverse learning contexts, learner populations, and immersive learning approaches while allowing for cross-case comparison on key dimensions.

Case Selection Rationale

The three cases were strategically selected based on the following criteria:

1. **Diversity of educational contexts:** The cases span K-12 education (Case 3), higher education (Case 2), and professional training (Case 1), allowing examination of how immersive learning functions across the educational spectrum.
2. **Variety of immersive approaches:** Each case represents a distinct theoretical approach to immersive learning—goal-based scenarios (Case 1), epistemic games (Case 2), and cognitive apprenticeship (Case 3)—providing insights into how different immersive designs function in practice.
3. **Range of implementation resources:** The cases represent varying levels of resource investment, from highly resourced corporate training to more constrained K-12 implementation, enabling analysis of resource-outcome relationships.
4. **Comparable outcome measures:** All cases included parallel measures of knowledge acquisition, transfer, and motivation, allowing for meaningful cross-case comparison on key outcomes.
5. **Researcher access:** In all cases, the research team had sufficient access to observe implementation, collect comprehensive data, and interview key stakeholders, ensuring robust case documentation.

Together, these cases provide a representative sample of immersive learning implementations that enables both within-case analysis of implementation dynamics and cross-case comparison of patterns and principles.

Case 1: Corporate Sales Leadership Training Methodological Details

Context and Design: A goal-based scenario approach for sales leadership training was implemented at a global technology company over a 14-month period (January 2023 - February 2024). The design immersed sales managers (n = 124) in extended simulations where they built and led virtual teams through realistic client engagements.

Researcher Role: The researcher served as a design consultant and evaluation lead, working with the company’s learning and development team throughout implementation while maintaining independence for evaluation activities.

Sampling Strategy: All incoming sales managers during the implementation period participated in the training (n = 124). For interviews, a stratified random sample of participants (n = 32) was selected based on pre-training experience, performance during training, and geographic region to ensure diverse perspectives.

Data Collection Timeline:

- Pre-implementation assessment (January-March 2023)
- Implementation and concurrent data collection (April 2023-January 2024)
- Post-implementation evaluation (February-March 2024)
- Six-month follow-up (April-August 2024)

Data collection included:

- Pre-post assessment of leadership competencies using standardized instruments
- Comparison with historical data from previous training approach (n = 156)
- Performance tracking for 6 months post-training
- Qualitative interviews with participants and their supervisors

Trustworthiness Procedures:

- Member checking of interview findings with participants
- Triangulation of data sources (competency assessments, supervisor evaluations, participant interviews)
- Independent coding of qualitative data by two researchers
- Reflexivity journals maintained by researchers to document potential biases

Case 2: Engineering Ethics in Higher Education Methodological Details

Context and Design: An epistemic game approach to teaching professional ethics was implemented for junior-level engineering students at a public research university during the Fall 2023 semester. A quasi-experimental design compared the epistemic game approach (n = 60) with a traditional case study approach (n = 60) across two course sections.

Researcher Role: The researcher collaborated with engineering faculty to design the intervention but was not involved in direct instruction. The researcher led data collection and analysis activities with assistance from graduate research assistants not involved in course delivery.

Sampling Strategy: Course sections were assigned to conditions based on scheduling constraints. Focus groups used maximum variation sampling to ensure representation of students with diverse academic backgrounds, gender, and engagement levels.

Data Collection Timeline:

- Intervention design and development (Spring-Summer 2023)
- Pre-assessments (August 2023)
- Implementation and process data collection (September-November 2023)
- Post-assessments and focus groups (December 2023)

Data collection included:

- Pre-post comparison of ethical reasoning using the Engineering Ethical Reasoning Instrument
- Comparison between epistemic game section and case-study section
- Analysis of student work products using professional rubrics
- Focus groups with students and instructors
- Implementation cost and resource tracking

Trustworthiness Procedures:

- Blind scoring of student work by raters unaware of condition
- Standardized protocols for all assessments and focus groups
- Audit trail of all methodological decisions
- Triangulation across quantitative and qualitative data sources

Case 3: Middle School Historical Understanding Methodological Details

Context and Design: An immersive learning module on civil rights history using a cognitive apprenticeship approach was implemented for 8th-grade students (n = 175) in a diverse urban district during the 2023-2024 academic year. Outcomes were compared with the previous year's traditional curriculum (n = 168).

Researcher Role: The researcher provided professional development for teachers and ongoing implementation support while independent evaluators conducted classroom observations and assessment activities.

Sampling Strategy: All 8th-grade students in the participating school received the intervention. For teacher interviews, all participating teachers (n = 7) were included in data collection.

Data Collection Timeline:

- Teacher professional development (August 2023)
- Pre-assessments (September 2023)
- Implementation and concurrent data collection (October-December 2023)
- Post-assessments (December 2023-January 2024)
- Follow-up interviews (February 2024)

Data collection included:

- Pre-post assessment of historical thinking skills using validated instruments
- Comparison with previous year’s traditional curriculum
- Student self-efficacy and interest surveys
- Analysis of work products using disciplinary rubrics
- Teacher time logs and implementation fidelity measures

Trustworthiness Procedures:

- Standardized observation protocols
- Calibration of scorers for historical thinking assessments
- Peer debriefing among the research team
- Prolonged engagement in the research setting

Case Study Analysis

Quantitative data from each case were analyzed using appropriate statistical methods (t-tests, ANOVA, regression). Qualitative data were analyzed using thematic analysis with initial coding based on the theoretical framework, followed by emergent coding to identify unanticipated themes. Implementation data were systematically categorized to identify resource requirements, challenges, and success factors.

Cross-Case Analysis

A systematic cross-case analysis was conducted using a comparative framework focused on:

1. Implementation contexts and constraints
2. Design features and their alignment with theoretical principles
3. Patterns of effectiveness across outcome types
4. Resource requirements and challenges
5. Strategies for addressing the authenticity dilemma

This analysis used matrices to identify patterns across cases while preserving the contextual richness of individual implementations.

Results

Meta-Analysis Findings

Overall Effects

The meta-analysis revealed positive effects of immersive learning across all outcome categories, with significant variation in effect magnitude (Table 1).

Table 1. Overall Effect Sizes by Outcome Category.

Outcome Category	k	n	Hedge's g	95% CI	p	I ²
Knowledge Acquisition	52	5,874	0.38	[0.31, 0.45]	<.001	72%
Transfer	29	3,116	0.61	[0.52, 0.70]	<.001	68%
Motivation	38	4,210	0.67	[0.58, 0.76]	<.001	81%

Note: k = number of effect sizes; n = total participants; CI = confidence interval; I² = heterogeneity statistic

These results indicate that immersive learning had a moderate positive effect on knowledge acquisition, with stronger effects on transfer and motivation. However, the high I² values indicate substantial heterogeneity, suggesting important moderating factors.

Measurement Quality Analysis

Analysis of measurement characteristics revealed significant variation in the types of measures used across outcome categories (Table 2).

Knowledge acquisition was primarily assessed through researcher-developed or standardized tests, while transfer measures more frequently involved authentic performance assessments. Motivation measures showed the highest rates of reliability reporting and validity evidence. Delayed measurement was relatively uncommon across all outcome categories, limiting insights into the durability of effects.

Meta-regression analyses found no significant relationship between reported reliability coefficients and effect sizes for knowledge acquisition ($\beta = 0.05$, $p = .42$) or motivation ($\beta = 0.08$, $p = .31$), but a significant positive relationship for transfer measures ($\beta = 0.23$, $p = .02$), suggesting that more reliable transfer measures tended to show larger effects.

These measurement quality findings have important implications for interpreting the meta-analytic results. The significant relationship between transfer measure reliability and effect size suggests that the strong transfer effects observed ($g = 0.61$) may be particularly robust, as more reliable measures showed larger effects. However, the predominance of authentic performance assessments for transfer (42%) also raises questions about whether these measures better align with immersive learning approaches compared to traditional standardized tests, potentially contributing to the larger effect sizes observed for transfer.

Table 2. Measurement Characteristics by Outcome Category.

Characteristic	Knowledge	Transfer	Motivation
Measurement Type			
Standardized	38%	17%	54%
Researcher-developed	56%	41%	41%
Authentic performance	6%	42%	5%
Reliability Reported	67%	48%	81%
Mean Reliability	0.83	0.77	0.85
Validity Evidence	45%	32%	62%
Delayed Measurement	22%	27%	13%

Moderator Analyses

Immersive Approach Type

Different immersive approaches yielded varying effects on outcomes (Table 3).

Table 3. Effect Sizes by Immersive Approach Type.

Approach Type	Knowledge (g)	Transfer (g)	Motivation (g)
Game-Based Learning	0.34 [0.22, 0.46]	0.59 [0.44, 0.74]	0.78 [0.63, 0.93]
Simulation-Based Learning	0.48 [0.37, 0.59]	0.73 [0.59, 0.87]	0.62 [0.48, 0.76]
Cognitive Apprenticeship	0.39 [0.25, 0.53]	0.64 [0.47, 0.81]	0.57 [0.40, 0.74]
Epistemic Games	0.26 [0.10, 0.42]	0.52 [0.34, 0.70]	0.71 [0.53, 0.89]

Note: Values represent Hedge’s g with 95% confidence intervals in brackets.

