From Locked Rooms to Open Minds: Escape Room Best Practices to Enhance Reflection in Extended Reality Learning Environments

Erica Kleinman
Ghost Lab
Northeastern University
Boston, Massachusetts, USA
e.kleinman@northeastern.edu

Rana Jahani
Northeastern University
Boston, Massachusetts, USA
jahani.r@northeastern.edu

Eileen McGivney
Art + Design and Communication Studies
Northeastern University
Boston, Massachusetts, USA
e.mcgivney@northeastern.edu

Mehmet Kosa
Marshall University
College of Engineering and Computer Sciences
Huntington, West Virginia, USA
kosam@marshall.edu

Seth Cooper
Northeastern University
Boston, Massachusetts, USA
se.cooper@northeastern.edu

Casper Harteveld
Northeastern University
Boston, Massachusetts, USA
c.harteveld@northeastern.edu

Abstract

Extended reality (XR) learning environments result in greater knowledge gains when coupled with opportunities to reflect on one's actions and learning. However, when and how one should prompt reflection in XR learning environments (XRLEs) to effectively enhance learning, without breaking immersion, remains an open question. In this work, we argue that we can extract insights on how to design effective, immersive reflection for XRLEs from the expertise of escape room game masters (GMs) who regularly provide reflective hints and prompts in complex, immersive problem solving environments. To explore what we can learn from GMs, we conducted exploratory semi-structured interviews with 13 escape room GMs and, via iterative open coding, captured their best practices in how they provide hints and give nudges to escape room players. From these results, we present a foundation and model of how GMs observe and intervene and discuss implications for XRLE-based reflection.

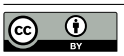
CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

Keywords

Reflection, Extended Reality, Learning Environments, Reflection Prompts, Escape Rooms

ACM Reference Format:
Erica Kleinman, Rana Jahani, Eileen McGivney, Mehmet Kosa, Seth Cooper, and Casper Harteveld. 2025. From Locked Rooms to Open Minds: Escape Room Best Practices to Enhance Reflection in Extended Reality Learning Environments. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3713811



CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713811

1 Introduction

Extended reality (XR) is an umbrella term for virtual, mixed, or augmented reality, which, in turn, are all environments that either simulate or blend the real, physical world with a digital one to varying degrees using technology such as head-mounted displays, phones, video screens, or computers [47]. The last decade saw substantial interest in leveraging XR in the domains of education and training [17]. For example, XR technology is often used to provide safe, hands-on training for manufacturing tasks, where real-world training can be both expensive and dangerous [33, 42, 91], and health and medicine, where real-world training could put patients at risk [89, 107]. In more traditional academic contexts, XR technology is used to take students on field trips to locations they otherwise would not be able to visit (e.g., the International Space Station) [66] and let them interact directly with lesson subject matter [43]. For the purposes of this paper, we refer to these environments that use XR to teach and train collectively as XR Learning Environments (XRLEs). Recent work, however, found that, while XRLEs are engaging, they often do not improve learning gains on their own [61]. Instead, learning supports must be integrated to ensure that the student's interaction with the XRLE goes beyond mere enjoyment and is an actual educational experience [46, 76].

One type of learning support showing great promise, and therefore garnering great interest, is reflective prompts, which can improve learning outcomes from XRLEs compared to lessons without reflection opportunities [22, 46, 68]. Reflection is not new to learning technology: digital and gamified learning environments often leverage reflection through prompting, data sharing, or visualizations [10, 96]. Open learner models (OLMs), a classic form of learning support technology, have also typically placed reflection support and reflective opportunities at the center of their design [14, 31, 63, 92]. Prior work demonstrated that including reflection in these digital learning contexts improves learning outcomes [74],

and many frameworks and guidelines exist for designing for such reflection in digital learning [48, 56, 58, 90]. Reflective prompting is also seeing new advancements into dynamic, automated forms, spurred by the recent advancements in artificial intelligence (AI) technology and the rise of large language models (LLMs). As a result, the last few years saw an increased interest in AI-powered reflection for learning [35, 73, 77], which already started to produce guidelines for design [1, 13].

XRLEs, however, take the immersion and interaction of digital learning a step further due to the way that a virtual environment truly surrounds the learner and, often, is not sitting stationary at a desk or computer but rather moving about [21, 38]. As such, existing designs for reflection may not apply to XRLEs as they do not account for this dynamic, immersive interaction. For example, they often rely on students to communicate their thoughts via written text [14, 74], which is difficult to replicate in XRLEs. In fact, prior work on reflection in XRLEs found that existing ways of prompting reflection break flow and immersion and risk interrupting the immersive learning process [46, 68]. Addressing these concerns, however, remains an open problem, as there have yet to be concrete and widely adopted suggestions for designing reflection for such immersive experiences.

In this work, we argue that we may be able to glean insights into designing reflection for XRLEs from escape room game masters. Escape rooms are immersive puzzle games that challenge players to solve a series of interconnected puzzles to escape from a seemingly locked, and usually themed, room [25]. They are typically facilitated by a game master (GM) who observes players' progress and interjects with guidance or assistance as needed to help players succeed. While escape rooms can exist in various formats, including board games or two-dimensional digital games, for the purposes of this work, we focus on escape rooms that take place in a physical, three-dimensional space (i.e., an actual room).

Given this definition, our argument that reflection insights can come from this environment is based on two key observations. First, escape rooms are essentially learning environments where players must exercise various meta-cognitive skills, including problem solving, communication, and critical thinking, to solve increasingly complex puzzles. This connection between escape rooms and learning has not gone unnoticed by educators either, with many educational escape rooms appearing in recent years [25, 95]. Second, the interjections that come from the GM of an escape room may come in the form of an explicit clue or suggestion, e.g., "check the drawer", but may also come in the form of a reflective prompt, e.g., "Do you remember where you have and have not searched yet?" In other words, these exchanges are moments in which players are prompted to reflect in an immersive learning environment. As such, game masters likely possess valuable knowledge on how to judge when learners in an immersive environment would benefit from reflection, how to prompt them without breaking their immersion or engagement, and how to determine if it worked or if more aid is needed—all information that would be needed to design effective reflection for XRLEs.

Based on these observations, we explore what insights for prompting reflection in XRLEs can be extracted from interviews with escape room game masters. While the insights gleaned from escape rooms could also apply to other immersive learning environments, e.g., digital games [15], here, we focus on applying them to XRLEs, specifically, due to the immersive and spatial nature of the two environments. While XRLEs, like escape rooms, typically require learners to move about in a three-dimensional space and take action with their actual bodies, leading to heightened immersion, educational games are typically presented via computer screens and lack the immersive and spatial component. Further, we recognize the problem of immersive reflective prompts to be more pressing in the XRLE context due to the aforementioned immersive and spatial nature, thus rendering them a more relevant context for this work.

