Creativity, Innovation and Entrepreneurship, Vol. 166, 2025, 1–9
https://doi.org/10.54941/ahfe1006104

A Software Engineering Decision Making Matrix for Assessing XR Suitability in Task Execution

Nicole Kosoris

Georgia Tech Research Institute, Atlanta, GA 30332, USA

ABSTRACT

XReality (XR) technologies, encompassing Virtual Reality (VR) and Augmented Reality (AR), offer transformative potential across domains like education and aviation, yet their implementation often lacks structured guidance, leading to mixed outcomes. This paper proposes a decision-making framework to evaluate XR suitability, addressing the absence of robust criteria for determining when and how to deploy XR systems. Through a comparative analysis of two case studies we identify factors influencing XR effectiveness. Drawing parallels with robotics' decision-making models, we introduce the DIVE acronym (Danger, Immersion, Verification, Expertise) to assess a proposed virtual system's suitability as an XR implementation. A novel decision matrix quantifies these criteria, incorporating subcomponents such as distraction mitigation, sensory immersion, action tracking, and expert guidance, weighted by scenario relevance and technology readiness. Applied to history lecture and aviation training scenarios, the matrix demonstrates how XR suitability varies with task demands, highlighting visual immersion for education and danger simulation for aviation. This framework equips program managers with a systematic tool to optimize XR integration, aligning technological capabilities with project goals to enhance adoption and efficacy in serious applications.

Keywords: Human systems integration, Virtual reality, Augmented reality

INTRODUCTION

There is a significant need to define a robust method for decision-making in using XReality (XR) across diverse projects, from training to operations. Numerous XR implementations have yielded mixed results, with some achieving success and others facing adoption challenges. Decisions to deploy XR technologies—like Augmented Reality (AR) heads-up displays or Virtual Reality (VR) training—often lack guidance on their suitability and optimal characteristics. A decision-making matrix could assist program managers in evaluating XR inclusion based on evidence.

The U.S. XR market was estimated at $25.7 million in 2024, with a global market of $140 billion in 2023, both projected to double by 2030 (Statista, 2025) (Research, 2025). Though 85% of the U.S. market focuses on entertainment, growth is expected in serious applications (healthcare, education, business). Despite this, XR development methods remain underdeveloped, with challenges often sidelining projects that seem promising (Learning, 2025) (ArborXR, 2025). Some issues await hardware advances, but others stem from choosing to use XR in poorly-fitted use cases, or from design failures that didn't understand what was most critical to emphasize.

We review two segments of XR cases: educational platforms, where metastudies report up to one-third failing to benefit students, contrasting with aviation's history of successful XR training. These examples inform a taskagnostic framework. A meta-review highlights XR's strengths in teaching abstract concepts (Luo et al., 2021), yet 13%–38% of educational studies fail due to poor fidelity and integration (Li Lee, 2024). Aviation's success, dating to 1977 (Orlansky String, 1977), leverages high-fidelity simulations for emergencies, reducing training time (Ortiz, 1993) and verifying skills (Bell Waag, 1998).

XR designers must carefully curate content to ensure the virtual world responds to immerse users in characteristics that primarily reinforce the central task and limit extraneous information. Studies like Bowman and McMahan's (Bowman McMahan, 2007) link training efficacy to realistic stimuli, while aviation's history (Board, 1992) shows danger, expert guidance, and verification drive success. Drawing on related matrices in other fields (Marr, 2025) (Robotnik, 2025) we will propose DIVE (Danger, Immersion, Verification, Expertise) for XR. Our matrix will allow us to evaluate XR suitability, demonstrated with a comparison of sample applications.

CONTEXTUAL XR CONSIDERATIONS

Systems engineering often assesses a technology's suitability for specific tasks. In robotics, for instance, a robust framework - summarized as the "4 D's": Dull, Dirty, Dangerous, and Dear - guides automation decisions (Marr, 2025)(Robotnik, 2025). Dull tasks involve repetition or frequent human distraction errors. Dirty and Dangerous are self-explanatory, while Dear tasks save significant time or cost via automation. The more D's a task meets, the stronger its case for automation. An XR decision-making matrix should start by assessing the necessity of simulation for the task, as well as the importance of physical precision. Training that involves symbolic abstraction depends less on physical precision, but tasks that can be fully abstracted are unlikely to require a virtual environment and thus are poor candidates for virtualization. The optimal candidates lie somewhere in between these extremes.