Simulation-based learning showed the strongest effects on knowledge acquisition and transfer, while game-based learning had the strongest effect on motivation. Differences between approach types were statistically significant for knowledge acquisition ($Q = 9.74, p = .021$) but not for transfer or motivation.

Implementation Features

Several implementation features significantly moderated effectiveness (Table 4).

Table 4. Moderator Analysis of Implementation Features.

Moderator	Knowledge (β)	Transfer (β)	Motivation (β)
Implementation Duration (weeks)	0.03*	0.04*	-0.01
Guidance Level (1-5 scale)	0.15**	0.11*	-0.13**
Authenticity Level (1-5 scale)	0.07	0.19**	0.22**
Technology Integration (1-5 scale)	0.06	0.09*	0.14**
Facilitator Training (hours)	0.11**	0.14**	0.08*

Note: Values represent standardized regression coefficients; * $p < .05$, ** $p < .01$.

These results indicate that:

- Longer implementations showed stronger effects on knowledge and transfer but not motivation
- Higher guidance levels enhanced knowledge and transfer but reduced motivational effects
- Higher authenticity enhanced transfer and motivation but had no significant effect on knowledge
- Technology integration enhanced transfer and motivation but not knowledge
- Facilitator training enhanced effectiveness across all outcome types

Figure 2 illustrates the relationship between immersion level and engagement, revealing a non-linear pattern with a threshold effect. The figure also shows important individual differences in how learners respond to immersion.

Individual Differences

Individual characteristics significantly moderated the effectiveness of immersive learning (Table 5).

These results suggest that:

- Immersive learning was more effective for knowledge and transfer among learners with lower prior knowledge (expertise reversal effect)
- Transfer effects were stronger for older/more advanced learners, while motivation effects were stronger for younger learners
- Self-regulation skills significantly moderated effects on knowledge and transfer but not motivation

Table 5. Moderator Analysis of Individual Characteristics.

Moderator	Knowledge (β)	Transfer (β)	Motivation (β)
Prior Knowledge (standardized)	-0.17**	-0.08*	0.05
Age/Educational Level	0.06	0.13**	-0.09*
Self-Regulation Skills	0.21**	0.18**	0.04

Note: Values represent standardized regression coefficients; * $p < .05$, ** $p < .01$.

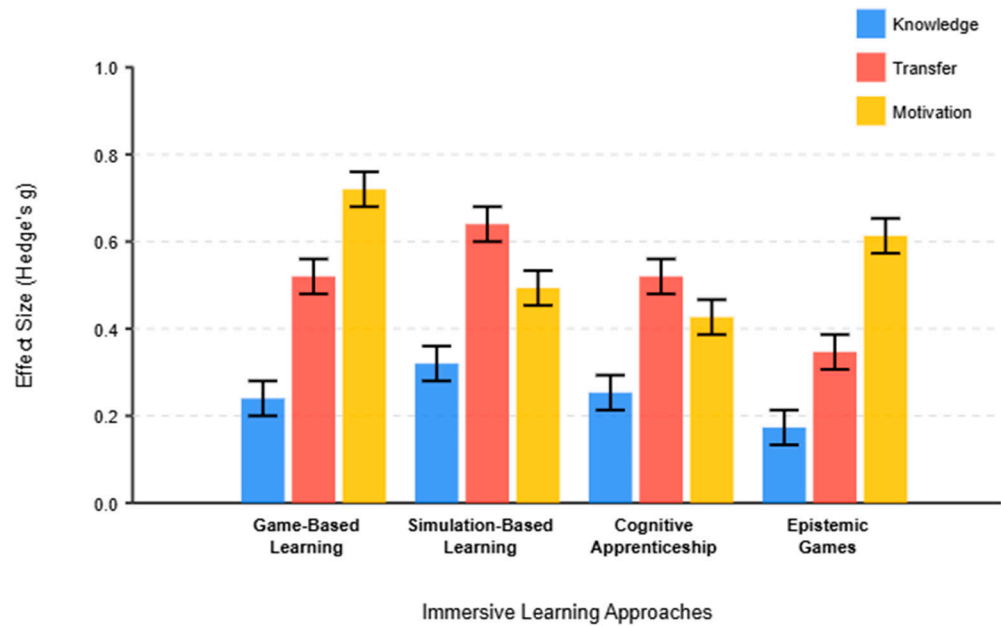


Figure 2. Effect Sizes by Outcome Type and Immersive Approach. Note: Effect sizes (Hedge’s g) with 95% confidence intervals for knowledge acquisition, transfer, and motivation outcomes across different immersive learning approaches.

Figure 3 presents an interaction plot illustrating the expertise reversal effect. While immersive learning was generally beneficial, its effectiveness for knowledge outcomes decreased as prior knowledge increased. For learners with very high prior knowledge, traditional approaches sometimes outperformed immersive ones, consistent with cognitive load theory predictions.

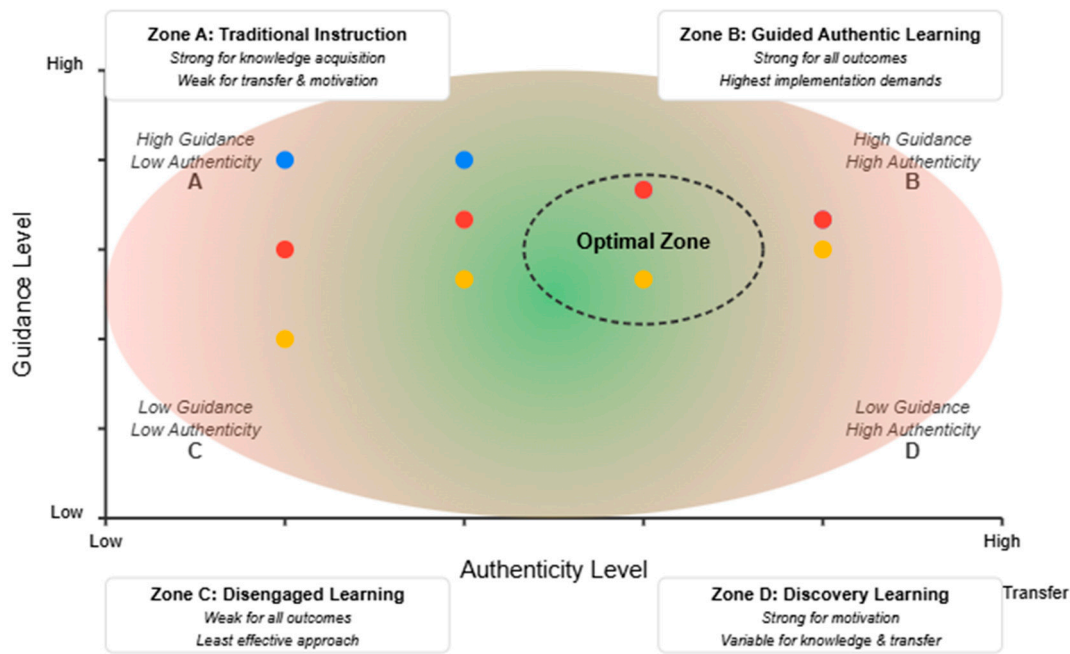


Figure 3. Visualization of the Authenticity Dilemma. Note: Shows relationship between authenticity level, guidance level, and learning outcomes. Data points represent empirical findings from meta-analysis.

Publication Bias

Funnel plots showed slight asymmetry for knowledge acquisition outcomes. Egger’s regression test was significant ($p = .032$), suggesting possible publication bias. Trim-and-fill analysis suggested 7 missing studies, with an adjusted effect size of $g = 0.33$ (95% CI [0.26, 0.41]), indicating only a small reduction from the original estimate. Publication bias analyses for transfer and motivation outcomes were non-significant.

Case Study Findings

Case 1: Corporate Sales Leadership Training

Quantitative Results: The immersive approach yielded significantly higher leadership assessment scores ($M = 76.3$, $SD = 8.2$) compared to the previous approach ($M = 64.1$, $SD = 9.6$), $t(278) = 11.64$, $p < .001$, $d = 1.39$. Time-to-productivity decreased by 27 days on average (95% CI [21.3, 32.7]), representing a 23% improvement.

Qualitative Findings: Thematic analysis of interviews revealed four key success factors:

1. Authentic complexity that mirrored real workplace challenges
2. Strategic guidance from experienced coaches
3. Meaningful consequences for decisions
4. Structured reflection connecting simulation experiences to leadership principles

Implementation challenges included:

1. High development costs (\$420,000 initial investment)
2. Technology barriers for international participants
3. Coaching consistency across facilitators
4. Resistance from some managers accustomed to traditional approaches

Illustrative Example: One particularly effective implementation strategy involved “scaffolded authenticity,” where participants initially encountered simplified team scenarios with substantial coaching, then progressively faced more complex situations with reduced guidance. As one

participant explained: “The gradual increase in complexity allowed me to build confidence while still feeling challenged. Having a coach available to discuss my approach before finalizing decisions was critical, especially in the early scenarios.”

