SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

When Immersive Learning Works: AI- and VR-Supported Feedback for Competency-Sensitive Skill Development in Higher Education

María del Pilar Gómez Carreño and Roberto Gómez Tobías
Tecnológico de Monterrey, Mexico

Article DOI: 10.68286/samyrad2026.365

ABSTRACT

Artificial intelligence (AI) and virtual reality (VR) may support different learning outcomes unevenly. This study examines competency-technology fit in an undergraduate business course through a randomized classroom study with individual allocation. A digital wheel assigned 34 students to control and experimental conditions (17 each). Both conditions received lectures, case activities, project preparation, and faculty guidance. The experimental condition added three 15-minute sessions using VirtualSpeech and Meta Quest 3 headsets for familiarization, rehearsal, and a final presentation. The course instructor assessed organizational behavior, human capital processes, and social intelligence individually using a competency-based rubric; satisfaction, engagement, and perceived learning were student-reported. Post-intervention ANOVA favored the experimental condition for organizational behavior, $F(1, 32) = 12.71$, $p = .0012$, and nominally for social intelligence, $F(1, 32) = 4.57$, $p = .0402$, but not human capital processes, $p = 1.0000$. Only organizational behavior remained significant after Holm correction within the competency family (adjusted $p = .0035$). No student-perception outcome survived correction within its family. Both groups improved on the rubric-based competency scores. The small sample and exploratory comparison across competencies warrant caution about effect magnitude and mechanism. The results motivate further tests of whether immersive feedback is especially useful for competencies requiring situated, observable performance; they do not establish competency-technology fit as a causal mechanism.

KEYWORDS Artificial intelligence, virtual reality, immersive learning, competency-based education, organizational behavior, social intelligence, formative feedback.

How to cite: Gómez Carreño, M. del P., & Gómez Tobías, R. (2026). When immersive learning works: AI- and VR-supported feedback for competency-sensitive skill development in higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.365

Introduction

Artificial intelligence (AI) and virtual reality (VR) have become prominent tools in higher education because they can make learning more interactive, adaptive, and performance oriented. AI can provide rapid feedback, personalized prompts, and analytic support, while VR can place learners in simulated environments where they can rehearse professional behaviors without the cost or risk of live organizational settings [1]-[5]. For business education,

this combination is especially promising because many managerial competencies are not merely declarative. They require interpretation, communication, social perception, judgment, and behavioral performance under conditions that resemble professional practice.

At the same time, the educational value of immersive technologies should not be assumed. The relevant question is not only whether AI and VR improve learning, but which competencies are likely to benefit from them and why. Competency-based education requires alignment among the competency being developed, the learning task, the feedback mechanism, the assessment evidence, and the form of faculty mediation [6], [7]. A technology-rich activity can therefore be pedagogically weak when the technology does not match the structure of the target competency, even if students find the experience novel or engaging.

Recent empirical findings make this distinction relevant to business education. Hubbard [8] compared written, video, and VR business cases: VR improved visual recall but reduced numerical recall, indicating outcome-specific trade-offs. In presentation training, Sichterman et al. [9] compared AI-only feedback with teacher-supported AI feedback after immersive practice and detected no significant between-condition differences. That finding does not establish equivalence, but it cautions against assuming that an additional instructional component necessarily improves every outcome. Together, these studies motivate closer attention to the learning target and the feedback arrangement.

The three course competencies offer a theoretically relevant comparison within a common instructional setting. Organizational behavior requires interpreting how a proposal affects workplace behavior and organizational wellbeing. Social intelligence is expressed through persuasive, audience-sensitive communication and the ability to elicit commitment. Human capital processes emphasizes the justification and integration of talent-management procedures. These are differences in the dominant demands of the assessed tasks, rather than claims that any competency is purely behavioral or purely conceptual. Immersive presentation practice may make the first two demands more directly rehearsable, while cases and guided analysis may already support much of the third.

The remaining question is how the affordances of immersive feedback align with different competencies assessed within the same business course. Competency-technology fit adds an outcome-specific proposition to experiential learning, feedback, and technology-integration accounts: the same instructional technology may offer different incremental benefits depending on what students must demonstrate and what its feedback makes observable. The present randomized classroom study examines that proposition through three faculty-rated competencies and a separate set of student-perception outcomes.