Serious Games literature suggests optimal gamification reinforces the task at hand, ensuring that the game's mechanics and objectives are closely aligned with the desired learning outcomes. This principle is supported by XR meta-studies, which highlight that failures often stem from poor instructional integration and low fidelity (Luo et al., 2021)(Li Lee, 2024). 'Fidelity' here means having accurately representative interactions and responsive mechanics that support the intended learning or task objectives.

Effective virtualization directly supports the task. This is currently challenging beyond visual and auditory fidelity, but haptics are advancing.

In some cases hybrid approaches, like AR overlays with physical props, can bypass haptic limitations. For example, consider a control box with a visual display in a noisy control room. A low-fidelity 3D-printed panel (e.g., buttons registering presses, not linked to the real system), paired with a VR display and simulated background noise, mimics key interactions with physical precision only where it counts cost-effectively. Contrast this with a task demanding precise physical manipulation in a wide and dynamic environment such as firearm training — where a VR "dummy" gun differing significantly in feel, aim, or response could disrupt learning, potentially worsening performance.

Tasks suited for virtualization often involve symbolic abstraction common to the field, enhancing capability through representation. Yet, virtualization broadly might be achieved via desktop tools, which are cheaper than immersive XR. This raises the question: when should one pursue XR over a simpler desktop solution?

The DIVE framework — Danger, Immersion, Verification, Expertise — offers support in decision making phases. These criteria assess XR's fit by examining risk simulation, sensory engagement, performance tracking, and expert guidance needs. Lets explore how each component of DIVE explains failures and successes in XR, using educational and aviation simulation paradigms as our reference. In particular, you'll see that educational XR failures often occur due to misaligned immersion and weak verification, where aviation is an XR success driven by effective danger simulation and robust tracking.

Danger is perhaps the most obvious. Within this consideration, we must balance whether said necessary danger can be adequately induced. Consider both whether the danger can be replicated symbolically (can you tell a story that generates the necessary feeling of danger), and whether you can avoid a danger inherent in attempting the scenario in the real world. Additional to this category, include "distraction". If an atmosphere of danger or distraction is a component of the operating task, and ignoring such is critical to the task, then there may well be significant value in undertaking this in VR.

In the context of the educational XR systems, danger or distraction was rarely a critical component of the tasks being taught. Many of these systems aimed to teach abstract concepts or procedural knowledge, where the environment was controlled and safe (Luo et al., 2021). Where distraction was a component, it was poorly considered; applications failed simply because students were distracted by the novelty of the VR training system rather than by task-relevant stressors. Instead, the inclusion of extraneous information overwhelmed learners, indicating a misalignment with the task's core needs.

In contrast, aviation training thrives on danger as a central component. Training scenarios often involve emergency situations, such as recovering from equipment failures or navigating hazardous conditions, which are impractical or unsafe to replicate physically (Orlansky String, 1977). XR systems allow trainees to experience these dangers symbolically - through simulated engine failures or turbulent weather - while maintaining safety. The ability to replicate high-stakes environments where distraction must be managed (e.g., alarms, crew communications) is critical to aviation training success.

Immersion has a great deal of dimensionality to it. Simulation literature traditionally defines immersion simply as the ability to replicate with precision, but here let us consider the term more generally to include a storied type of immersion (Kosoris Gandy, 2023). The kinds of physical immersion we can reliably produce at present are relatively limited, but we can, for example, embody visual interaction using head movement.

For educational XR systems, meta-reviews noted that low interactive fidelity and poor integration of instructional strategies led to failures (Li Lee, 2024). We see here that, although high demand for immersion is a key component of whether one should create an XR training system, poorly replicating immersive characteristics can cause the platform to fail. This highlights, rather than diminishes, immersion as a necessary component of XR. For instance, if a chemistry VR module provided stunning visuals but failed to allow precise interaction with molecular models, the immersion did not reinforce the task. The presentation of extraneous information further diluted the learning experience, as students were overwhelmed by irrelevant sensory stimuli rather than focused on key concepts.

Aviation XR systems, however, excel in leveraging immersion. The highfidelity visual and auditory environments replicate the real-world context with precision (Bowman McMahan, 2007). This immersion is critical because pilots must develop situational awareness in complex, dynamic settings. For example, a 1993 study showed that VR-trained students mastered flight maneuvers faster due to the realistic sensory stimuli provided (Ortiz, 1993).