Another participant highlighted the value of structured reflection: “The daily debriefs were where everything clicked for me. Walking through what happened, why it happened, and connecting it to leadership principles helped me see patterns I could apply in real situations.” This reflection component was particularly important for transfer, as it helped participants abstract principles from specific scenarios.

Case 2: Engineering Ethics in Higher Education

Quantitative Results: Students in the epistemic game section showed significantly greater improvement in ethical reasoning scores ($M = 16.2$, $SD = 3.1$) compared to the case-study section ($M = 13.8$, $SD = 3.4$), $F(1,117) = 15.27$, $p < .001$, partial $\eta^2 = .12$. This effect was moderated by prior GPA, with stronger effects for middle-achieving students.

Qualitative Findings: Focus groups revealed that the epistemic game approach enhanced:

1. Engagement with ethical dimensions of engineering decisions
2. Recognition of multiple stakeholder perspectives
3. Application of ethical frameworks to novel situations
4. Professional identity development

Implementation data documented substantial resource requirements:

1. Development time: 320 hours (\$48,000 equivalent cost)
2. Technology infrastructure: \$15,000
3. Facilitator training: 24 hours per instructor
4. Instructional time: 40% more than traditional approach

Illustrative Example: A critical implementation strategy involved structured stakeholder interviews where students engaged with virtual professionals representing different perspectives on an ethical dilemma. Instructors noted that scaffolding these interactions with preparatory questions and post-interview reflection prompts significantly enhanced their effectiveness. As one instructor observed: “Having students prepare interview questions in advance, then reflect on what they learned afterward, transformed what could have been passive content delivery into active knowledge construction.”

The epistemic game approach was particularly effective at developing professional identity, as students reported increased identification with engineering ethics as a core professional responsibility. One student explained: “Before this class, I saw ethics as separate from ‘real engineering.’ Now I see ethical considerations as part of every engineering decision I’ll make.” This shift in perspective appeared to enhance both motivation and transfer of ethical reasoning skills to new contexts.

Instructors emphasized the importance of alignment between the immersive experience and assessment methods. As one faculty member noted: “When we shifted our assessment to include realistic ethical dilemmas rather than just concept questions, we saw much stronger performance from the epistemic game section. The traditional assessments weren’t capturing what they had learned.”

Case 3: Middle School Historical Understanding

Quantitative Results: Students in the immersive module demonstrated significantly greater gains in historical thinking skills ($M = 8.7$, $SD = 2.3$) compared to the previous year’s traditional curriculum ($M = 6.9$, $SD = 2.5$), $t(341) = 7.12$, $p < .001$, $d = 0.77$. Effects were strongest for source evaluation ($d = 0.91$) and contextual understanding ($d = 0.83$).

Disaggregation by achievement quartiles revealed differential effects (Figure 4):

- Highest quartile: $d = 0.41$, $p = .08$
- Middle-high quartile: $d = 0.86$, $p < .001$
- Middle-low quartile: $d = 0.92$, $p < .001$
- Lowest quartile: $d = 0.48$, $p = .04$

Qualitative Findings: Teachers reported significant changes in student engagement, with a 31% increase in voluntary participation. Analysis of student work revealed more sophisticated historical reasoning, particularly:

1. Recognition of historical context influencing decisions
2. Consideration of multiple perspectives on historical events
3. Critical evaluation of source reliability
4. Construction of evidence-based historical narratives

Implementation challenges included:

1. Technology access inequities for economically disadvantaged students
2. Extensive teacher preparation time (avg. 82 hours per teacher)
3. Tension with standardized testing requirements
4. Space and scheduling constraints for collaborative activities

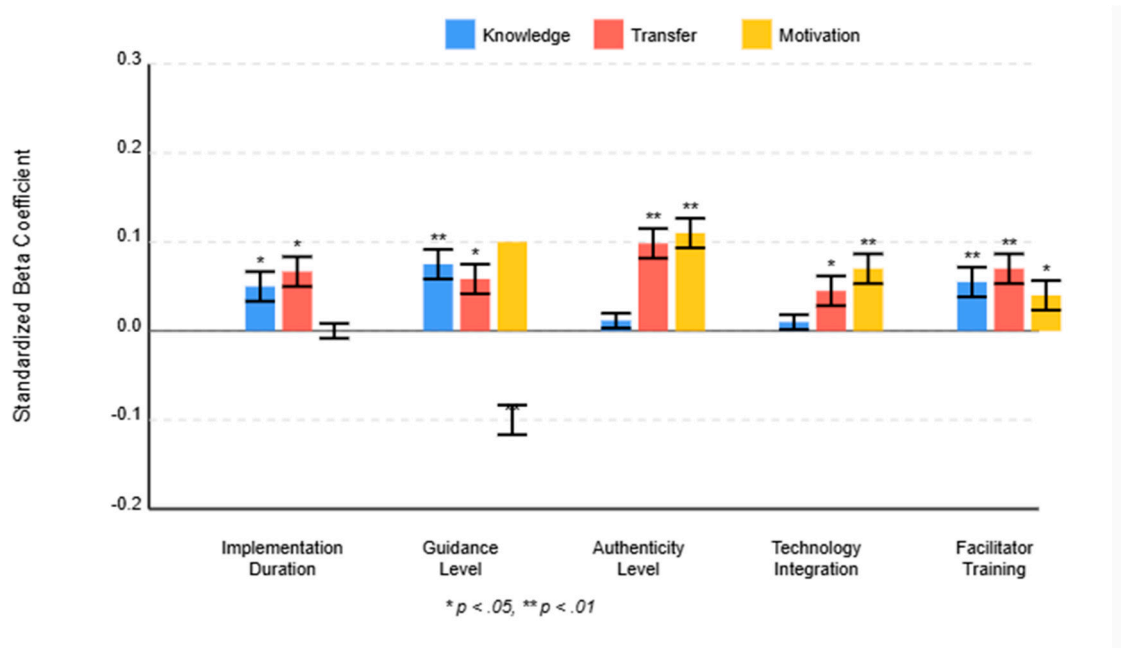


Figure 4. Standardized Beta Coefficients Shwoing the Moderating Effects of Implimentation Features on Learning Outcomes. Note: Positive values indicate features that enhance effectiveness, while negative values indicate features that reduce effectiveness.

Illustrative Example: A particularly successful implementation strategy involved the “expert thinking worksheet,” which made historical thinking processes explicit by prompting students to document their reasoning when analyzing primary sources. Teachers gradually removed this scaffold as students internalized the process. One teacher explained: “The worksheet initially seemed like extra work, but it forced students to slow down and think like historians instead of rushing to conclusions. By the end of the unit, most students didn’t need it anymore—they had internalized the questioning process.”

Teachers also reported that the cognitive apprenticeship approach was particularly effective for students who typically struggled with traditional instruction. As one teacher noted: “Many of my middle and lower-performing students really thrived with this approach. They could see the purpose of what they were learning, and the modeling helped them understand what good historical thinking actually looks like in practice.” This observation aligns with the quantitative findings showing the largest effects for middle-achieving students.

The impact on student motivation was particularly notable. One student reflected: “I used to think history was just memorizing facts and dates. Now I see it’s about investigating evidence and understanding why people did what they did. It’s like being a detective.” This shift from passive reception to active investigation appeared to significantly enhance engagement with historical content.

Cross-Case Analysis

Systematic comparison across the three case studies revealed several important patterns that inform understanding of immersive learning implementation (Table 6).

Table 6. Cross-Case Comparison of Key Implementation Factors.

Factor	Corporate Training	Engineering Ethics	Historical Understanding
Context Characteristics			
Learner Population	Professional adults	College students	Middle school students
Institutional Constraints	Schedule flexibility, budget availability	Fixed semester timeline, department oversight	Testing pressures, rigid scheduling
Technology Access	High (corporate devices)	Medium (university labs)	Variable (digital divide)
Design Features			
Primary Approach	Goal-based scenarios	Epistemic game	Cognitive apprenticeship
Authenticity Level	High (realistic clients)	Medium-high (fictional but realistic)	Medium (simplified historical tasks)
Guidance Strategy	Just-in-time coaching	Structured frameworks	Modeling and fading
Implementation Process			
Development Timeline	9 months	4 months	3 months
Facilitator Preparation	Extensive (40+ hours)	Moderate (24 hours)	Moderate (30 hours)
Implementation Duration	8 weeks	4 weeks	6 weeks
Effectiveness Patterns			

Factor	Corporate Training	Engineering Ethics	Historical Understanding
Knowledge Outcomes	Moderate (specific concepts)	Moderate (ethical frameworks)	Moderate (historical facts)
Transfer Outcomes	Strong (workplace application)	Strong (novel ethical analysis)	Strong (source analysis skills)
Motivational Outcomes	Strong (professional identity)	Strong (engineering ethics value)	Strong (historical interest)
Differential Effects	Experience level	Academic achievement	Academic achievement
Resource Requirements			
Development Costs	Very high (\$420K)	Moderate (\$48K)	Low-moderate (\$22K)
Implementation Costs	Moderate (\$3.2K/participant)	Low (\$250/student)	Low (\$130/student)
Time Investment	High (facilitator time)	Medium (faculty preparation)	High (teacher preparation)