Three research questions guide the analysis. RQ1 asks whether the experimental condition produced stronger post-intervention competency outcomes than the control condition. RQ2 asks whether those differences varied across competencies in a pattern consistent with competency-technology fit. RQ3 asks whether student perceptions followed the same pattern as competency outcomes or revealed a more nuanced relationship among satisfaction, engagement, and perceived learning.

The theoretical proposition is that immersive rehearsal and formative feedback may be more useful when learners can practice, observe, and revise the behavior targeted by assessment. Structured conceptual or procedural demands may receive less additional support from presentation rehearsal. We use this proposition to interpret an exploratory pattern across outcomes, not as a confirmatory test of an independently prespecified hierarchy of competency fit.

Conceptual background and competency-technology fit model

Competency-sensitive immersive learning

Immersive learning is most defensible when the target outcome requires action in context. Constructivist and experiential learning perspectives emphasize that competence develops through active experience, feedback, reflection, and repeated performance [10], [11]. VR can support this process by creating controlled practice environments, while AI can accelerate formative feedback by identifying observable features of performance and offering suggestions for improvement. This combination is especially relevant when students must rehearse communication, leadership, negotiation, presentation, or socially situated judgment.

A competency-sensitive approach requires distinguishing technology affordances from competency requirements. VR offers presence, simulation, and repeated rehearsal; AI offers feedback, personalization, and rapid analytic response. These affordances are strongest when the competency can be practiced through observable behavior and when feedback can be linked to performance indicators. They may be less decisive for competencies that depend primarily on conceptual classification, procedural knowledge, or content mastery already supported by cases, lectures, worked examples, and guided analysis.

Feedback, faculty mediation, and assessment alignment

Feedback is one of the strongest mechanisms through which learning environments shape performance, but feedback is most effective when it is timely, specific, understandable, and connected to criteria that students can act on [12]. In immersive AI-supported environments, automated feedback can increase the frequency of practice-feedback cycles. Yet automated feedback should not be treated as a complete judgment of competence. Faculty mediation remains necessary to interpret feedback, connect it to disciplinary standards, challenge superficial responses, and ensure that platform-generated indicators are translated into meaningful learning decisions.

The Technological Pedagogical Content Knowledge framework situates technology within pedagogy and disciplinary content [13]. Here, immersive practice was embedded in project preparation and faculty guidance. Automated feedback introduced an additional source of performance information, while the course instructor remained responsible for interpreting disciplinary expectations and assessing competencies. This arrangement is part of the instructional condition being evaluated.

Competency-technology fit as a design proposition

Competency-technology fit is defined here through four related dimensions: behavioral observability, or whether the target performance can be seen or heard; interpersonal situatedness, or dependence on an audience or social context; rehearsability, or the opportunity to practice the target action; and feedback responsiveness, or whether feedback can guide a concrete revision. These dimensions connect competency demands to the specific affordances of immersive presentation practice. They extend general accounts of learning through experience and feedback by identifying why the same activity might support some assessed outcomes more directly than others.

Table 1 applies these dimensions to the three competencies and juxtaposes the proposed fit with the observed outcome pattern. The fit classifications function here as theory-informed interpretive judgments; the present analysis does not establish prospective confirmation of a competency hierarchy. Accordingly, the table provides an exploratory explanation to be tested prospectively. The study does not manipulate competency fit or estimate a formal treatment-by-competency interaction.

Competency fit and assessment evidence

Competency-technology fit also has an assessment implication. A technology may create an engaging experience yet still fail to produce evidence that corresponds to the intended competency. For outcomes centered on social interaction or behavioral performance, assessment should capture what students can enact, adjust, and justify in context. For more structured procedural outcomes, written analyses, process accuracy, or concept application may provide a closer representation of competence. The assessment method therefore participates in the fit relationship rather than serving as a neutral measurement step.

This perspective is consistent with feedback research emphasizing that learners benefit when performance criteria are visible and actionable [12]. In the present study, the same logic connects rehearsal to competency evidence: immersive practice is educationally meaningful only when feedback points students toward the behaviors and interpretations that the course intends to develop. This reinforces the human role in the model because faculty must decide which indicators are instructionally relevant and which platform outputs are merely proxies for more complex professional performance.