To guide this exploratory work, we developed the following research question: *"What can we learn from how game masters provide hints and nudges to escape room players to enhance reflection in XRLEs?"* While terminology and practices are not consistent across escape rooms, for the purposes of this work, we refer to assistance that was specifically requested as a "hint" and assistance that was provided at the discretion of the GM as a "nudge," consistent with how the majority of our participants differentiated between the two forms of aid. To answer our question, we conducted semi-structured interviews with 13 game masters with various levels of experience in the United States. From the interviews, we extracted ten themes sorted into three categories related to observations of player *progression*, players' *experience* of the escape room, and the *delivery* of aid. From these themes, we synthesize and discuss a foundation of how these three categories, and the themes within, describe what GMs look for in players, how they interject to provide aid, and how these findings can inform the design of reflective prompting in XRLEs. We translate these findings into ideas for the use case of a human (i.e., educator) and the use case of technology prompting, guiding, assessing, and responding to reflection in XRLEs. Through these discussions, we contribute suggestions for innovative design opportunities of how XRLEs can better integrate reflection to enhance learning in such immersive and spatial learning environments.

2 Related Work

In the following sections, we review relevant literature on how reflection is elicited via prompts, visualization, and conversation in learning environments. We then transition into a discussion of extended reality learning environments (XRLEs) and the role of reflection within them. We conclude with a brief introduction to and justification for examining escape rooms.

2.1 Eliciting Reflection in Learning

Reflection is a metacognitive process by which learners explore and review their experiences to reach new understandings [9]. It is a key part of learning, particularly when the goals include deeper competence and confidence beyond the subject's surface features. Through reflection, learners can overcome challenges [88] and adapt to new situations [45], ultimately making them stronger learners. Prior work already demonstrated that reflection improves learning outcomes in various contexts [2, 85, 98]. As such, there is interest in effectively triggering reflection [19] and evaluating its quality and impact [57] to better leverage it to improve learning experiences and outcomes.

In traditional learning environments, reflection is often elicited through "prompts" or brief interventions that request the learner

write or dictate a response to a question (or set of questions) or elaborate their thoughts on an idea [62]. For example, Rakovik et al. [82] had biology students write short essay responses to five reflective prompts regarding their activities in and around the class after their first exam. Such prompts can also vary in terms of specificity. Kramarski and Kohen [49] illustrate this point in their work, which provided students in a teaching program with generic (e.g., stop and reflect) or specific (e.g., reflect back and ahead on the experience) prompts to examine and compare the effectiveness of the two styles on the learners' self-awareness. In earlier work, Davis [19] also examined the differing impact of generic vs. directed prompts on middle school students, finding that both the quality of reflection and learning gains varied depending on the style of the prompt. Interestingly, the results of these two studies contradict one another, with one finding generic prompts to be better for reflection and the other finding them to be worse [19, 49], suggesting that there is a time and place for both approaches.

As learning technology evolved, so did the ways of eliciting reflection from learners. One trend of note is visualizations of student or community (i.e., the class) performance in a learning context. Commonly utilized in learning analytics and open learner modeling [8, 30], but also becoming more popular in game-based learning as well [96], the aim is to provide students with data that allows them to contextualize their progress in terms of class goals or classmates and motivate better learning behaviors. In doing so, this approach is conceptually similar to using data to promote reflection in other areas of human-computer interaction, such as personal informatics [6, 67]. An example of this approach can be seen in the recent work of Maram et al. [63], who developed an open learner model for an educational game that teaches parallel programming. The model uses process visualizations to show students where they performed poorly and expose them to alternative ideas via the data of their classmates. Their user study highlights how access to this information prompted students to reflect on their learning process and consider how they would adapt.

Conversing with an AI is also an effective way to elicit deep reflection [30]. As a result, many digital learning environments opt for dialogue-style interactions with a chatbot or similar agent as an alternative to traditional prompting (with or without any visualized component) [50, 101]. Winkler et al. [99], for example, presented Sara, a conversational agent embedded within an online learning environment, who, as they discuss in their work, helped students reflect on their understanding of the topic through scaffolded dialogue. In another example, Wolfbauer et al. [100] developed Rebo, a chatbot that prompts users to reflect on their work experiences as learning opportunities, which they found was successful in doing so with 98% of their participants.

All of this work demonstrates the various and creative ways that learners are prompted to reflect in analog and digital learning environments. However, most of the work until now has focused on what one might call "traditional" digital learning environments, where a student is situated in front of a monitor. Extended reality learning environments (XRLEs) add a layer of complexity by immersing the learner such that they are situated within the environment itself. This change in interaction modality makes many existing approaches to integrating reflection difficult to implement, warranting a need for domain-specific considerations.

2.2 Extended Reality Learning Environments (XRLEs)

Extended reality (XR) refers to technologies that surround the user in 3D images and sound, making them feel immersed in either a fully digital or mixed physical and digital environment, encompassing virtual, mixed, and augmented reality. XR can enable learning opportunities that are difficult or impossible in traditional classrooms, such as transporting students to inaccessible places and visualizing phenomena at scales not possible in real life [78]. The primary affordances of XR for learning are its immersive capability that makes learners feel present in a different place, and the agency it gives them through its novel forms of interactivity with digital objects and environments [60]. These affordances can enable immersive learning experiences that situate learning in meaningful tasks and environments previously only possible through experiences like apprenticeships, field trips, and exchange programs [21, 54]. In fact, XRLEs can improve upon physically situated learning experiences not only because they make them more accessible but also because the experience can be designed to maximize learning. For example, XRLEs can scaffold learners' experience by giving them progressively more challenging tasks, overlaying information onto objects and environments, and automatically measuring their learning by collecting multi-modal performance data.

Despite the promise of XR to deliver these rich learning experiences, reviews and meta-analyses find that using XR at times increases learning outcomes over using other media, but at others has no effect or even a negative effect [28, 81, 102]. One explanation is that the stimulus and arousal of using XR lead to extraneous cognitive load, making it difficult for learners to focus or to process the information they receive in the environment [65]. For example, Makransky et al. [61] found learners retained less when using a lab simulation in VR than a 2D version.