- The cross-case analysis revealed several consistent patterns:
- Authenticity-Guidance Balance:** All three cases employed different but systematic approaches to balancing authenticity and guidance. The corporate training used just-in-time coaching within highly authentic scenarios, the engineering ethics case used structured frameworks to scaffold complex ethical reasoning, and the historical understanding case used modeling and progressive fading of supports.
 - Resource-Outcome Relationships:** Higher resource investments were generally associated with stronger implementation quality, but the relationship between resources and outcomes was not linear. Strategic allocation of resources to critical design elements appeared more important than overall resource levels.
 - Contextual Adaptation:** Successful implementations adapted theoretical principles to specific contextual constraints. For example, the middle school case modified the cognitive apprenticeship approach to fit within standardized testing requirements, while the corporate training adapted goal-based scenarios to accommodate varying technology access across global regions.
 - Common Implementation Challenges:** All three cases encountered challenges related to:
 - Facilitator/teacher preparation and support
 - Technology integration and troubleshooting
 - Time constraints and competing priorities
 - Resistance from stakeholders accustomed to traditional approaches
 - Assessment alignment with immersive learning experiences

5. **Successful Strategies:** Cross-cutting successful strategies included:
 - Explicit connection of immersive experiences to learning objectives
 - Progressive complexity with corresponding scaffolding
 - Structured reflection connecting experiences to principles
 - Ongoing facilitator support and communities of practice
 - Clear communication with all stakeholders about rationale and expectations
6. **Differential Effectiveness:** Across all three cases, immersive learning showed consistent patterns of differential effectiveness:
 - Moderate effects on knowledge acquisition across all contexts
 - Strong effects on transfer and application in all contexts
 - Strong motivational effects, particularly related to identity development
 - Most beneficial for learners in the middle achievement/experience range

Synthesis of Findings

Integration of meta-analysis and case study findings revealed several consistent patterns:

1. **Differential effects by outcome type:** Both datasets confirmed stronger effects on transfer and motivation than on knowledge acquisition.
2. **The authenticity-guidance balance:** The optimal balance differed by outcome type, with higher guidance enhancing knowledge and transfer but potentially reducing motivation.
3. **Individual difference effects:** Prior knowledge and self-regulation skills consistently moderated effectiveness, with stronger benefits for certain learner profiles.
4. **Implementation requirements:** All case studies documented substantial resource requirements for high-quality implementation, consistent with moderator analyses showing the importance of implementation duration and facilitator training.
5. **Context-specific adaptation:** Effective designs required significant adaptation to specific educational contexts, learner characteristics, and content domains.

Discussion

Interpretation of Key Findings

This mixed-methods investigation provides nuanced insights into the effectiveness, moderators, and implementation requirements of immersive learning.

Differential Effectiveness by Outcome Type

The stronger effects on transfer and motivation compared to knowledge acquisition align with the theoretical foundations of immersive learning. The authentic contexts and tasks that characterize immersive learning appear particularly effective for helping learners understand when and how to apply knowledge (transfer) and for generating intrinsic interest in the content (motivation).

However, the moderate effects on knowledge acquisition ($g = 0.38$) suggest that immersive learning should not be viewed as a universal replacement for direct instruction. Instead, it appears most valuable when transfer and motivational outcomes are primary goals or when combined with appropriate direct instruction components.

Alternative Explanations and Critical Analysis

Several alternative explanations for these findings warrant consideration:

1. Novelty Effects vs. Inherent Advantages

The stronger motivational effects observed with immersive learning might reflect novelty effects rather than inherent advantages of immersion. However, several lines of evidence argue against this interpretation:

- The meta-analysis included studies with implementation durations ranging from 1 to 16 weeks, and moderator analyses showed no significant decline in motivational effects with longer implementations.
- Case studies with repeated measures over time showed sustained engagement rather than declining interest patterns typical of novelty effects.
- Qualitative data from case studies highlighted specific features of immersive learning (authenticity, agency, meaningful contexts) that supported motivation through established psychological mechanisms.

Nevertheless, the possibility of novelty contributing to motivational effects cannot be entirely dismissed, particularly for short-term implementations or technology-rich approaches that are uncommon in the educational context.

2. Measurement Artifacts

The pattern of stronger effects for transfer and motivation could potentially reflect measurement artifacts rather than true differences in effectiveness. Transfer and motivation are often measured with more authentic, performance-based assessments, while knowledge acquisition typically relies on standardized tests designed for traditional instruction.

The measurement quality analysis provides some support for this concern—transfer was more frequently assessed through authentic performance measures (42%) than knowledge acquisition (6%). However, subgroup analyses based on measurement type showed that the pattern of stronger transfer effects persisted even when comparing studies using similar measurement approaches. Additionally, the case studies, which used parallel measurement approaches across outcome types, showed the same pattern of stronger transfer and motivation effects.

The significant positive relationship between transfer measure reliability and effect size ($\beta = 0.23$, $p = .02$) suggests that measurement quality influences reported outcomes. This finding has important implications for interpreting the strong transfer effects ($g = 0.61$) observed in the meta-analysis. On one hand, the relationship suggests that more reliable transfer measures yield stronger effects, potentially indicating that the true effect on transfer is robust. On the other hand, it raises questions about whether studies with less reliable measures might underestimate transfer effects, contributing to variability in the literature.

3. Selection and Implementation Bias

The positive effects might reflect selection bias in which only successful implementations are published or studied. The publication bias analysis indicated some evidence of this for knowledge outcomes but not for transfer or motivation. Additionally, the case studies documented challenges and implementation variations that would likely be underreported in published literature.

More concerning is the possibility of implementation bias—immersive learning may receive more careful design attention and resources than comparison conditions. The case studies documented substantial resource investments in immersive approaches, raising questions about whether equivalent investment in traditional approaches might yield similar improvements. This possibility cannot be fully addressed with the available data and remains an important consideration for interpreting the findings.

The Authenticity Dilemma

The findings empirically validate the authenticity dilemma proposed by Nachtigall et al. (2024). The meta-analysis demonstrates that authenticity level positively predicted transfer and motivation

but not knowledge acquisition, while guidance level positively predicted knowledge and transfer but negatively predicted motivation.

This tension suggests that designing effective immersive learning requires intentional decisions about the balance between authenticity and guidance based on primary learning goals. The case studies further illustrate how this balance can be achieved through strategic scaffolding, expert facilitation, and structured reflection integrated within authentic contexts.

Figure 5 illustrates how the authenticity dilemma manifests across different outcome domains, highlighting the need for deliberate design decisions based on primary learning objectives.

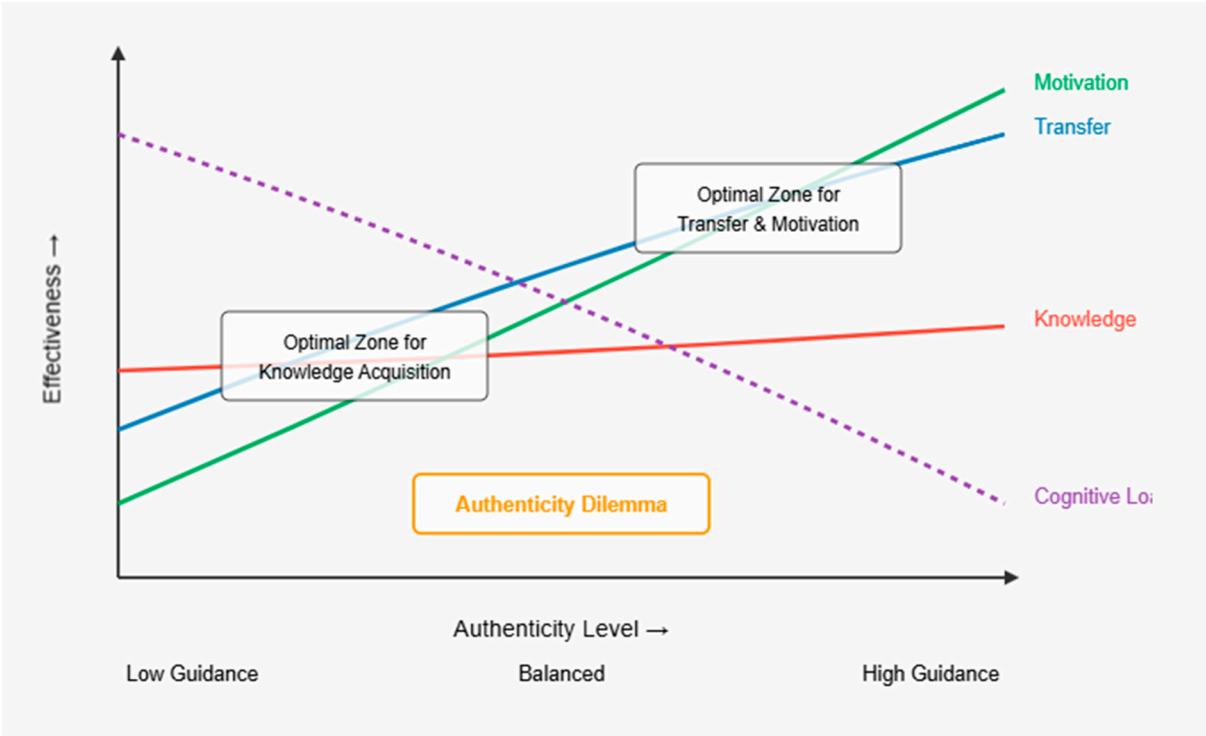


Figure 5. Authenticity Dilemma: Balancing Authenticity and Guidance. Note: Figure illustrates how different levels of authenticity affect learning outcomes, creating a dilemma for instructional designers.