Table 1. Competency-affordance alignment and interpretive fit

Competency	Learning demand	Proposed fit	Observed evidence
Organizational behavior	Situated behavioral interpretation	High	Robust; p=.0012
Human capital processes	Structured procedural/conceptual	Mid-low	No difference; p=1.000
Social intelligence	Social interpretation and interaction	High	Nominal only; p=.0402

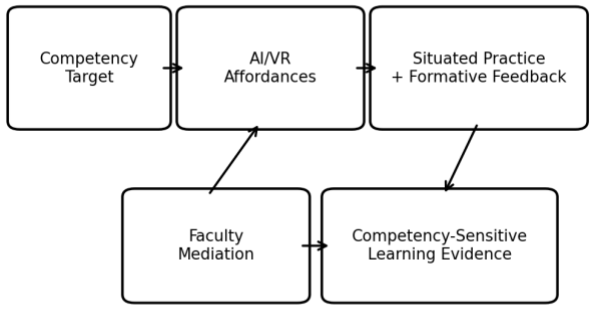


Fig. 1. Proposed competency-sensitive alignment model for AI- and VR-supported formative feedback.

Fig. 1 summarizes the proposed relationship. A target competency defines the learning demand; AI and VR contribute simulation, rehearsal, and feedback; situated practice creates learning opportunities; and faculty

mediation preserves disciplinary meaning and assessment responsibility. The proposed selective benefit of this alignment remains a question for prospective testing.

Research design and methods

Research design, context, and participants

The study used a randomized classroom design with individual allocation [14] in an undergraduate course on organizational behavior and human capital processes at a private university in Mexico. The analytic sample comprised 34 students, with 17 in each condition. Students had previously formed project teams through Canvas, but these teams were not the allocation unit. A freely available online digital wheel randomly assigned individual students to the experimental or control condition. Each student presented their project individually, and the course instructor assessed competencies at the individual level. Instruction and the evaluated presentations were conducted in Spanish.

The course lasted five weeks, with four scheduled classroom hours per week (20 hours in total). The sample reflects the available classroom cohort, rather than a random sample of the wider student population. Pre-intervention ratings preceded the intervention; post-intervention competency assessment and student-perception measurement followed the instructional sequence. The analysis therefore concerns immediate course-related outcomes in a single classroom context.

Instructional conditions and faculty mediation

Both conditions shared lectures, case activities, project preparation, faculty guidance, and pre/post measurement. The experimental condition used VirtualSpeech with Meta Quest 3 headsets during three 15-minute sessions embedded in regular classes. The first session introduced the headset and interface; the second involved presentation rehearsal with virtual avatars; and the third comprised the final presentation. Thus, the sequence covered 45 scheduled minutes across familiarization, rehearsal, and final presentation. During the corresponding periods, control students engaged in face-to-face preparation and presentation rehearsal. Table 2 summarizes the conditions.

The virtual avatar provided formative feedback, followed by an automated performance evaluation. The results interface displayed content, eye contact, filler words, speaking pace, body language, and listenability indicators, together with an overall score and presentation duration. For example, speaking pace was expressed in words per minute, filler words as a count, and eye contact as a percentage. These indicators provided feedback on presentation performance. They were not the competency outcome scores analyzed in the study; those scores were assigned by the course instructor using the rubric.

Both groups worked within the same scheduled course hours and corresponding preparation periods. The conditions differed in the medium and source of feedback: immersive avatar-supported feedback was added in the experimental condition, while faculty guidance remained part of both. Equal scheduled time does not establish an identical amount of individual rehearsal or feedback. The comparison therefore evaluates the integrated instructional condition, without separating the contributions of immersion, automated feedback, and faculty mediation.

Table 2. Instructional conditions and evidence sources

Component	Control	Experimental	Evidence
Lectures/case activities	Yes	Yes	Course activities
Face-to-face project preparation	Yes	Yes	Final project
AI-enabled immersive rehearsal	No	Yes	Platform practice
Automated formative feedback	No	Yes	Platform feedback
Simulated professional scenarios	No	Yes	Immersive activity
Faculty mediation	Yes	Yes	Guided instruction
Pre/post competency ratings	Yes	Yes	Instructor rubric

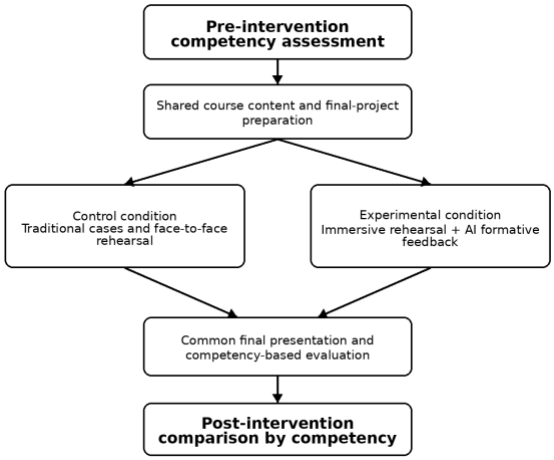


Fig. 2. Classroom sequence with individual random allocation and condition-specific learning activities.