Reflection, however, has been shown to help learners manage their cognitive load in VR while increasing their intrinsic motivation to learn, suggesting it can mitigate the information retention issues of previous work. One study showed that a reflective activity increased learning outcomes over only viewing a VR experience straight through: participants in the reflective condition removed their headset and wrote a summary of what they had just viewed six times throughout the VR experience [76]. Another study found a generative learning activity in which participants worked in pairs to teach what they had learned to a peer immediately following a VR or computer-based simulation increased learning retention and transfer most for the VR participants [46]. Qualitative research on classroom VR implementation also found that learners can better understand what they see and do in VR field trips when given opportunities to reflect and discuss with their peers immediately following the VR experience [66].

While these prior studies illustrate the benefit of reflection in supporting learning in XRLEs, questions remain about the most effective ways to incorporate it within an XR learning experience rather than prompt reflection afterward. One study utilized segmentation in a VR learning experience about climate change, pausing and encouraging users to think about what they had just seen and heard, but this did not improve learning outcomes over engaging straight through without pauses [79]. Therefore, more research

is needed to understand when and how such prompting is most effective in XRLEs.

2.3 Escape Rooms

In the mid-2010s, escape rooms emerged as a popular pastime, challenging small groups of players to a game that required them to solve puzzles and answer riddles to escape from a "locked" room [25]. The games grew in popularity, and escape rooms can now be found in most major cities. Research into escape rooms and how they can be used for serious or high-impact domains has grown since [71, 72]. Examinations into the rooms themselves and their players found that the experiences offer more than just entertainment. Like many gaming experiences, escape rooms can nurture real-world skills, especially those related to collaboration, communication, and teamwork [75, 95].

Important for our purposes, escape rooms promote and require critical thinking and problem-solving skills [5, 20, 29, 32, 103], making them comparable to learning environments. Unlike traditional learning environments, however, escape room players are immersed within the environment of the room, similar to learners' state of immersion within an XRLE. Escape rooms also almost always feature a "game master," a type of facilitator who watches what happens in the room from a control room and, most notably, provides hints or nudges to the players when they ask or become stuck in their progress. These prompts are integral to the experience and are meant to guide players or help push them forward, not unlike reflective prompts in learning environments. To illustrate, a common structure for reflective prompts in learning is "please review/evaluation...what worked/did not work?" [45, 57]. Escape rooms may, similarly, feature hints along the lines of "think about [something noticed/already done] what worked/did not work?". More importantly, these escape room aids are exchanged so that they do not break player engagement. For these reasons, we turn to escape rooms and game masters to seek insights that may address the problems with integrating reflective prompts in XRLEs that we discussed previously.

3 Methods

To explore our idea and answer our research question, we conducted semi-structured interviews with current and former escape room game masters. Through interviews, we could ask targeted questions regarding our research interest (prompting reflection via nudges and hints) but follow up on what they say to get deeper insights into their experiences and expertise.

3.1 Participants

Participants were recruited through social media, email, escape room special interest groups, and the researchers' personal networks. Participants were required to be over 18 years of age, located in the United States, capable of communicating in spoken English, and either currently or formerly have worked as a GM for an escape room. No other inclusion or exclusion criteria were applied. Participants were not compensated, and the study was a strictly voluntary activity. A total of 13 participants were recruited for the study. We discuss saturation of the data below.

The GMs had a range of experience that can be seen in Table 1. Experience ranged from 5 months to 7 years and most had experience overseeing rooms of various difficulties About half had also helped design rooms at some point. Beyond what is presented in the table, no other demographic information (i.e., age) was considered relevant for the current study.

3.2 Interview Protocol

Interviews were one-on-one and conducted over Zoom between May and July 2024. Audio was recorded. All participants were assigned a random ID number. Interviews lasted between 15 and 52 minutes with an average of 32 minutes. University IRB approved the protocol.

The interview (see Table 2 to view questions) was intended to capture various elements of GM assistance in escape rooms, including when it is given, how GMs determine it is needed (both for hints and nudges), how it is phrased or delivered to the player, and what type of assistance it is.

Interviews were semi-structured, meaning that the attending researcher would ask follow-up questions when necessary, and participants could speak for as long as they desired. Questions were skipped if they were deemed irrelevant to the participant (i.e., if they did not provide nudges when players did not ask). Participants were also allowed to skip any question they did not wish to answer or ask for further explanation for any question they did not understand. At the beginning of the interview, the attending researcher asked the GMs for their years of experience as a GM, the number of rooms completed as a player, whether or not the GM had ever designed a room, and the average difficulty of rooms GMed (see Table 1). After the interview, the author allowed the participants to ask any remaining questions before thanking them for their time.

3.3 Data Analysis

After each interview, the first author noted initial emerging themes. Apparent saturation was observed quickly: by the eighth interview, no new themes were observed. Five additional interviews were conducted to confirm, and no additional themes emerged. All authors reviewed the themes and determined that saturation [26] had been reached. 13 participants is consistent with common standards for similar, qualitative research [27, 44, 97].

Interviews were transcribed using Adobe Premiere, cleaned manually, and analyzed using iterative open coding [87]. For this purpose, all interview transcripts were imported into a spreadsheet with each individual interview transcript contained within its own tab and each utterance (of either the interviewer or GM) in its own line. The transcribed utterances were contained in the left-most cell of each line and the remaining cells were used for coding. The unit of analysis was an entire utterance by a participant. The first and second authors conducted the analysis, first reviewing a representative sub-set containing 30% of the data [12]. Independent of each other, they labeled each participant utterance, using the empty cells to the right, with initial emerging themes, quotes, or notes or memos and organized and consolidated their labels into initial codes. Together, they then discussed their findings, collapsed overlapping codes, and generated a combined list of 16 initial themes.

Table 1: Demographic information for the GMs interviewed. For the communication category, "verbal" refers to information conveyed by speaking over audio equipment (i.e., a speaker or a walkie-talkie). Some GMs mentioned stepping into the room to explain something in extreme situations, but this was not the standard form of aid. "# played" refers to the number of rooms they have attempted as a player.

GM	Yrs as GM	Hours	# Played	Designer?	Communication	Difficulty of Rooms GMed
1	5 Months	Full Time	1	No	Verbal and Text	Beginner to Advanced
2	6 Years	Full Time	50	Yes	Verbal	Beginner to Advanced
3	3 Years	Part Time	3	No	Verbal	Intermediate
4	1.5 Years	Part Time	12	Yes	Text	Beginner to Advanced
5	1.5 Years	Part Time	37	No	Text	Intermediate
6	3 Years	Full Time	40	No	Live Actors	Beginner to Advanced
7	7 Years	Part Time	225	Yes	Verbal and Text	Beginner to Advanced
8	2 Years	Part Time	36	Yes	Text	Beginner to Advanced
9	3 Years	Full Time	175	Yes	Verbal	Advanced
10	3 Years	Full Time	50	Yes	Text	Intermediate to Advanced
11	3 Years	Part Time	25	No	Verbal and Text	Intermediate to Advanced
12	1 Years	Full Time	7	No	Verbal and Text	Intermediate to Advanced
13	2.5 Years	Part Time	10	No	Verbal and Text	Beginner to Advanced

Table 2: The interview outline for the escape room game masters.