Individual Differences as Boundary Conditions

The significant moderating effects of individual characteristics highlight important boundary conditions for immersive learning effectiveness. The expertise reversal effect (stronger benefits for learners with lower prior knowledge) suggests that immersive approaches may be particularly valuable for introductory or foundational learning experiences.

Similarly, the moderating effect of self-regulation skills suggests the need for differentiated support within immersive environments. The case studies demonstrated successful strategies for such differentiation, including adaptive coaching, optional scaffolds, and peer collaboration structures.

Figure 4 from the middle school case study provides a particularly clear illustration of differential effects across achievement levels, showing that middle-performing students benefited most from the immersive approach. This pattern suggests that immersive learning may help address the “missing middle” phenomenon often observed in educational interventions, where high and low achievers receive targeted support while middle-performing students receive less attention.

Implementation Requirements and Challenges

Both the meta-analysis and case studies highlight substantial resource requirements for effective implementation. The significant moderating effects of implementation duration, facilitator training, and technology integration emphasize that immersive learning quality depends heavily on implementation factors beyond the core design.

This finding has important implications for educational policy and practice, suggesting that successful adoption of immersive learning requires institutional commitment to capacity building, infrastructure development, and sustained implementation support (see Figure 6).

Theoretical Implications

The findings extend existing theoretical frameworks in several important ways:

1. **Refinement of the Authenticity Dilemma:** This study provides empirical validation for this theoretical construct while identifying specific moderators that influence how this dilemma manifests across contexts and outcomes.
2. **Integration with Expertise Reversal Effect:** The interaction between prior knowledge and immersive learning effectiveness connects situated learning theory with cognitive load perspectives, suggesting boundary conditions for immersive approaches.
3. **Expanded Understanding of Transfer:** The strong transfer effects observed across studies support the theoretical claim that authentic contexts enhance transfer, while highlighting the importance of explicit reflection and abstraction processes within immersive experiences.
4. **Theoretical Integration:** The findings support the integrated theoretical model (Figure 1) by demonstrating how multiple theoretical perspectives (situated cognition, cognitive apprenticeship, productive failure, cognitive load) collectively explain the complex patterns of effectiveness observed across different contexts and learner populations.

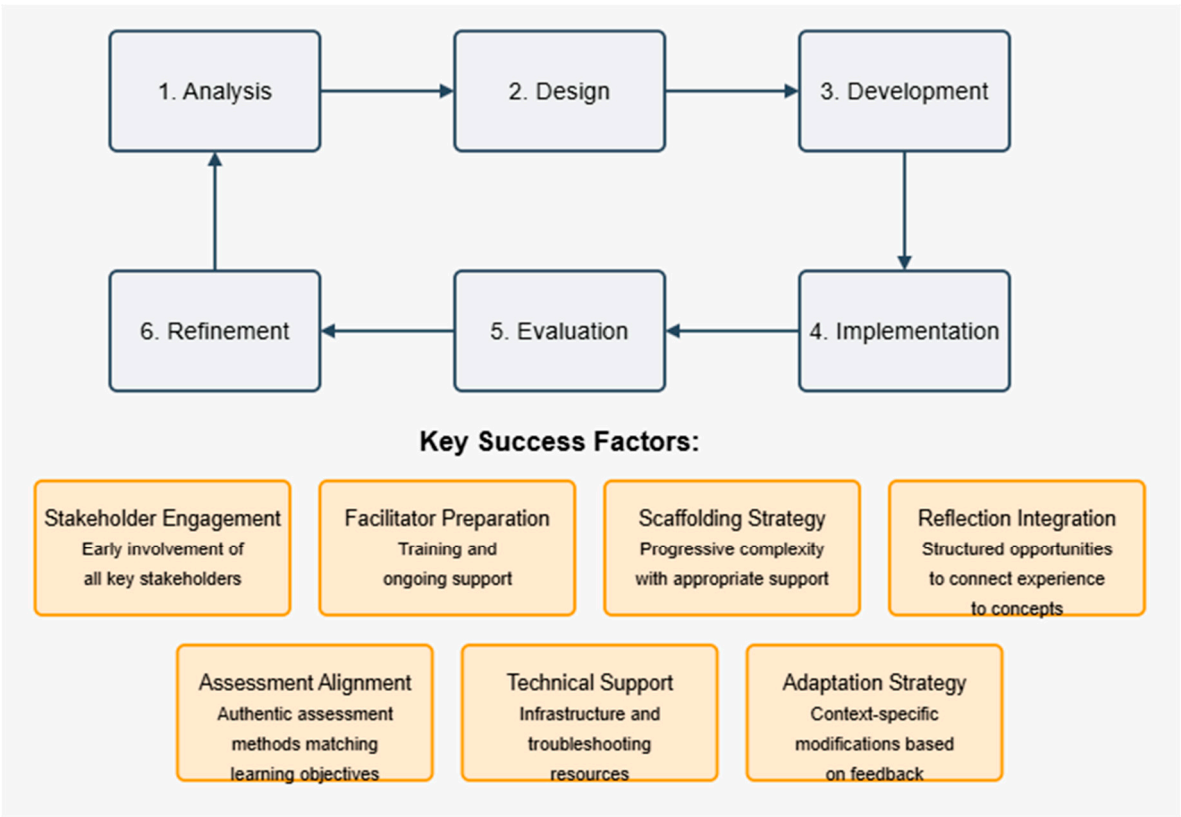


Figure 6. Implementation Framework for Immersive Learning. Note: This framework outlines the cyclical process and key success factors for effective immersive learning.

Practical Implications

The findings yield several practical implications for educational designers, practitioners, policymakers, and institutional leaders:

1. **Alignment with Learning Goals:** Immersive learning should be selected and designed based on primary learning goals, with stronger justification when transfer and motivation are priorities.
2. **Strategic Scaffolding:** Effective immersive learning requires thoughtful scaffolding that supports learning without undermining authenticity, such as just-in-time guidance, embedded tools, and structured reflection opportunities.
3. **Implementation Planning:** Successful adoption requires comprehensive planning that addresses facilitator preparation, technology infrastructure, time allocation, and institutional alignment.
4. **Cost-Benefit Analysis Framework:** Based on the findings, we propose a structured framework for assessing the potential return on investment for immersive learning implementations (Table 7).

Table 7. Cost-Benefit Analysis Framework for Immersive Learning.

Component	Considerations	Estimation Approach
Development Costs		
Design expertise	Instructional design time	Hours × hourly rate
Content development	Materials, scenarios, assessments	Hours × hourly rate + materials
Technology development	Programming, platform costs	Vendor quotes or internal development costs
Pilot testing	Testing and refinement	Hours × hourly rate + participant compensation
Implementation Costs		
Facilitator training	Preparation time, materials	Hours × hourly rate + materials
Facilitation time	Instruction, coaching, feedback	Hours × hourly rate
Technology infrastructure	Hardware, software, maintenance	Purchase/licensing costs + support
Participant time	Time beyond traditional instruction	Hours × hourly rate (opportunity cost)
Expected Benefits		
Knowledge gains	Improvement in core content knowledge	Effect size × standard deviation × value per SD
Transfer improvements	Enhanced application of knowledge	Effect size × standard deviation × value per SD
Motivation enhancements	Increased engagement and persistence	Retention improvements × value per retained learner

Component	Considerations	Estimation Approach
Time-to-proficiency reduction	Faster achievement of performance standards	Days saved × daily productivity value

This framework provides a structured approach to evaluating whether immersive learning investments are likely to yield sufficient returns in specific contexts. The case studies illustrate how this analysis might be applied:

- In the corporate training case, the high development cost (\$420,000) was justified by substantial time-to-proficiency reductions (27 days × daily productivity value) across a large number of participants.
- In the engineering ethics case, moderate development costs (\$48,000) were justified by improvements in ethical reasoning transfer and professional identity development, outcomes highly valued by the engineering program.
- In the middle school case, the lower development costs (\$22,000) and strong effects on historical thinking made the approach cost-effective despite the substantial teacher preparation time required.

For policymakers and institutional leaders, findings suggest several important considerations:

1. **Strategic Investment:** Rather than broad implementation, strategic investment in immersive learning for specific high-priority outcomes (especially transfer and professional identity development) likely offers the highest return on investment.
2. **Capacity Building:** Sustainable implementation requires investment in educator capacity through professional development, communities of practice, and ongoing support structures.
3. **Infrastructure Development:** Technology infrastructure, physical space flexibility, and scheduling accommodations may be necessary to support immersive learning implementation.
4. **Assessment Alignment:** Traditional assessment systems may not capture the unique benefits of immersive learning, suggesting the need for expanded assessment approaches that value transfer and application.