Verification is, quite simply, the ability to track the user. A need to know what the user is looking at or doing and the precise context of movement and action make VR and AR an excellent choice. Centering Verification in your application needs is an excellent reason to pursue XR.

In educational XR systems, meta-reviews indicated that many systems failed to track whether students were mastering the intended skills, focusing instead on delivering content (Li Lee, 2024). Educational applications have the best possible case for verification – if you are attempting to teach someone, determining whether you succeeded should be central to your design. A history VR module that immerses students in a virtual ancient city but fails to assess whether they understood key events or concepts is an educational failure, no matter how accurate its representation.

Aviation XR systems, by contrast, heavily rely on verification and assess performance against strict standards (Bell Waag, 1998). For instance, simulators verify that a pilot correctly responds to an engine failure by monitoring their sequence of actions and timing. This ability to confirm skill mastery is critical for certification and safety, making Verification a cornerstone of aviation XR success.

Expertise, the final and most tentative DIVE category, is the ability to deliver immersive-dependent expert guidance. It's only relevant for specialized, interactive tasks where expertise is rare and valuable—not mundane or non-interactive ones. Currently, Expertise-driven XR is rarely cost-effective, except in niche cases like guiding a user through an explosive detonation scene, highlighting key details for event reconstruction. We retain this category for its AI-enhanced future potential.

In educational XR, expert guidance was seldom prioritized; platforms delivered general knowledge, not specialized instruction (Luo et al., 2021). Meta-reviews show many failures stemmed from lacking tailored direction, leaving students lost in complex virtual settings (Li Lee, 2024).

Aviation XR has embedded expert guidance effectively, although often via shortcuts like instructors joining trainees in simulators (Board, 1992). In another case, a system might flag a missed checklist item in an emergency simulation, mimicking an expert's role. This delivery of expertise ensures that trainees not only perform tasks but understand the underlying principles, enhancing skill transfer.

The DIVE framework directly informs the evaluation of XR implementation by addressing key decision criteria. For complex tasks, such as those requiring intricate procedural knowledge, Danger and Immersion ensure that XR systems replicate critical environmental stressors or sensory cues, while Verification confirms user proficiency across varied scenarios. Safety concerns that can be mitigated are also assessed under Danger, allowing hazardous scenarios to be practiced without physical risk. Human performance requirements are met through Immersion and Expertise, which tailor sensory and instructional inputs to enhance cognitive and motor skills, avoiding the educational pitfalls of extraneous information. To prevent misalignment between training and real-world application, Immersion and Verification prioritize fidelity and measurable outcomes.

Finally, as a portion of cost-benefit analysis, we can consider each component of DIVE. Danger mitigation costs can be assessed, but the increase in Immersion needed to simulate danger will drive costs up. Expertise guides cost-benefit analyses by identifying when XR's immersive and tracking capabilities justify development expenses, when expert support will be costly or unavailable. Verification provides a means of proving the utility of the tool, but increases in centrality of Verification will drive cost up – as the kind of verification becomes more environmentally responsive it will prove more useful but also be more costly to implement. Leveraging DIVE in your cost-benefit analyses will ensure your system's XR applicability to real-world needs while balancing resource constraints.

DECISION MAKING MATRIX

The decision-making matrix now demonstrates its practical application by supporting consideration of XR implementation of two proposed systems. The matrix evaluates proposed XR applications using the DIVE framework, with each criterion broken down into specific sub-components: Danger assesses whether the simulation centers on the mitigation of dangerous or distracting elements; Immersion examines the need for auditory, visual, haptic, or storied immersion; Verification considers the centrality of tracking actions or knowledge; and Expertise evaluates the benefit of co-situated expert guidance. For each sub-component, the matrix captures two inputs: the relevance of that sub-component to the scenario (i.e., how critical it is to the task) and the readiness of current XR technology to meet the required standard. By weighting and scoring these factors, the matrix provides a nuanced assessment of XR suitability, supporting program managers in aligning technology choices with project goals and resource constraints.