Part 1: Regarding when assistance is given and how GMs determine it is needed
For nudges provided when the players do not ask
1. How often do you typically provide nudges when not explicitly asked?
2. How do you identify when a participant needs a nudge?
3. How do the participants typically respond to these unprompted nudges?
4. How do you determine if the nudge worked or not?
5. If it does not work how do you respond?
6. Have there been moments where you provided a nudge that the participants did not want?
For hints provided when the players ask for them
7. How often do participants request hints?
8. In what types of situations do participants typically request hints?
9. How do the participants typically respond to the hints?
10. How do you determine if the hint worked or not?
11. How often would you say you provide nudges unprompted vs. provide requested hints?
Part 2: Regarding how the assistance is phrased and delivered
1. When giving a hint or nudge, how do you phrase it?
2. What, in terms of the task at hand, do you include in the hint or nudge?
3. Is there anything specific that you do not include or try not to say?
4. When deciding how to structure the hint, what, about the players, do you take into account?
5. Have you ever encountered a scenario where you unintentionally gave players the answer?
6. Can you think of a recent scenario where you gave a hint you thought was very successful?
7. If a hint or nudge does not work, how do you respond?
Part 3: Regarding the type of assistance
1. What types of assistance do you typically give?
2. What do you take into account when you decide what type of assistance to provide?
3. Do you typically stick to one kind of assistance or does it change?

The first and second authors then independently coded a separate 30% of the transcripts with these themes to assess their Inter-Rater Reliability (IRR) via Cohen's Kappa [16]. Once again, the themes were applied to the empty cells to the right of the utterance in the spreadsheet. The IRR calculation was then performed manually by the first author. The resulting Kappa value was .64, indicating moderate agreement [53]. The two authors discussed disagreements and identified several conceptual overlaps, resulting in 14 themes. The researchers then performed a second IRR check with the updated themes, resulting in a Kappa of .798, a strong level of agreement [53]. The first author then coded the remaining data and performed a final iteration on the themes, collapsing and re-organizing them based on the number of appearances and remaining conceptual overlaps. The result is 10 themes, which were then discussed among

all authors and organized into three conceptual categories. This final list of themes can be seen in Table 3 and is discussed in the next section.

4 Results

The single most common concept that emerged from our interview data was the importance of adaptation. GMs had to be ready to adapt their entire intervention strategy and to adapt individual aids on the fly, depending on what players needed. Our analysis, further, revealed ten themes concerning GMs' best practices in terms of what they look for and how they intervene when they are engaging with this adaptive aid process. We organized these into three interrelated categories based on common concepts, which we present below: Progression, Experience, and Delivery. An overview can be seen in Table 3.

4.1 Theme Category One: Progression

The first category is *Progression*. We define *progression*, here, as a focus on the observed player progress, or their advancement through a task or the entire room. This category contains three themes related to the ways in which GMs would observe progress, players' awareness of their own progress, or how interventions could prompt or guide progress.

4.1.1 Theme One: Progression is both an indicator that aid is needed and that it was successful. GMs revealed that sometimes players would simply stop making progress and would either explicitly ask for help or require the GM to interject: *"those are the people who will...sort of spend 2 or 3 minutes kind of staring at things or staring at nothing and then go, 'what are we supposed to do here?'"* (GM 6). While this example demonstrates a case where progress stops before it even begins, others discussed cases where players would stop making progress in the middle of a puzzle, typically because they were unable to determine the next step: *"so there's steps to get...from one process to the next. And, if they're wandering around and, you know, if the next step they need is basically just to, you know, open a drawer or, you know, look somewhere that they haven't looked yet, then, I will...wait 2 or 3 minutes and then give them a clue"* (GM 11). GM 11's quote highlights the need to wait before chiming in, as players' progress may not be stalled, and they may just be thinking.

GMs suggested that tracking progress was important, not only to see when help is needed, but also to make the correct judgment in what help to give: *"I have a couple times assumed that a group was further along in a puzzle than they actually were...and I sent help for where I thought they were in the puzzle, which included maybe an object that they hadn't picked up or something they hadn't noticed yet"* (GM 8). Additionally, almost all GMs revealed that they knew their aid worked if the players made progress afterward. For example: *"if they actually solve the puzzle. Yeah. If they solve it, and continue to the next. Because the puzzles in our rooms, it's like a series of puzzles. They have to start on the first one, then the second, then the third. Like this...So once they actually solve the puzzle and continue to the next one, I can hear it, see it actually, the hint working"* (GM 1). Some GMs suggested that the speed with which they can turn the help into progress is also an indicator of how good a hint is: *"I sent it to them and they immediately got it. And I felt amazing, and I immediately added it to our prefabricated hints, so that it can be reused as many times as we want, because it just worked that well"* (GM 8). This theme and these quotes highlight how observable progress is an easily measured benchmark for GMs to determine the state of players and when they are ready for aid. They also emphasize the need to pay close and constant attention to players' actions to ensure that aid is provided at the right time and in the right way.

4.1.2 Theme Two: The sensation of a time crunch will prompt aid. Building upon progress on its own, GMs also revealed that progress could be measured in relation to time. Most escape rooms provide players with one hour to solve all the puzzles and "escape" the room (with some variants offering more time), and in their interviews, GMs revealed that both they and players were cognizant of this time crunch. In some cases, players who were tracking their time remaining would begin to request more hints if they felt they were falling behind: *"I think that we generally saw...a pickup in requests for hints as there was a time crunch. So once it's getting closer to the end, I would say that's the only time where I would say, situationally, more hints would be requested"* (GM 7). Similarly, several GMs described tracking time, and what players' progress should look like as time ticked down to decide whether and when to interject: *"I know, like, okay, if they haven't gotten this in like five minutes or I know they need to be in to the next puzzle in like the next ten minutes, like I can kind of use that as, as a way to gauge, like, okay, they really need a nudge now because if I wait, they're not going to get out"* (GM 4) and *"we interfere here and we send the hints to help them even...even if they don't like it, because they took a long time and we need them to...to explore the room more"* (GM 1). The emphasis on time was motivated not only by players' desire to succeed but also by GMs' desire to have players succeed and ultimately have a good time: *"We want them to experience the room, in our escape rooms. So we're not necessarily we don't want to sit there and just watch them, like, not get anything for 40 minutes"* (GM 4) and "*you want to put them in the right direction to sort of avoid frustration*" (GM 13). However, GMs also revealed that, despite their desire to help players succeed, some did not want hints, no matter how far behind they may be: *"it is wild to see people lose because they have decided that interacting with me is a concession, that they will not make for themselves"* (GM 6) and *"then there's groups that will die before receiving a hint. And it doesn't matter if they are going to succeed or fail, they will not take anything. No matter how many times they're prompted, they will not accept or ask for a hint"* (GM 7). These quotes reveal the significant role that time plays in players' and GMs' perceptions of progress and the need to give or receive aid and raise questions about what may happen in an environment that does not have a time limit.