Measures and competency operationalization

Competency outcomes were individual performance evaluations made by the course instructor responsible for the challenge, using the competency-based rubric. Its five performance levels were Destacado (outstanding), Avanzado (advanced), Sólido (solid), Básico (basic), and

Incipiente (emerging), from highest to lowest. The analytic dataset contained one instructor-assigned score for each competency at each measurement occasion on a 1–5 scale, with higher scores indicating stronger performance. The three competencies were analyzed separately; no overall competency index was constructed.

The rubric described two criteria for each competency. Organizational behavior concerned the clarity of the proposed improvement and justification of its contribution to organizational wellbeing and performance, including relevant evidence or indicators. Human capital processes concerned a well-grounded talent-management proposal and integration of prior learning and current trends. Social intelligence was operationalized through persuasive communication adapted to the managerial audience and a conclusion that encouraged commitment and collaborative action. These criteria assess the application of the course competencies within an individual project presentation; they do not constitute an exhaustive measure of each broader construct.

Satisfaction, engagement, and perceived learning were collected through student surveys, separately from faculty assessment. Each was represented by its corresponding course-specific item on a five-point agreement scale (1 = strongly disagree; 5 = strongly agree). For example, the pre-intervention engagement item asked whether students felt motivated and committed to course activities and materials; the perceived-learning item asked whether their learning in the course had improved. The post-intervention items concerned the instructional experience. These indicators were analyzed individually, without creating a multi-item perception scale.

The course instructor applied the rubric as the sole assessor; inter-rater agreement is therefore not applicable to this assessment design. Internal-consistency coefficients are likewise not applicable to individual single-item perception indicators. The recorded rubric-based ratings constitute the competency outcomes analyzed in Tables 3, 4, and 7. Results are reported at group level without student identifiers.

Data analysis and interpretive strategy

Paired t tests examined within-condition pre/post competency changes, and one-way ANOVA compared post-intervention outcomes between groups (df1 = 1, df2 = 32). Means, sample standard deviations, and pointwise 95% t-based confidence intervals (CIs) are reported for both groups at each occasion. Post-intervention mean-difference CIs use the pooled-variance model underlying the ANOVA. Holm adjustment was applied separately to the three competency comparisons and the three perception comparisons, using unrounded p values; the six paired competency tests also received a separate Holm sensitivity check. Within-group effect sizes use d_{av}: the pre-minus-post mean difference divided by the square root of the average of the pre and post variances [15]. Negative d_{av} values therefore indicate improvement. Partial eta squared (ηp^2) summarizes each post-intervention ANOVA.

The five-point ratings were treated as numerical scores. Confidence intervals are unadjusted for multiplicity and describe sampling precision under these models. Holm correction limits familywise false-positive risk within each specified family, but does not remove the imprecision of a small sample or the possibility that statistically selected effects overstate their population magnitude. Nominal findings that do not survive correction are treated as exploratory. Post-intervention ANOVA does not adjust for baseline differences; this constraint informs interpretation.

Results

Competency outcomes

Table 3 reports competency scores before and after the intervention. At baseline, experimental-group means exceeded control means by 0.12 points for organizational behavior and human capital processes and by 0.24 points for social intelligence. These differences provide context for the post-intervention comparisons.

Table 3. Descriptive statistics for competency ratings

Outcome / group	Pre M (SD) [95% CI]	Post M (SD) [95% CI]
Organizational behavior Control	3.24 (0.66) [2.89, 3.58]	4.18 (0.39) [3.97, 4.38]
Organizational behavior Experimental	3.35 (0.70) [2.99, 3.71]	4.71 (0.47) [4.46, 4.95]
Human capital processes Control	3.29 (0.69) [2.94, 3.65]	4.47 (0.51) [4.21, 4.74]
Human capital processes Experimental	3.41 (0.51) [3.15, 3.67]	4.47 (0.51) [4.21, 4.74]
Social intelligence Control	3.24 (0.66) [2.89, 3.58]	4.35 (0.49) [4.10, 4.61]
Social intelligence Experimental	3.47 (0.72) [3.10, 3.84]	4.71 (0.47) [4.46, 4.95]

Note. n = 17 per group at each occasion. M = mean; SD = sample standard deviation. Brackets show unadjusted 95% t-based confidence intervals for the mean. Scores range from 1 to 5.