The decision matrix is structured to quantify the suitability of XR systems by evaluating each DIVE sub-component across two dimensions: Relevance, which measures the importance of the sub-component to the proposed scenario on a scale of 0 to 10 (where 10 is critical), and Technology Readiness, which assesses the capability of current XR technology to meet the necessary standard, also on a scale of 0 to 10 (where 10 is fully capable). The score for each sub-component is calculated as the product of Relevance and Technology Readiness, divided by 10 to normalize the contribution (i.e., Score = $(R \times T)/10$). Each sub-component is assigned a weight reflecting its relative importance to the scenario, with weights summing to 1.0 to ensure balanced evaluation. The total score for a system is the weighted sum of all sub-component scores, providing a clear metric for comparing XR solutions. To illustrate, we apply the matrix to two contrasting scenarios: a history lecture, representing an educational application where XR has faced adoption challenges, and aviation training, where XR is a proven success. These examples demonstrate how the matrix differentiates suitability based on scenario-specific demands.

History Lecture

For the history lecture, visual immersion is paramount (Relevance = 9) to vividly recreate historical settings, with VR-System 1 excelling due to superior graphics (Tech Readiness = 9) compared to AR-System 2, which is less capable in delivering high-fidelity visuals (T = 5). Auditory immersion, such as ambient city sounds, is moderately relevant (R = 5), with VR-System 1 slightly stronger (T = 8 vs. T = 7). Haptic immersion is irrelevant (R = 0), as tactile interaction with artifacts is not required, rendering technology readiness scores moot. Dangerous elements are negligible (R = 1), as the classroom poses no physical risk, though distraction, such as simulating a bustling marketplace to teach focus, has some value (R = 4), with ARSystem 2 better at overlaying dynamic environments (T = 8 vs. T = 7). Action verification, like navigating the virtual city, is unnecessary (R = 0), so tracking capabilities do not contribute. Knowledge verification is critical (R = 8), with AR-System 2 offering stronger tracking (T = 8 vs. T = 7). Expert guidance, such as a virtual historian's narration, is moderately important (R = 6), with AR-System 2 slightly ahead (T = 7 vs. T = 6). Weights prioritize visual immersion, knowledge verification, and expert guidance, reflecting educational goals. VR-System 1's higher total score (24.9 vs. 22.4) highlights its strength in visual immersion. for XR implementation, as indicated by its relatively weak score.

Table 1: Decision matrix for history lecture.

Criterion	Sub-Component	Wt	VR (R/T/S)	AR (R/T/S)
Danger	Dangerous Elements	0.06	1/8/0.8	1/6/0.6
	Distraction	0.12	4/7/2.8	4/8/3.2
Immersion	Auditory	0.12	5/8/4.0	5/7/3.5
	Visual		9/9/8.1	9/5/4.5
	Haptic	0.00	0/4/0.0	0/5/0.0
Verification	Actions	0.00	0/8/0.0	0/9/0.0
	Knowledge	0.18	8/7/5.6	8/8/6.4
Expertise	Expert Guidance	0.22	6/6/3.6	6/7/4.2
Total Score			**24.9**	**22.4**

Aviation Training

In aviation training, dangerous elements are critical (R = 9) to simulate high-stakes scenarios like engine failures, with VR-System 1 better equipped (T = 8) than AR-System 2 (T = 6), which will have a limited feeling of presence for users. We weighted Danger as a single element to more directly compare systems, but it would be recommended in this case to split out the system's alleviation of danger from its ability to accurately give users a feeling of danger; they are separate but both useful considerations. Distraction, such as cockpit alarms, is highly relevant (R = 8), with ARSystem 2 slightly stronger (T = 8) in overlaying dynamic cues. Visual and auditory immersion are essential (R = 9 and R = 8) for realistic cockpit displays and sounds, with VR-System 1 leading in visuals (T = 9) and both systems performing well in audio. Haptic immersion, like control yoke feedback, is moderately relevant (R = 6) but limited by technology (T = 4 for VR, T = 5 for AR). Action verification, such as correct switch flips, is paramount (R = 9), with AR-System 2 excelling (T = 9). Knowledge verification, like recalling procedures, is less critical (R = 6), with AR-System 2 slightly better (T = 8). Expert guidance, via virtual instructors, is valuable (R = 7), with AR-System 2 marginally ahead (T = 7). Weights emphasize dangerous elements, visual immersion, and distraction, reflecting aviation's demands. VR-System 1's higher score (45.3 vs. 44.5) highlights its strength in simulating dangerous scenarios and visuals; this system's score is very significantly higher than our historical application, aligning with aviation's proven XR success.

Table 2: Decision matrix for aviation training.