4.1.3 Theme Three: Aid can direct players' attention or engagement when it is insufficient or incorrect. Similar to the first theme, GMs also discussed the need to observe and track players' focus and engagement. Specifically, if players begin to drift away from the puzzle they are working on or become overly focused on unrelated parts of the room, it is an indicator that they may need help: *"it could be that they are spending a lot of time on something that is not going to help them at all. Just really digging a rabbit hole down there and they need a nudge to get away from that into something that maybe they thought of before or something that is like next to whatever they're working on"* (GM 13). Unlike the first theme,

Table 3: The analysis resulted in ten themes, which we organize into three interrelated higher-level categories based on common concepts and present alongside the number of times each theme appeared in the data and the number of GMs who mentioned it.

# of Obs	*# of GMs*	*Theme*
		Theme Category 1: Progression
39	13	Progression is both an indicator that aid is needed and that it was successful
17	10	The sensation of a time crunch will prompt aid
45	14	Aid can direct players' attention or engagement when it is insufficient or incorrect
		Theme Category 2: Experience
46	13	Emotions determine when aid is needed and whether or not it was appropriate
13	8	Aid is diegetic, and sometimes the requests are too
81	14	Player characteristics and differences will influence when/how aid occurs
		Theme Category 3: Delivery
34	13	Aid is incremental, scaffolded, and elaborates as needed
19	10	Aid should not include the answer, unless it must
17	8	Aid should make the players feel like they came up with the solution themselves
22	10	Players should have agency in when/how they receive aid

here, GMs are specifically watching players' behaviors in terms of what puzzles they are attempting to solve and whether or not they become distracted by unrelated elements. In some cases, it seemed like this was caused by set dressing in the room that confused players: *"there are elements of the the space that will be designed in a way that has more potential for red herring...the players would just start reading books, and in that game, it's not about the books"* (GM 6) or because players did not realize the pieces they were interacting with were not part of the same puzzle or task: *"if they are combining information from two separate puzzles that isn't intended to be mixed, then I'll say like this information over here, just ignore that for now"* (GM 10).

For these reasons, GMs described directing players' attention usually to aspects of the room they may have overlooked. For example, GM 2 described directing players with, *"Hey, guys, you have 15 minutes remaining. Did you notice that clock on the wall?"*, illustrating how directing attention also relates to a time crunch. GMs further discussed how players would often miss one piece of a puzzle, either not noticing an important piece or not looking at it closely enough: *"sometimes...you just have to tell them that something exists...if they just looked at the front of a chest closer, that's all they need to know"* (GM 4) and *"how do you make them realize that they haven't...done the thing you're telling them to do? Sometimes there are specific words that we can like, highlight and make bigger so that they reread it"* (GM 11). This illustrates how GMs track and direct players' attention to help them reach the answers on their own, rather than provide an answer directly. This desire influenced other themes, which we discuss further below.

4.2 Theme Category Two: Experience

The second category is *Experience*. We define *experience*, here, as a holistic combination of sensations, emotions, and perceptions players may have within the room, as inferred by the GMs. This category contains three themes related to the complicated interaction between the players' inferred experience and the GMs' decisions and actions, and how the two inform each other.

4.2.1 Theme Four: Emotions determine when aid is needed and whether or not it was appropriate. The first theme in this category deals with the explicit observation of emotional states. GMs discussed how negative emotional states such as frustration or lack of interest were good cues for them to interject. They further revealed that these emotional states could be inferred from gesture or movement: *"there's something that I describe as the hands. It's the kind of like tossing up of the hands, they don't know what's going on anymore"* (GM 8) or, more subtly, from a change in behavior or tone in the room: *"so people would just not be talking and they would just be staring at something or kind of not knowing what to say. So if it was a very quiet group, at a certain point, then it was a good indicator that they're completely stuck"* (GM 7). Additionally, GMs also emphasized the importance of listening to how players speak to each other since verbal cues could also indicate negative emotional states: *"If they start saying 'no, I already tried that' in a terse tone, or say things like, 'here, give it to me, you're doing it wrong'"* (GM 5). All GMs seemed to agree that such negative emotional states were a sign that aid was needed, and they would either chime in with a nudge or remind the participants that they could ask for a hint if they needed it. One participant, GM 3, even suggested that interjecting with a nudge or a hint was a strategy they used to break up an argument or fight: *"if it seems like they're going to argue with each other, we'll do interrupts with hints"*.

Emotions, however, were not only an indicator of when aid was needed but also a way of evaluating whether it was appropriate to the players' desires and needs. When asked how players typically responded to help, GMs described a variety of positive and negative emotional responses. Several GMs described players making self-deprecating comments either seriously or jokingly: *"Sometimes it's the self-deprecation of, 'oh, my God, I'm so stupid, I should have seen that'. Weirder still is sometimes they'll bust their friends like 'you idiots! Why didn't you see that?' Says the person who also didn't see that"* (GM 6) and *"typically they would laugh and say, 'oh my gosh, we are so dumb'. That phrase is the one that I heard most in my entire time working there"* (GM 5). Other GMs described players reacting

to the help with joy or laughter: *"they sometimes laugh. It's like a pretty good response"* (GM 1) and *"we were literally having a group high five...they were so excited"* (GM 9). Others described behavioral reactions such as increased motivation: *"if you've done a really good job and given a really good clue or hints, there's the, you know, I got it and, you know, and they're excited again. And the energy's back up"* (GM 4). Together, these examples illustrate how GMs often infer emotions from subtle indicators and learn to make the correct choices in response.