Table 4 reports the post-intervention competency ANOVAs. Organizational behavior favored the experimental condition, $F(1, 32) = 12.71$, nominal $p = .0012$, $\eta p^2 = .284$, and remained significant after Holm correction (adjusted $p = .0035$). Its post-intervention mean difference was 0.53 points (95% CI [0.23, 0.83]). Social intelligence showed a nominal difference, $F(1, 32) = 4.57$, $p = .0402$, $\eta p^2 = .125$, but did not survive correction (adjusted $p = .0805$). Human capital processes showed no detected difference,

$F(1, 32) = 0.00$, $p = 1.0000$, $\eta p^2 = .000$. Only organizational behavior therefore provides a multiplicity-robust between-condition result in this sample.

Table 4. Anova results for post-intervention competency outcomes.

Outcome	F	p (nominal)	Interpretation
Organizational behavior	12.71	.0012	Holm-robust; ηp^2=.284
Human capital processes	0.00	1.0000	No difference; ηp^2=.000
Social intelligence	4.57	.0402	Suggestive; ηp^2=.125

Student perceptions

Table 5 presents the pre/post student-perception descriptives. These ratings describe the students' instructional experience and are distinct from the instructor-assessed competency outcomes.

Table 5. Descriptive statistics for student perceptions.

Outcome / group	Pre M (SD) [95% CI]	Post M (SD) [95% CI]
Satisfaction Control	3.41 (0.51) [3.15, 3.67]	4.29 (0.47) [4.05, 4.54]
Satisfaction Experimental	3.41 (0.51) [3.15, 3.67]	4.53 (0.51) [4.26, 4.79]
Engagement Control	3.35 (0.79) [2.95, 3.76]	4.29 (0.59) [3.99, 4.60]
Engagement Experimental	3.47 (0.62) [3.15, 3.79]	4.65 (0.49) [4.39, 4.90]
Perceived learning Control	3.59 (0.62) [3.27, 3.91]	4.47 (0.51) [4.21, 4.74]
Perceived learning Experimental	3.59 (0.62) [3.27, 3.91]	4.82 (0.39) [4.62, 5.03]

Note. n = 17 per group at each occasion. M = mean; SD = sample standard deviation. Brackets show unadjusted 95% t-based confidence intervals for the mean. Scores range from 1 to 5. Model-based intervals are not truncated to the response range.

Table 6 summarizes post-intervention perception ANOVAs. Perceived learning nominally favored the experimental condition, $F(1, 32) = 5.05$, $p = .0316$, $\eta p^2 = .136$. Neither engagement, $F(1, 32) = 3.60$, $p = .0668$, $\eta p^2 = .101$, nor satisfaction, $F(1, 32) = 1.94$, $p = .1733$, $\eta p^2 = .057$, met the nominal .05 threshold. Holm-adjusted p values were .0948 for perceived learning, .1337 for engagement, and .1733 for satisfaction. No perception outcome survived correction; the perceived-learning result is exploratory.

Table 6. Anova results for post-intervention perceptions

Perception	F	p (nominal)	Interpretation
Satisfaction	1.94	.1733	No difference; ηp^2=.057
Engagement	3.60	.0668	Not significant; ηp^2=.101
Perceived learning	5.05	.0316	Suggestive; ηp^2=.136

Pre-post gains within groups

Paired t tests showed increased competency ratings in both conditions. Table 7 reports the t statistics and d_{av} values; their negative signs reflect pre-minus-post coding. All six paired comparisons remained below .001 after Holm correction across these tests. The absolute d_{av} values ranged from 1.72 to 2.27. These large standardized changes reflect both the observed score increases and the relatively small dispersion of five-point ratings. They describe within-condition change and do not estimate the incremental treatment effect.

Both instructional conditions were accompanied by improved rubric-based performance. The between-condition evidence is more selective: organizational behavior favored the experimental condition after multiplicity adjustment, social intelligence provided a nominal signal, and human capital processes showed no detected difference. Improvement within each group should not be interpreted as proof of superiority of either instructional approach.