Criterion	Sub-Component	Wt	VR (R/T/S)	AR (R/T/S)
Danger	Dangerous Elements	0.20	9/8/7.2	9/6/5.4
	Distraction	0.15	8/7/5.6	8/8/6.4
Immersion	Auditory	0.15	8/8/6.4	8/7/5.6
	Visual		9/9/8.1	9/7/6.3
	Haptic	0.10	6/4/2.4	6/5/3.0

Continued

Table 2: Continued

Criterion	Sub-Component	Wt	VR (R/T/S)	AR (R/T/S)
Verification	Actions	0.10	9/8/7.2	9/9/8.1
	Knowledge	0.05	6/7/4.2	6/8/4.8
Expertise	Expert Guidance	0.05	7/6/4.2	7/7/4.9
Total Score			**45.3**	**44.5**

CONCLUSION

Ultimately, developing and utilizing a robust decision-making matrix offers a comprehensive approach to integrating XR technology. It supports program managers by equipping them with a strong tool for evaluating the inclusion of XR elements, helping ensure these technologies align effectively with the project's objectives and constraints. By contrasting the failures of educational XR systems with the successes of aviation training, we see that the DIVE criteria—Danger, Immersion, Verification, and Expertise—capture the critical factors influencing outcomes. Educational systems often failed due to misaligned immersion and insufficient verification and expertise, while aviation systems succeeded by addressing all four criteria, ensuring realistic, verifiable, and expert-guided training in highstakes scenarios. This framework, grounded in systems engineering and informed by decision-making models from fields such as robotics, provides a structured path for optimizing XR implementation across diverse applications.

REFERENCES

ArborXR. (2025). VR Training Challenges. https://arborxr.com/blog/vr-training-challenges/. Accessed March, 2025.

Bell, H. H., Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. The International Journal of Aviation Psychology, 83223–242.

Board, U. A. F. S. A. (1992). Report of the Support Panel of the Air Force Scientific Advisory Board Ad Hoc Committee on Concepts and Technologies to Support Global Reach-Global Power 1995–2020, Tech. Rep. Washington, DC, U. S. Air Force.

Bowman, D. A., McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? Computer, 40736–43.

Kosoris, N., Gandy, M. (2023). Immersion, Presence and Transference: Bringing Story Thinking to Simulation Design. In 2023 Intermountain Engineering, Technology and Computing (IETC) (pp. 83–87).

Learning, R. (2025). 5 Problems with Virtual Reality Training They Don't Want You To Know. https://roundtablelearning.com/5-problems-with-virtual-reality-training-they-dont-want-you-to-know/. Accessed March, 2025.

Li, X., Lee, J. J. (2024). Failure Stories and Surprising Findings: Learning from When VR Did Not Work. In M.-L. Bourguet, J. M. Krуğer, D. Pedrosa, A. Dengel, A. Pena-Rios, J. Richter (Eds.), Immersive Learning Research Network, (pp. 22–36). ChamSpringer Nature Switzerland.

Luo, H., Li, G., Feng, Q., Yang, Y., Zuo, M. (2021). Virtual reality in K-12 and higher education: a systematic review of the literature from 2000 to 2019. J. Comput. Assist. Learn. 373887–901.

Marr, B. (2025). The 4 D's of Robotisation: Dull, Dirty, Dangerous and Dear. https://bernardmarr.com/the-4-ds-of-robotisation-dull-dirty-dangerous-and-dear/ . Accessed March, 2025.

Orlansky, J., String, J. (1977). Cost-effectiveness of Flight Simulators for Military Training: Vol. I. Use and Effectiveness of Flight Simulators. IDA Paper No. P-1275. Arlington, VA, Institute for Defense Analysis.

Ortiz, G. A. (1993). Transfer of Learning Effectiveness: PC-Based Flight Simulation. Journal of Aviation/Aerospace Education & Research, 32. 10.15394/jaaer.1993.1097.

Research, G. V. (2025). Extended Reality (XR) Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By End-use (Retail, Education, Automotive, Industrial & Manufacturing), By Application, By Region, And Segment Forecasts, 2023-2030. https://www.grandviewresearch.com/industry-analysis/extended-reality-xr-market-report. Accessed March, 2025.

Robotnik. (2025). Advantages and Benefits of Industrial Robots: Improving the 4 Ds. https://robotnik.eu/advantages-and-benefits-of-industrial -robots-improving-the-4-ds/. Accessed March, 2025.

Statista. (2025). Market size of augmented and virtual reality worldwide from 2016 to 2027 (in billion U. S. dollars). https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/. Accessed March, 2025.