4.2.2 Theme Five: Aid is diegetic, and sometimes the requests are too. Escape rooms typically feature a theme or story such as "the Old West", "secret agents", or "spaceships". When asked if the theme played a role in how help was administered, many GMs discussed how nudges and hints are often diegetic, meaning they fit cohesively within the theme or narrative. For example: *"it was not intended to be a person on the other side of the screen...people would ask for help by saying, 'computer, we need assistance.' And it was supposed to be like, oh, we've got this sentient top of the line AI. And so sometimes I would send things like 'my database indicates that, Romans 623 is a Bible verse'"* (GM 8) and *"we try to provide hints, like...in a manner where it's something...probably the Mad Hatter would say"* (GM 3, who worked in an *Alice in Wonderland* themed room). One GM, GM 6, even revealed that, sometimes, the players themselves would also get caught up in the theme of the room, and the way they would ask for hints would become diegetic: *"Often when they ask for hints, they are asking in an immersed kind of way. So they'll say, 'hey [character name], have you ever seen a book like this before?'"*. This theme highlights the narrative experience within the context of the escape room and an attempt on the parts of all involved to maintain the consistency and immersion of that experience.

4.2.3 Theme Six: Player characteristics and differences will influence when/how aid occurs. This theme, which is the most prominent in the dataset as seen in Table 3, encapsulates the central idea from all GMs that no two groups of players would require the same aid at the same moment in time. GMs further clarified that aid had to be customized or adapted based on what players were experiencing within the room. Age was a commonly mentioned determinant factor, with almost all GMs providing help differently to groups of children versus groups of adults: *"there are some teams of students who will come in, and students generally, it seems like, don't care as much and so they'll ask for a lot more clues...but I guess if there's like a team of people college aged or in their 20s who are playing, they'd probably ask for less"* (GM 11) and *"I find it's people like their 20s to 30s are usually the most prideful when it comes to asking for help, because they think they should be able to do it on their own"* (GM 10). Beyond just the timing of help, how it is phrased may also be impacted: *"with adults, I'll send something a little more vague, and with little kids, it's very obvious, like, 'go over to this thing'"* (GM 8). Other characteristics determining when and how aid was administered included competitiveness: *"I'm less likely to give a group of, like, all men who seem to be very competitive, a nudge. But if it's a group that's just there to have fun..."* (GM 11), past experience with escape rooms: *"If a group was open to receiving extra hints, it was usually groups who were brand new to escape rooms..."* (GM 5), or experience with similar games: *"if they're wearing a Dungeons and Dragons shirt or a video game shirt...you kind of know that they're probably going to be okay"* (GM 4).

This theme highlights the significance of individual differences when administering aid to players. Notably, however, it is not just the differences that define players outside of the room but also the differences that emerge between groups as they progress through their experience, which require the GMs to judge and adapt to the situation dynamically.

4.3 Theme Category Three: Delivery

The third category is *Delivery*. We define *delivery*, here, as the considerations and dynamics involved in the communication of aid by GMs and the subsequent inferred perceptions players have of that aid. This category is distinct from the previous as it focuses solely on the provided aid, whereas the previous category considers the entire escape room environment. This category contains four themes related to how aid is structured and how GMs try to ensure certain perceptions from the players towards the aid they give.

4.3.1 Theme Seven: Aid is incremental, scaffolded, and elaborated as needed. GMs described dynamically adapting to players, and help was almost never a "one and done" situation but rather an incremental process of starting vague and becoming more direct or specific based on how players responded. For example: *"If I send them a nudge and they read it and maybe they don't quite know how to apply that information, I might send something a little more clear or something for another part of the puzzle that when they have both pieces, they can connect it and come to the conclusion"* (GM 8). This was partially influenced by the desire to let players reach the answer on their own, which meant that assistance often began with a vague statement, i.e., *"We try to give them as little as possible...whatever the least we can possibly say that will get them to where they need to be is our goal as a game master"* (GM 9). However, the GMs also revealed that players would frequently not understand such statements or that they would only get them part way to the next step, therefore requiring further aid: *"I might have to take 2 or 3 messages to get you there, because some puzzles are just more complicated"* (GM 4). GMs also revealed that they had to think quickly about what caused miscommunications or confusion so they could accurately adapt the aid to the players: *"we'll have to kind of cater the clues to the person. So if they're not getting it right away, I have to then figure out, okay, well, was it the wording or was it...some of them aren't English speakers for their first language. Some of them speak other languages like French, Mandarin, Spanish, we've even had a team from Italy come and play, so that might also be why. So sometimes we might translate the clues for them to help them"* (GM 12). This theme highlights the challenge in determining what element of help may not be working and how GMs must think about adapting assistance to the players, similar to how an educator may adapt their teaching style to the needs of a struggling student.

4.3.2 Theme Eight: Aid should not include the answer unless it must. When asked what they try to avoid including in their assistance, most GMs explicitly stated that they did not want to give away the answer. For example: *"obviously we try to avoid giving like the, the straight, straight up...answer"* (GM 3), *"I would try not to give the complete answer"* (GM 5), and *"we just don't give them an*

answer...we don't tell them codes" (GM 2). This trend is unsurprising, as giving the answer away would likely ruin the fun of the game. However, several participants revealed situations where they would give players the answer. In some cases, this happened if the GM had given multiple clues to the players, but the players still did not understand the solution, suggesting limits to the scaffolding revealed by the prior theme. For example: *"If you have to give them the answer, if you're out of clues and the last thing you have is the answer and they say, 'yes, I would like the answer' I would say they probably didn't understand the puzzle"* (GM 12). In other cases, the GM would give the answer if the players had effectively solved the puzzle but were stuck on a small or superfluous detail: *"I would just tell them, I would just tell them the answer in most cases. If they have gotten like, most of the numbers or they've gotten the order wrong, but like most of the numbers, I would just tell them like, this is correct"* (GM 11). Notably, however, GM 11 did go on to mention that participants would not always accept the answer: *"sometimes they'll say like, 'not yet' and they'll wait and try it a few more minutes and then ask for it"* suggesting that players had the agency to accept the answer on their terms. In all cases, when GMs discuss ultimately giving the answer, it comes from wanting the groups to succeed and enjoy the room, and therefore, a desire to just get them past the roadblock in their way.

4.3.3 Theme Nine: Aid should make the players feel like they came up with the solution themselves. Several GMs highlighted how the players' sense of competence was central to having a good experience. As GM 2 put it *"a big thing about escape rooms is...some people don't like them because they feel like they make them feel stupid"*. This statement highlights the reason that emotional responses to help can be so varied or even negative: players want to feel smart, they want to figure it out on their own. To help players feel smart, most of the GMs would phrase their assistance in the form of a question: *"our hints are always given in the form of questions, and I think that's very important...so...rather than just kind of ignoring everything they've done so far and saying 'this is what you got to do instead,' trying to gently steer them in the correct direction by asking questions so they can still come to those conclusions themselves"* (GM 10) and *"I usually form it as a question, because then I guess they saw it as, 'okay, she's asking a question that's obviously supposed to get me to think about this,' but then, when they think about, it's like, 'oh, yes, I answered this question then I got it,' and that gave them the confidence to continue on. It made them feel less dumb"* (GM 5). GM 6, in particular, described how rewarding it was for all involved for the players to feel like they had reached the answer on their own: *"in telling me what they know, sometimes, just the process of explaining it out loud to another person, they will come up with the solution and they'll go, 'oh my God, thank you' and I'll go, 'cool! I did literally nothing.' I mean, what a win that is, right? Because that was entirely their idea"*. Ultimately, GMs believe making the players feel smart helps to keep them in a positive mood, more comfortable, and also more immersed in and engaged with the environment.