Table 7. Pre-post competency changes within groups

Group	Competency	t	p	d_{av}
Control	Organizational behavior	-5.19	<.001	-1.72
Control	Human capital processes	-7.63	<.001	-1.94
Control	Social intelligence	-5.37	<.001	-1.91
Experimental	Organizational behavior	-7.95	<.001	-2.27
Experimental	Human capital processes	-5.28	<.001	-2.07
Experimental	Social intelligence	-9.06	<.001	-2.04

Note. df = 16 for each paired test. d_{av} uses the pre-minus-post difference standardized by the square root of the average pre/post variances; negative values indicate higher post scores.

Integrative competency-fit pattern

The observed pattern is consistent with a competency-sensitive interpretation, but remains preliminary. Organizational behavior produced the strongest between-condition evidence. Social intelligence showed a smaller nominal effect, while human capital processes showed no detected advantage. Because the same course and instructor provided the setting for all three outcomes, this contrast motivates further investigation of how presentation rehearsal supports different task demands.

A significant result for one competency and a nonsignificant result for another do not by themselves establish that the treatment effects differ. Competency type was not manipulated, and no formal treatment-by-competency interaction was estimated. The pattern

therefore motivates a prospective test of the proposed fit dimensions rather than establishing competency-technology fit as a causal mechanism.

Research-question synthesis

RQ1 receives selective support: only organizational behavior showed a between-condition advantage robust to Holm correction. RQ2 yields an exploratory pattern that is compatible with the proposed fit interpretation, without demonstrating effect heterogeneity or mechanism. RQ3 indicates that no student-perception outcome survived multiplicity correction. The findings support further testing of a conditional proposition rather than a general superiority claim for immersive technology.

Discussion

From technology effect to conditional effectiveness

The findings motivate a competency-sensitive interpretation of AI- and VR-supported formative feedback, with the strongest evidence for organizational behavior. The assessed task required students to explain and justify behavioral improvements in an organizational context. Social intelligence showed a nominal effect in the same direction, but it did not survive Holm correction. Human capital processes showed no detected difference; this is compatible with, but does not establish, a smaller incremental benefit for more structured learning demands.

Competency-technology fit directs attention to the relation among affordances, task demands, feedback, and assessment. In this study, presentation practice made some communication and behavioral features observable, while the instructor assessed their disciplinary relevance. A plausible interpretation is that this alignment supported the organizational-behavior task. Establishing that interpretation as a mechanism would require process measures and direct tests of fit, beyond the present outcome comparisons.

The null result for human capital processes is therefore not a residual finding to be minimized. Together with the strong organizational-behavior effect and the weaker social-intelligence signal, it creates the contrast that makes the study theoretically informative. The evidence does not justify claiming that competency-technology fit has been causally demonstrated; competency type was not manipulated and the sample is small. It does, however, support a more precise testable proposition than a generic technology-effect account: the incremental educational value of AI and VR may depend on what learners are being asked to become competent at doing.

Human mediation as part of the instructional architecture

Faculty mediation remained part of both conditions. The platform introduced avatar-delivered feedback and automated performance indicators, while the instructor retained responsibility for course goals, disciplinary expectations, and competency assessment. This division of roles is consistent with TPACK's emphasis on integrating technology with pedagogy and content [13]. Because the study did not vary teacher support independently, it cannot estimate the separate contribution of faculty mediation.

This human mediation is especially important when AI-generated feedback concerns socially situated performance. Automated indicators can identify observable patterns, but they do not exhaust the meaning of organizational behavior or social intelligence. Faculty judgment is required to connect performance signals to context, professional norms, ethical considerations, and the learning objectives of the course. From this perspective, the relevant innovation is a human-AI instructional architecture in which automated feedback increases practice opportunities without transferring educational accountability to the system.

Perceived learning, satisfaction, and productive effort

The perception results add an important qualification. Perceived learning favored the experimental condition at the nominal level, but the effect did not remain statistically significant after Holm correction; satisfaction and engagement did not meet the nominal significance threshold. These results caution against treating satisfaction or self-reported engagement as proxies for educational effectiveness. Immersive learning can introduce adaptation demands, unfamiliar interaction patterns, and additional effort even when students perceive value in the experience. Prior work similarly emphasizes that emotion, cognitive response, and engagement in VR are intertwined with instructional design rather than automatic consequences of immersion [16].