The findings connect to broader educational trends, including competency-based education’s focus on demonstrated application, personalized learning’s attention to individual differences, and advances in educational technology that enable increasingly authentic simulations. These connections suggest that immersive learning will continue to gain prominence as education systems evolve to meet changing societal needs.

Ethical Considerations

This investigation highlights several important ethical considerations in immersive learning design, implementation, and research:

Psychological Safety

Immersive learning often involves emotional engagement and risk-taking that requires careful attention to psychological safety. In the corporate training case, participants sometimes experienced stress when facing challenging simulated leadership scenarios. Effective implementations included

clear framing of the experience as a learning opportunity, debriefing protocols after emotionally charged activities, and options to modify participation if needed.

Privacy and Data Ethics

Technology-enhanced immersive environments often collect extensive data on learner behavior. The engineering ethics case raised concerns about data privacy in the virtual stakeholder interview system, which was addressed through clear consent procedures, anonymized data storage, and transparency about data usage. Educational leaders implementing immersive learning should develop explicit policies regarding data collection, storage, usage, and sharing.

Equity and Access

The middle school case highlighted significant concerns about equitable access to immersive learning opportunities. Students with limited technology access at home were disadvantaged when activities extended beyond class time. Successful implementation required providing alternative access options (extended lab hours, paper alternatives) and careful consideration of technology requirements. More broadly, the substantial resource requirements for high-quality immersive learning raise concerns about potentially exacerbating existing educational inequities if implementation is concentrated in well-resourced settings.

Cultural Sensitivity

Immersive learning often involves role-play, perspective-taking, and engagement with diverse viewpoints. The historical understanding case revealed the importance of cultural sensitivity when engaging students with historically marginalized identities in civil rights history. Effective implementation required teacher preparation for facilitating potentially sensitive discussions, consultation with community members, and thoughtful framing of historical contexts.

Limitations and Future Research

Several limitations of this study suggest directions for future research:

1. **Long-term Outcomes:** Few studies in the meta-analysis measured outcomes beyond immediate post-tests. Future research should examine the durability of immersive learning effects through longitudinal designs.
2. **Implementation Variability:** While implementation features were documented as moderators, more detailed analysis of implementation quality and fidelity would further clarify the conditions for effectiveness.
3. **Domain Specificity:** The case studies covered diverse domains, but systematic investigation of how immersive learning effectiveness varies across subject areas would enhance design principles.
4. **Equity Dimensions:** Both the meta-analysis and case studies identified potential concerns about differential access and effectiveness. Future research should explicitly examine how immersive learning can be designed for equity.
5. **Measurement Challenges:** The field would benefit from development of validated instruments specifically designed to assess the unique outcomes of immersive learning, particularly complex transfer and identity development.
6. **Cost-Effectiveness Comparisons:** More rigorous comparative studies of cost-effectiveness relative to other educational interventions would inform strategic investment decisions.
7. **Integrative Approaches:** Further research should examine how immersive learning can be effectively integrated with direct instruction to maximize benefits across all outcome types.

The moderate knowledge acquisition effects suggest that hybrid approaches might be particularly promising.

Conclusion

This mixed-methods investigation demonstrates that immersive learning, when thoughtfully designed and implemented, offers advantages over traditional instruction for developing transferable knowledge, complex skills, and intrinsic motivation. The effect sizes observed in the meta-analysis ($g = 0.38$ for knowledge, $g = 0.61$ for transfer, $g = 0.67$ for motivation) represent educationally meaningful impacts, while the case studies illustrate practical implementation strategies across diverse contexts.

However, the findings also highlight important nuances often overlooked in discussions of immersive learning. The significant moderating effects of implementation features, individual characteristics, and the authenticity-guidance balance emphasize that effectiveness depends heavily on how, for whom, and under what conditions immersive learning is implemented.

Immersive learning connects to broader trends in educational practice and technology, including competency-based education’s focus on demonstrated application, personalized learning’s attention to individual differences, and advances in educational technology that enable increasingly authentic simulations. These connections suggest that immersive learning will continue to gain prominence as education systems evolve to meet changing societal needs.

As educational institutions and organizations increasingly recognize the limitations of traditional instruction for developing the adaptive expertise needed in today’s complex environments, immersive learning will continue to gain prominence. The challenge lies not in blindly advocating for immersion, but in developing the design expertise, evidence base, and implementation capacity to leverage these approaches effectively.

For researchers, these findings suggest the need for more nuanced investigation of implementation factors, individual differences, and long-term outcomes. For practitioners, the results provide empirical guidance for strategic decision-making about when and how to implement immersive approaches for maximum impact. For policymakers and educational leaders, the findings highlight the importance of strategic investment, capacity building, and assessment alignment to support effective implementation.

By addressing both theoretical and practical dimensions of immersive learning, this research contributes to the ongoing transformation of educational practice to better prepare learners for complex, rapidly changing environments.

Conflicts of Interest and Informed Consent Declarations

I declare that I have no conflicts of interest.
All participants provided written informed consent.

Appendix A: Meta-Analysis Coding Protocol

Study Identification and Screening Form

Citation Information:

- Study ID: [AUTO-GENERATED]
- Authors: _____
- Year: _____
- Title: _____
- Journal/Source: _____
- Volume/Issue/Pages: _____

Inclusion Criteria Checklist:

- Published between January 2010 and December 2024

- Peer-reviewed publication
- Experimental or quasi-experimental design with comparison group
- K-12, higher education, or professional training context
- Measures at least one target outcome (knowledge, transfer, or motivation)
- Sufficient statistical information for effect size calculation
- English language publication

Decision:

- Include
- Exclude
 - Reason for exclusion: _____

*Study Coding Form***Study Characteristics:**

- Sample size (treatment): _____
- Sample size (control): _____
- Total sample size: _____
- Educational context:
 - K-12 Education
 - Higher Education
 - Professional Training
- Subject domain: _____
- Participant age range: _____
- Country: _____
- Study design:
 - True experimental (random assignment)
 - Quasi-experimental (non-random assignment)
- Attrition rate (%): _____
- Study quality rating (1-5): _____ [Based on standardized quality assessment tool]

Intervention Characteristics:

- Type of immersive approach:
 - Game-based learning
 - Simulation-based learning
 - Cognitive apprenticeship
 - Epistemic game
 - Other: _____
- Duration (weeks): _____
- Total instruction time (hours): _____
- Technology use (1-5 scale): _____ [1=minimal to 5=extensive]
- Degree of authenticity (1-5 scale): _____ [1=low to 5=high]
- Level of guidance (1-5 scale): _____ [1=minimal to 5=extensive]
- Implementation setting:
 - Classroom/formal learning environment
 - Laboratory setting
 - Online/distance learning
 - Workplace

- Other: _____
- Facilitator training (hours): _____
- Key design features (open text): _____

Comparison Condition:

- Type of comparison:
 - Traditional instruction
 - Alternative active learning
 - No treatment
 - Other: _____
- Description of comparison condition: _____

Outcome Measures:

- Knowledge acquisition:
 - Measure description: _____
 - Reliability coefficient (if reported): _____
 - Timing of measurement:
 - Immediate
 - Delayed (specify timeframe): _____
 - Effect size data:
 - Treatment group (M, SD, n): _____
 - Control group (M, SD, n): _____
 - Test statistic (t, F, etc.): _____
 - p-value: _____
 - Calculated effect size (Hedge's g): _____
 - Standard error: _____
- Transfer:
 - Measure description: _____
 - Transfer type:
 - Near transfer
 - Far transfer
 - Reliability coefficient (if reported): _____
 - Timing of measurement:
 - Immediate
 - Delayed (specify timeframe): _____
 - Effect size data:
 - Treatment group (M, SD, n): _____
 - Control group (M, SD, n): _____
 - Test statistic (t, F, etc.): _____
 - p-value: _____
 - Calculated effect size (Hedge's g): _____
 - Standard error: _____
- Motivation:
 - Measure description: _____
 - Motivation dimension:
 - Interest

- Self-efficacy
- Value
- Engagement
- Other: _____
- Reliability coefficient (if reported): _____
- Timing of measurement:
 - Immediate
 - Delayed (specify timeframe): _____
- Effect size data:
 - Treatment group (M, SD, n): _____
 - Control group (M, SD, n): _____
 - Test statistic (t, F, etc.): _____
 - p-value: _____
 - Calculated effect size (Hedge's g): _____
 - Standard error: _____

Moderator Variables:

- Learner characteristics:
 - Prior knowledge level:
 - Low
 - Medium
 - High
 - Mixed
 - Not reported
 - Self-regulation skills (if measured): _____
 - Other individual differences measured: _____
- Implementation features:
 - Implementation fidelity (1-5 scale): _____ [1=low to 5=high]
 - Level of teacher/facilitator involvement (1-5 scale): _____ [1=low to 5=high]
 - Technology integration quality (1-5 scale): _____ [1=low to 5=high]

Additional Notes:

- Key findings reported by authors: _____
- Limitations reported by authors: _____
- Coder comments: _____

Appendix B: Case Study 1 Research Instruments

Case Study 1: Corporate Sales Leadership Training

Pre-implementation Assessment Protocol

Organizational Context Analysis:

1. Current sales leadership training approaches
2. Identified performance gaps
3. Organizational readiness for immersive learning
4. Technology infrastructure assessment
5. Budget and resource constraints

Stakeholder Interview Guide:

Leadership Interviews (30-45 minutes)

1. What are the primary challenges your sales leaders face when transitioning into leadership roles?
2. What knowledge, skills, and attitudes are most critical for successful sales leadership?
3. How would you describe the effectiveness of current training approaches?
4. What metrics do you currently use to evaluate sales leadership performance?
5. What expectations do you have for a new immersive training approach?
6. What concerns do you have about implementing an immersive learning approach?
7. What resources (time, budget, personnel) can be allocated to this initiative?