4.3.4 Theme Ten: Players should have agency in when/how they receive aid. When asked if they gave players nudges or only hints, all GMs stated, to some extent, that they asked the players what they wanted. To illustrate: *"we give all of our players an option as to the kind of level of hint giving they want. Some players will, of course, select 'only if they ask'. Others, though, will say, 'yeah, it's okay if you kind of jump in or, you know, at least ask before you give a hint', something like that"* (GM 4). In some cases, GMs would clarify that they would *always* wait for the player to request help before providing it: *"we might just poke our heads in and check in and be like, 'how's it going? Everything still going well?' And then it's kind of...providing them the opportunity to ask for a hint, but not just jumping and being like, 'here's a hint'"* (GM 10). GMs emphasized player choice as important for ensuring the players enjoyed themselves and got the desired experience out of the room. Several GMs disclosed situations where, when defying player agency and giving a hint when it was not wanted, they were met with animosity. GM 6 had experience with this situation: *"when I give them a nudge, it's sort of the ego 'we need to win this without needing any hints'...Any time players are antagonistic towards me or treat me as the antagonist, I try to give them lots of physical space...I will not interrupt them"*. Other GMs also described allowing players more agency in such situations: *"if we see that the guest, you know, didn't appreciate the additional information, we'll...let them struggle a little bit more on things"* (GM 9). Together, these examples highlight the significance of honoring the agency of a player.

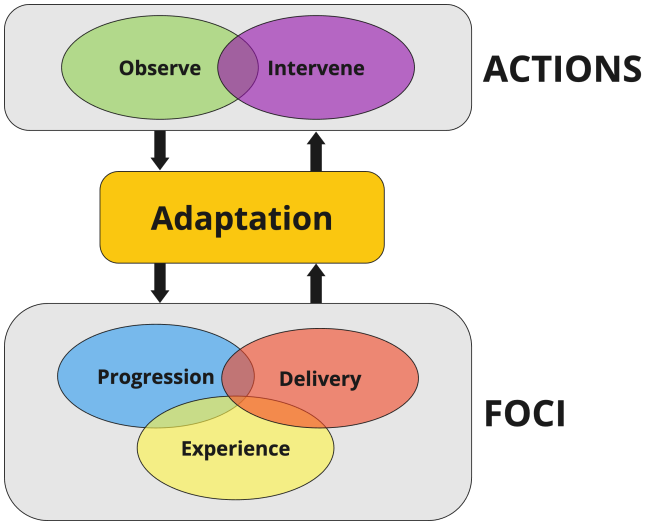


Figure 1: From the results, we can recognize a model of how the GMs provide aid, which can inform reflection design. The model presents the GM as an actor capable of two actions (observing or intervening), with adaptation acting as a mediator in that they must both adapt their interventions and adapt what they watch for and how they interpret it, and focused on three foci, in terms of what they observe or interact with (progression, experience, and delivery).

5 Discussion

XRLEs' immersive nature sets them apart from other digital learning environments, and makes it difficult to adopt best practices from those domains. Because escape rooms are highly immersive yet cognitively demanding like XRLEs, the insights from GMs' best practices are informative for improving reflective opportunities in

XRLEs, perhaps moreso than insights from other digital learning environments. To date, most research on reflection in XRLEs utilizes activities outside the immersive environment, such as having students complete an exercise after using a VR experience [76]. However, this work presents exploratory considerations for embedding reflection within XRLEs to support learning throughout the experience. The model we present in Figure 1 formalizes how GMs provide aid in escape room environments by leveraging the available actions of observing and intervening and focusing on the foci of progression, experience, and delivery, all while constantly adapting this process based on their observations and the needs of the players.

In this section, we discuss these concepts further in terms of how they connect our results to insights for XRLEs, and we summarize nine actionable insights in Table 4. We note, however, that XRLEs are not inherently the same as escape rooms. While escape rooms are typically (though not always) built for entertainment and group play, and feature a live human overseer, XRLEs are meant for education (and may not be entertaining in any capacity), are more often meant for individual interaction, and may not feature any real-time human oversight. Given these differences, we acknowledge that our results and the directions we present below may be limited to certain types of XRLEs or need to be adapted based on the goals and design of the XRLE. For example, while our participants stressed that sensations of a time crunch prompted aid, it would be more beneficial for learners in an XRLE to slow down and think rather than speed up. We emphasize, instead, that the similarities between both mediums (featuring spatial, immersive interactive experiences) mean that XRLEs, inherently, have the *potential* to support these ideas, pending how they are designed, and currently, some XRLEs do feature more playful experiences [4] or are meant to be completed in a group context [34]. Our goal is to set the signposts for further research and development for XRLE-based reflection, and we make no claims about the impact on learning based on these results alone.

We also envision the ideas we present here as being informative to both human-mediated (i.e., an instructor) and technology-mediated (i.e., an AI agent) reflection. Currently, many XRLEs are deployed in contexts where an instructor or educator oversees the experience [41, 105] and in such contexts, it would be possible for these overseers to leverage the considerations we present to provide reflective, or better reflective, opportunities to their students. On the topic of technology-mediated reflection, we acknowledge that the GMs' insights, which reflect distinctly human instincts, may not be feasible for current AI to replicate perfectly. Nevertheless, AI has excelled at tracking and evaluating learner progress for quite some time [55], and recent AI is getting better at detecting more subtle indicators, such as emotions [64]. LLMs, known to be capable of human-like communication, likewise show great promise in prompting reflection in human-like ways [35, 73, 77]. While some of what the GMs discuss cannot be replicated technologically at this time, we anticipate that future development of this technology will produce systems capable of performing the required tasks to oversee reflection in XRLEs and present our considerations as guidance for such development. Future work can also explore how other forms of technology, such as biosensors for detecting focus or emotions [23], can augment this process further.