For instructors, the implication is not to maximize friction or difficulty, but to distinguish productive effort from technical burden. Practice should require students to use feedback, revise performance, and reflect on evidence. Technical complexity that does not serve these processes should be minimized. This distinction is particularly relevant as institutions adopt AI-enabled systems: ease of use matters, but the educational objective is not frictionless completion; it is meaningful rehearsal and improvement.

Hybrid rather than replacement logic

The findings suggest a provisional hybrid design in which conventional instruction supports conceptual framing and guided analysis, while immersive feedback supports rehearsal of observable performance. This is a pedagogical implication to be tested, rather than an empirically established optimal sequence. The study evaluates one integrated classroom implementation and does not compare alternative sequencing or replacement strategies.

Selective adoption may be more defensible than placing every outcome inside an immersive environment. A competency map can identify tasks for which presentation practice makes relevant behavior visible and feedback actionable. Communication, social interpretation, and leadership-related performance are plausible candidates; more structured tasks may require different supports. These are design proposals derived from the framework, not additional outcomes demonstrated by the present sample.

Toward a research agenda for competency-sensitive immersive learning

The competency-sensitive perspective suggests a research agenda that goes beyond comparing one technology-enhanced group with one conventional group. Future studies can classify competencies according to theoretically relevant attributes such as observability, social situatedness, ambiguity, feedback responsiveness, and dependence on repeated rehearsal. Those attributes can then be modeled as moderators of the incremental benefit produced by immersive AI-supported practice. Such designs would permit the field to test whether competency-technology fit explains variation across outcomes more effectively than broad labels such as 'soft skills' or 'technology-enhanced learning.'

A second research direction concerns the decomposition of the instructional architecture. The present study evaluates an integrated condition in which immersion, AI feedback, and faculty mediation operate together. Subsequent research could compare immersive rehearsal with and without AI feedback, AI feedback with and without immersion, or different levels of faculty scaffolding. This would help distinguish the affordance of the medium from the instructional value of feedback and human mediation, while retaining the ecological validity of authentic classroom implementation.

Limitations and future research

The sample was small (n = 34; 17 per condition), drawn from a single course, and assessed immediately after the intervention. Effect estimates are therefore imprecise and may differ substantially in larger samples. Large standardized effects should be interpreted alongside the raw-score differences and confidence intervals, particularly because the outcomes were five-point ratings closely aligned with the course tasks. Small sample size does not itself change a test's nominal significance threshold, but unstable estimates and selection of significant results can exaggerate apparent effect magnitude. Holm correction addresses specified families of comparisons, not these broader sources of uncertainty.

Individual random allocation supports the group comparison, but chance baseline differences remain possible in a small class. The experimental group had somewhat higher initial competency means, and the post-intervention ANOVA does not adjust for those differences. Future tests should prespecify baseline-adjusted models and formal comparisons of treatment effects across competencies. The integrated condition also does not separate immersion, automated feedback, and instructional mediation. Equal scheduled class time cannot establish equal individual practice or feedback exposure. The social-intelligence and perceived-learning findings remain nominal signals because neither survived its family's Holm correction.

The assessment design used one course instructor and separate single-item perception indicators. Replication should include independent performance assessment and broader multi-item measures of student perceptions, together with multiple instructors, larger cohorts, and delayed follow-up. Process evidence on rehearsal, uptake of avatar feedback, and subsequent revisions would help distinguish the proposed fit mechanism from alternative explanations. Prior VR experience, digital self-efficacy, and task complexity are also relevant candidates for future measurement.

Practical implications for AI-supported higher education

For educational leaders, the practical implication is to avoid technology-first adoption. AI and VR investments should begin with a competency map. If the target outcome requires performance, social interaction, communication, behavioral interpretation, or repeated rehearsal in context, immersive practice with AI-supported feedback is pedagogically defensible. If the target outcome is primarily procedural or conceptually structured, AI and VR may still be useful, but different supports may be more appropriate and more economical.

For faculty, the central design principle is alignment. Learning tasks should require students to use feedback, not merely receive it. Assessment should capture observable performance and disciplinary criteria rather

than rely primarily on satisfaction. Automated suggestions should be translated into discipline-specific improvement through instructor mediation. Under these conditions, AI and VR can support a responsible form of educational innovation in which technology expands opportunities for practice while human educators retain responsibility for meaning, standards, and assessment.