Current Sales Managers Interviews (30-45 minutes)

1. Describe your experience transitioning into your leadership role.
2. What were the most challenging aspects of becoming a sales leader?
3. How did previous training prepare you for your role? What was missing?
4. What scenarios or situations would you want new leaders to practice before taking on the role?
5. How comfortable are you with technology-enhanced learning approaches?
6. What support would you need to participate in an immersive learning experience?

Training Needs Analysis Survey:

Rate the importance of developing competency in each area (1=not important, 5=extremely important):

1. Strategic account planning
2. Sales team coaching and development
3. Performance management
4. Pipeline management
5. Cross-functional collaboration
6. Customer relationship management
7. Business acumen and financial analysis
8. Change management
9. Virtual team leadership
10. Negotiation strategy

Rate your current proficiency in each area (1=novice, 5=expert)

[Same 10 items]

Open-ended questions:

1. What situations do you find most challenging in your sales leadership role?
2. What resources or support would help you become more effective?
3. What would an ideal sales leadership development program include?

*Implementation Measurement Instruments***Leadership Competency Assessment Tool:**

Assessment dimensions (1-5 scale with behavioral anchors):

1. Strategic thinking and planning
2. Coaching and development
3. Performance management
4. Team leadership
5. Business acumen
6. Customer focus

7. Cross-functional collaboration
8. Change management
9. Decision making
10. Communication

Sample item - Coaching and Development:

1 = Provides minimal feedback to team members

3 = Provides regular feedback and some development opportunities

5 = Creates comprehensive development plans, provides actionable feedback, and actively supports growth

Simulation Engagement Tracking:

- Time spent in simulation (total and by module)
- Decision points encountered and choices made
- Resources accessed during simulation
- Help requests and system support utilized
- Completion rate and progression pace

Post-Module Reflection Prompts:

1. What was the most challenging decision you faced in this module?
2. How did you approach this challenge and what informed your decision?
3. What would you do differently if faced with a similar situation in the future?
4. How does this experience connect to your current or future role?
5. What support or resources would help you apply these lessons in your work?

Post-Training Survey:

Experience rating (1=strongly disagree, 5=strongly agree):

1. The simulation scenarios were realistic and relevant to my role
2. I was able to apply my existing knowledge in meaningful ways
3. The feedback I received helped me improve my approach
4. The experience challenged me to think differently about leadership
5. I feel more confident in my ability to handle similar situations
6. The technology enhanced rather than distracted from my learning
7. The coaching support enhanced my learning experience
8. I would recommend this training to colleagues

Open-ended questions:

1. What aspects of the training were most valuable to you?
2. What aspects could be improved?
3. How do you plan to apply what you've learned?
4. What additional support would help you implement these skills?

6-Month Follow-up Performance Tracking:

- Time to productivity metrics (days to achievement of performance standards)
- Team performance indicators (sales results, conversion rates, etc.)
- Manager evaluation of leadership effectiveness (standardized assessment)
- Team member feedback on leadership (360° assessment)
- Self-reported application of training (structured interview)

Qualitative Interview Protocol (Post-Implementation)

Participant Interviews (45-60 minutes):

1. Describe your overall experience with the immersive training program.
2. How did the simulation compare to real-world leadership challenges you've encountered?
3. Which aspects of the simulation most effectively prepared you for your role?
4. How has your approach to leadership changed as a result of this experience?
5. What specific situations have you handled differently based on what you learned?
6. What aspects of the experience have been most valuable in your daily work?
7. What additional scenarios or elements would have enhanced your learning?
8. How did the technology affect your learning experience?
9. How did the coaching component contribute to your development?
10. What recommendations would you make for improving the program?

Supervisor Interviews (30-45 minutes):

1. What changes have you observed in the participant's leadership approach?
2. How would you compare the effectiveness of participants who completed this training versus previous approaches?
3. What skills or competencies seem to have been most influenced by the training?
4. Have you observed any challenges in applying the learning?
5. How has the training affected team performance and dynamics?
6. What organizational factors have supported or hindered application of the learning?
7. What recommendations would you make for improving the program?

Implementation Cost and Resource Tracking Form:

- Development costs (itemized by category)
- Technology infrastructure costs
- Facilitator/coach time allocation and costs
- Participant time allocation and opportunity costs
- Ongoing support and maintenance costs
- Total cost per participant

Appendix C: Case Study 2 Research Instruments

Case Study 2: Engineering Ethics in Higher Education

Quasi-Experimental Design Protocol

Participant Assignment:

- Assignment method: Section-based assignment (two course sections)
- Control for selection bias: Demographic and academic background survey
- Sample size calculation: Power analysis for medium effect ($d=0.5$), power=0.8, $\alpha=0.05$

Experimental Group (Epistemic Game Approach):

- 4-week immersive learning module
- Engineering ethics investigation scenario
- Virtual stakeholder interviews
- Document analysis activities
- Collaborative ethical analysis
- Professional framework application
- Final ethical recommendations

Control Group (Case Study Approach):

- 4-week traditional ethics instruction

- Same learning objectives and content
- Traditional case study analysis
- Lecture-based ethical framework instruction
- Individual written ethical analyses
- Class discussions of ethical dilemmas
- Final ethics position papers

Implementation Fidelity Checklist:

- Core epistemic game elements present
- Appropriate facilitation provided
- Technology functioning properly
- Authentic ethical frameworks applied
- Collaborative processes supported
- Reflection activities completed
- Assessment aligned with learning approach

Assessment Instruments

Engineering Ethical Reasoning Instrument (Pre/Post):

- 25-item assessment measuring:
 - Recognition of ethical issues (5 items)
 - Stakeholder perspective-taking (5 items)
 - Application of ethical frameworks (5 items)
 - Ethical decision-making process (5 items)
 - Justification of ethical positions (5 items)
- Format: Scenario-based questions with constructed responses
- Scoring: Standardized rubric (1-5 scale per dimension)
- Validation: Previously validated with engineering students ($\alpha=.87$)
- Example item: "A senior engineer asks you to remove a safety concern from your report because addressing it would delay an important product launch. Analyze the ethical dimensions of this situation and describe how you would respond."

Work Product Analysis Rubric:

- Quality of ethical issue identification (1-5 scale)
- Depth of stakeholder perspective consideration (1-5 scale)
- Appropriateness of ethical framework application (1-5 scale)
- Logical reasoning and analysis (1-5 scale)
- Professional communication of recommendations (1-5 scale)
- Consideration of consequences (1-5 scale)
- Integration of technical and ethical considerations (1-5 scale)

Student Engagement Survey:

Rate your agreement with each statement (1=strongly disagree, 5=strongly agree):

1. I found myself thinking about the ethical scenarios outside of class time
2. I was mentally engaged during the learning activities
3. The scenarios captured my interest
4. I could see the relevance of these activities to my future career
5. I invested significant effort in the learning activities
6. I enjoyed discussing ethical issues with my peers

7. I felt personally involved in resolving the ethical dilemmas
8. I wanted to perform well on the assignments

Implementation Process Documentation:

- Development time log (hours by activity)
- Technology development and integration costs
- Facilitator training time and activities
- Instructional time allocation (by activity)
- Technical support requirements
- Student time on task (logged in learning management system)

Focus Group Protocol

Student Focus Groups (60-90 minutes, 6-8 students per group):

Opening Questions:

1. Describe your overall experience with the ethics module.
2. What aspects of the learning experience stood out to you?

Key Questions:

3. How did this approach to learning ethics compare to other educational experiences you've had?
4. How did engaging with virtual stakeholders affect your understanding of ethical issues?
5. What was challenging about analyzing ethical situations in this format?
6. How did collaboration with peers influence your ethical reasoning?
7. How did the immersive nature of the experience affect your engagement?
8. In what ways did this experience change how you think about ethical issues in engineering?
9. How do you anticipate applying what you've learned in your future work?

Closing Questions:

10. What would you change about the learning experience?
11. What additional thoughts would you like to share about learning ethics in this way?

Instructor Focus Group (60-90 minutes):

1. How did implementing this approach compare to previous teaching experiences?
2. What changes did you observe in student engagement and learning?
3. What challenges did you encounter in facilitating the immersive approach?
4. How did your role as an instructor change in this approach?
5. What support or resources were most helpful in implementation?
6. What additional support would have improved the implementation?
7. How sustainable is this approach for future teaching?
8. What modifications would you recommend for future iterations?

Appendix D: Case Study 3 Research Instruments

Case Study 3: Middle School Historical Understanding

Historical Thinking Skills Assessment

Pre/Post Assessment of Historical Thinking Skills:

- 20-item assessment measuring:
 - Sourcing (evaluating document origins)
 - Contextualization (placing events in historical context)
 - Corroboration (comparing multiple sources)
 - Historical perspective-taking

- Historical significance assessment
- Format: Document-based questions with constructed responses
- Scoring: Standardized rubric (1-5 scale per dimension)
- Validation: Previously validated with middle school students ($\alpha=.83$)
- Sample item: "Examine these two accounts of the Montgomery Bus Boycott. How do they differ? Why might they present different perspectives? What can we learn from comparing them?"

Student Work Product Rubric:

- Historical accuracy (1-5 scale)
- Use of evidence from primary sources (1-5 scale)
- Consideration of multiple perspectives (1-5 scale)
- Contextualization of historical events (1-5 scale)
- Recognition of cause and effect relationships (1-5 scale)
- Quality of historical narrative construction (1-5 scale)
- Exhibition of historical empathy (1-5 scale)

Self-Efficacy and Interest Measures

Historical Thinking Self-Efficacy Scale:

Rate your confidence in your ability to (1=not at all confident, 5=extremely confident):

1. Analyze primary source documents
2. Identify biases in historical accounts
3. Understand historical events from multiple perspectives
4. Place historical events in their proper context
5. Evaluate the reliability of historical sources
6. Create evidence-based historical explanations
7. Recognize connections between past and present
8. Engage in historical debates using evidence

Historical Interest Inventory:

Rate your agreement with each statement (1=strongly disagree, 5=strongly agree):

1. I enjoy learning about history
2. Understanding history is important to me
3. I am curious about how people lived in the past
4. I find historical debates and controversies interesting
5. I often wonder about the causes of historical events
6. I like to imagine what life was like in different time periods
7. I see connections between historical events and current issues
8. I would choose to learn more about history outside of school

Implementation Documentation

Teacher Time and Activity Log:

- Preparation time (hours by activity)
- Instructional time (hours by activity)
- Assessment time (hours)
- Collaboration time (hours)

- Technology troubleshooting time (hours)
- Total implementation time

Implementation Fidelity Checklist:

- Cognitive apprenticeship elements implemented:
 - Modeling of historical thinking
 - Coaching during investigations
 - Scaffolding provided
 - Articulation of reasoning
 - Reflection on historical thinking
 - Exploration of primary sources
- Historical investigation components:
 - Primary source analysis
 - Virtual interviews with historical figures
 - Evidence-based narrative construction
 - Museum exhibit creation
- Assessment alignment:
 - Authentic tasks
 - Process and product evaluation
 - Self-assessment opportunities

Technology Access and Equity Assessment:

- Student technology access (home and school)
- Accommodations provided for students with limited access
- Technical issues encountered
- Digital literacy support provided
- Alternative access methods utilized

Student Engagement Tracking

Behavioral Engagement Measures:

- Classroom participation rates (teacher observation protocol)
- Assignment completion rates
- Time on task during class activities (structured observation)
- Extension activity participation
- Resource access metrics (digital platform analytics)

Teacher Observation Protocol:

For each 15-minute interval, record:

1. Number of students actively engaged with learning materials
2. Number of voluntary student contributions to discussion
3. Number of student-initiated questions
4. Instances of students building on peers' ideas
5. Examples of historical thinking observed

Student Reflection Journal Prompts:

1. What did you find most interesting about today's historical investigation?
2. What questions do you still have about this historical topic?
3. How did examining primary sources change your understanding of events?
4. What was challenging about today's historical thinking activities?

- 5. How has your approach to analyzing historical sources changed?
- 6. How does this historical topic connect to other things you’ve learned or experienced?

Teacher Interview Protocol

Mid-Implementation Interviews (30-45 minutes):

- 1. How are students responding to the cognitive apprenticeship approach?
- 2. What aspects of the implementation are working well?
- 3. What challenges have you encountered in implementation?
- 4. How are you adapting the approach to meet student needs?
- 5. What differences are you observing compared to previous instructional approaches?
- 6. What support would help improve implementation?

Post-Implementation Interviews (45-60 minutes):

- 1. How would you describe the overall impact of this approach on student learning?
- 2. What changes have you observed in students’ historical thinking skills?
- 3. How did student engagement compare to previous instructional approaches?
- 4. Which components of the cognitive apprenticeship approach were most effective?
- 5. How did the implementation affect different groups of students?
- 6. What modifications would you make in future implementations?
- 7. How sustainable is this approach within your current teaching context?
- 8. What institutional factors supported or hindered implementation?
- 9. How has this experience influenced your approach to teaching historical thinking?
- 10. What advice would you give to other teachers implementing this approach?

Appendix E: Cross-Case Analysis Protocol

Cross-Case Analysis Protocol

Implementation Factors Comparative Analysis Framework

Design Features Matrix:

- Authenticity components implemented
- Guidance/scaffolding approaches
- Technology integration level
- Content/context alignment
- Assessment approach

Implementation Process Comparison:

- Development time and resources
- Facilitator preparation
- Learner orientation
- Implementation challenges
- Adaptation strategies

Resource Requirements Comparison:

- Development costs (standardized categories)
- Technology infrastructure requirements
- Personnel time allocation
- Sustainability factors
- Cost-effectiveness metrics

Outcome Patterns Analysis Framework

Effect Size Standardization:

- Conversion of all outcome measures to comparable effect sizes
- Categorization by outcome type (knowledge, transfer, motivation)
- Disaggregation by learner characteristics

Qualitative Outcome Pattern Analysis:

- Cross-case thematic analysis of qualitative data
- Identification of common success factors
- Documentation of implementation barriers
- Context-specific moderating factors

Moderator Analysis Template:

- Individual differences effects
- Implementation quality effects
- Contextual factors effects
- Design feature effects

*Case Synthesis Protocol***Synthesis Workshop Agenda:**

1. Individual case presentation (key findings)
2. Cross-case pattern identification
3. Rival explanation analysis
4. Contextual factor discussion
5. Implementation principles development
6. Research limitations assessment
7. Practical implications formulation
8. Future research direction identification

Integration Framework:

- Triangulation of quantitative and qualitative findings
- Explanatory mechanisms identification
- Boundary condition definition
- Implementation guidance development
- Research-to-practice translation

References

1. Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
2. Cheng, K. H., Yen, S. H., & Chen, J. S. (2022). Augmented reality in K-12 education: A meta-analysis of student learning effectiveness and a systematic review of implementation challenges. *Computers & Education*, 183, 104494.
3. Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, 90(4), 499-541.
4. Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79-122.
5. Clark, R. E. (2007). Learning from serious games? Arguments, evidence, and research suggestions. *Educational Technology*, 47(3), 56-59.

6. Collins, A., & Kapur, M. (2022). Cognitive apprenticeship. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 156-174). Cambridge University Press.
7. Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66-69.
8. Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.
9. Gee, J. P. (2003). What video games have to teach us about learning and literacy. *ACM Computers in Entertainment*, 1(1), 1-4.
10. Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111-127.
11. Ifenthaler, D., & Kim, Y. J. (Eds.). (2019). *Game-based assessment revisited*. Springer International Publishing.
12. Joksimović, S., Dowell, N., Gašević, D., Mirriahi, N., Dawson, S., & Graesser, A. C. (2023). Comprehensive analysis of automated feedback and learning analytics to enhance problem-based learning. *Computers & Education*, 191, 104644.
13. Kallio, J. M., & Halverson, R. (2020). Distributed leadership for personalized learning. *Journal of Research on Technology in Education*, 52(3), 371-390.
14. Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31.
15. Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45-83.
16. Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
17. Kuhn, S., & Weissman, S. (2023). The fidelity paradox in simulation-based medical education: A systematic review and framework. *Medical Education*, 57(6), 577-592.
18. 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.
19. Nachtigall, V., Shaffer, D. W., & Rummel, N. (2024). The authenticity dilemma: Towards a theory on the conditions and effects of authentic learning. *European Journal of Psychology of Education*, 39(4), 3483-3509.
20. 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.
21. Reich, J., Buttimer, C. J., Fang, A., Hillaire, G., Hirsch, K., Larke, L. R., Littenberg-Tobias, J., Moussapour, R. M., Napier, A., Thompson, M., & Slama, R. (2019). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences*, 116(29), 14900-14905.
22. Schank, R. C. (1994). Goal-based scenarios: A radical look at education. *Journal of the Learning Sciences*, 3(4), 429-453.
23. Shaffer, D. W. (2006). *How computer games help children learn*. Palgrave Macmillan.
24. Shaffer, D. W., & Resnick, M. (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, 10(2), 195-216.
25. Stoddard, J., Brohinsky, J., Behnke, D., & Shaffer, D. (2022). *Designing epistemic games for informed civic learning*. Harvard Education Press.
26. Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261-292.

27. Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249-265.
28. Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.