For curriculum designers, the study suggests a practical sequence: establish conceptual foundations through conventional instruction; identify competencies that require enactment or social interpretation; introduce immersive rehearsal where it can make performance visible; use AI feedback to accelerate formative cycles; and return to faculty-mediated assessment for contextual judgment. This sequence operationalizes competency-sensitive adoption and avoids the assumption that every learning objective benefits from the same technological treatment.

At the program level, competency-sensitive mapping can also improve adoption decisions. Rather than purchasing immersive technologies and subsequently searching for uses, schools can identify high-value competency targets first, specify the evidence needed to demonstrate improvement, and then select technologies whose affordances match those requirements. This reverses the usual adoption sequence and provides a more defensible basis for scaling AI and VR across curricula.

Conclusion

This randomized classroom study examined AI- and VR-supported feedback in an undergraduate business course. Both conditions showed improved rubric-based competency ratings. Organizational behavior was the only post-intervention outcome that favored the experimental condition after Holm correction; social intelligence and perceived learning showed nominal signals, while human capital processes showed no detected difference. These results offer preliminary evidence motivating prospective tests of competency-technology fit. They do not establish a causal fit mechanism or demonstrate that the treatment effects differ across competencies. The practical proposition is to align immersive rehearsal and feedback with specific, observable task demands while retaining faculty responsibility for assessment.

The study contributes a concrete classroom comparison and a testable account of why immersive feedback might support learning outcomes differently. Larger studies with direct tests of the proposed fit dimensions are needed to determine where this instructional approach offers a reproducible advantage.

References

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education - where are the educators?" Int. J. Educ. Technol. High. Educ., vol. 16, no. 39, 2019.

[2] H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," Int. J. Educ. Technol. High. Educ., vol. 20, no. 22, 2023.

[3] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson, "Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design," J. Comput. Educ., vol. 8, pp. 1-32, 2021, doi: 10.1007/s40692-020-00169-2.

[4] B. M. Kyaw et al., "Virtual reality for health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration," J. Med. Internet Res., vol. 21, no. 1, e12959, 2019.

[5] C. Dede, "Immersive interfaces for engagement and learning," Science, vol. 323, no. 5910, pp. 66-69, 2009.

[6] D. Carlgren, "Competency-based curriculum transition: A conceptual framework," J. Competency-Based Educ., vol. 6, no. 2, e1229, 2021, doi: 10.1002/cbe2.1229.

[7] C. Rivers, S. Gibson, E. Contreras, T. Livingston, and P. Hanson, "Competency-based education: An evolutionary higher education business model," J. Competency-Based Educ., vol. 4, no. 1, e1179, 2019, doi: 10.1002/cbe2.1179.

[8] T. D. Hubbard, "Transforming business education: The impact of virtual reality on learning outcomes in case studies," Front. Virtual Real., vol. 6, Art. no. 1461690, 2025, doi: 10.3389/frvir.2025.1461690.

[9] B. Sichterman et al., "Can AI feedback stand alone for fostering students' presentation performance? A comparison of AI-only versus teacher-supported AI feedback," Interactive Learning Environments, vol. 34, no. 3, pp. 1280–1297, 2026, doi: 10.1080/10494820.2025.2521338.

[10] D. A. Kolb, Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ, USA: Prentice Hall, 1984.

[11] D. H. Jonassen, "Instructional design models for well-structured and ill-structured problem-solving learning outcomes," Educ. Technol. Res. Dev., vol. 45, no. 1, pp. 65-94, 1997.

[12] J. Hattie and H. Timperley, "The power of feedback," Rev. Educ. Res., vol. 77, no. 1, pp. 81-112, 2007.

[13] P. Mishra and M. J. Koehler, "Technological pedagogical content knowledge: A framework for teacher knowledge," Teach. Coll. Rec., vol. 108, no. 6, pp. 1017-1054, 2006.

[14] W. R. Shadish, T. D. Cook, and D. T. Campbell, Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA, USA: Houghton Mifflin, 2002.

[15] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum, 1988.

[16] M. Vesisenaho et al., "Virtual reality in education: Focus on the role of emotions and physiological reactivity," J. Virtual Worlds Res., vol. 12, no. 1, 2019, doi: 10.4101/jvwr.v12i1.7329.