5.1 Progression

Three themes related to progress through a task or environment fell under the category of "progression." Within this category, themes were focused on observing and interpreting players' completion of tasks. When this observation resulted in a perception that progression was insufficient or unfocused, aid would be offered (or requested). Similarly, observations of progression continuing or focus returning after the aid were indicators that it was successful. This aligns with prior research on how awareness of one's progress or performance [11, 24], and especially awareness in the context of others [45, 63], can prompt individuals to want to perform better in both lifestyle [6] and learning [30] activities. Additionally, such information can help learners think about their approach and what aspects of a problem require their focus [45]. As a result, data dashboards or similar tools that present one's score or progress in the context of the community are commonly integrated into technology to motivate users [96]. Open learner models and learning analytics dashboards, for example, frequently present students with such information [8].

The focus of GMs on progression is informative for how to design reflection activities to occur in response to the learner's achievement on the tasks in the immersive environment rather than in a standardized manner. When learners struggle on a task, they can be prompted to pause and assess their strategies and what they have learned so far to devise a more successful strategy. This approach is likely helpful for learners in XRLEs who may be overwhelmed with distracting stimuli or, as the GMs suggest, aspects of a problem that are not helping them progress. On the other hand, reflection prompting can come when learners successfully progress through the tasks in an XRLE to help learners consolidate what they learned and help them devise deeper structures of their problem-solving strategies that go beyond the surface features specific to the environment. Aligning with the GMs, instructors could monitor students' progress to understand whether they are adequately progressing through the tasks in an XRLE and intervene in response to similar cues. Alternatively, XRLEs could be built with automated systems that assess data on the learner's actions or biometric data to determine when a learner needs to be redirected or should reflect on their success, and an agent within the XRLE could engage the learner in reflection when needed. As the GMs describe, the learner's response to these prompts would also help an instructor or AI agent to assess if the prompting was successful based on their subsequent progression.

This approach that adapts to learners' progression is a departure from current methods of reflection prompting in XRLEs that provide standardized times or breaks for reflection, such as segmenting a VR experience at pre-determined time points [76, 79, 80]. Further, this approach allows learners to reflect in action rather than reflect on their actions after completing a task or XR experience [66]. While time is the most universal manner of measuring progress (i.e., how long students spend on individual tasks), our results provide insights into other, subtler cues that these instructors or automated systems could look for to measure progress. Relative to other digital learning environments, progression in XRLEs is a unique case as their immersive and spatial nature means progress may be subtle, observable from, for example, the learners' attention

Table 4: From our discussion of the escape room GMs' best practices, we synthesize nine actionable implications for the design of reflection in XRLEs.

Construct	***Implications***
Progression Progression is an important indicator for when escape room players need aid and also when learners need to stop and reflect, whether they are struggling to complete a task, are off task, or are being successful and should reflect on what they are doing well to uncover strategies.	Reflective prompting should respond to whether a player is progressing or stalling in an XRLE and intervene when necessary. Consistent monitoring of progress will help determine if further reflection is needed or the intervention was successful to move the players along. Reflective prompting can help steer learners who are not making progress due to being overwhelmed by distractions in XRLEs. Reflection can help them think through more effective strategies or review what they have learned so far to point them in a more productive direction. Reflection activities can also support learners when they are successful and making progress by asking them to think through the strategies they have used and what they have learned.
Experience Hinting in escape rooms aims to enhance the player's experience, and reflective prompting should do the same by embedding it in the XRLE world and adapting the learner's affective state and characteristics.	Reflective prompting and activities should be native to the story, for example, coming from a character aligned to the theme or a task that fits into the narrative. Reflective prompts should respond to the emotional state of the learner, adapting when to intervene and what prompts to use based on whether a learner is frustrated, bored, or excited. Reflection design and prompts are not one-size-fits all but need to be tailored to learners' characteristics like age, literacy, and education level.
Delivery Escape room GMs structure their hints and nudges to best support players without giving away the solutions. Reflecting in XRLEs aims to do the same, helping learners arrive at their own solutions.	Reflective prompts should provide scaffolding, starting with broad or vague questions and continuing to more narrowly direct attention only for those learners who need it. Reflective prompts should be delivered in a way that emphasizes the learner's competence and their ability to complete the tasks themselves, thereby increasing intrinsic motivation. Learners should have some degree of agency in how often or when they receive support and prompts, which will also contribute to their intrinsic motivation. However, designers should still ensure every learner engages in sufficient reflection regardless of whether they think they need it.

but not necessarily from explicit interaction with the environment. In this capacity, they are closer to escape rooms than most other digital learning environments, and observation is more key here than in other contexts. With regards to observing when progression had halted, our GMs described cues such as low energy levels (i.e., everyone slows down or becomes less animated), lack of focus on a single task and/or random actions, and a tendency to "just stand there". All of these cues could be directly observable in an XRLE, and new technology is experimenting with detecting these cues, like attention, via algorithmic or biometric means [83, 93]. Such technology in XRLEs can be implemented to power an automated system or to support human judgment.

5.2 Experience

Under the category of "experience" were three themes related to players' emotions, sensations, and perceptions. Within this category, themes were focused on how to maintain a desired experience and what to observe to ensure that this maintenance process was adapted correctly.

The most frequently occurring theme in our dataset was the idea that aid had to adapt in both delivery and timing to the players, who came to the room as different people and changed throughout play, in order to ensure that everyone would receive a desirable experience. In contrast, XRLEs typically feature the same type of reflective prompt for all learners [76]. While this is largely the result of technical limitations, and is starting to see change with the advent of new AI technology [70], the fact remains that prompts

could be customized better. Further, existing reflective designs do not typically account for the fact that learners change throughout their experience, despite the literature discussing this point [94]. One aspect that changes frequently throughout any experience is emotions. Almost all GMs discussed observing behavioral and emotional cues to determine when it was a good time to provide aid. These best practices of GMs have several implications for the design of reflective activities in XRLEs. Similar information can be utilized to understand when and how to intervene with reflection prompting as a GM's nudging practices in escape rooms. Asking learners to reflect when they feel overwhelmed or frustrated will help them manage their experience and maintain intrinsic motivation for learning within an XRLE.

Similar to escape room players, learners in an XRLE may exhibit complex, and potentially negative, emotions. Frustration, for example, may occur if they are unable to properly complete a learning task due to misunderstanding information [59]. Prior work showed how frustration is a good opportunity for reflection [69] and how reflection is an effective practice for overcoming frustration that would otherwise interfere with learning. However, prior work focused on reflecting on frustration afterward and has not considered how frustration could indicate opportunities for *in-action* reflection, i.e., reflection at the moment, rather than after the fact [88]. In other words, based on the insights from the GMs, instructors overseeing XRLEs should consider prompting in-action reflection when frustration is detected . This would complement the ways reflective prompting adapts to a learner's progression, adding a focus on the affective dimensions of learning to the personalization of reflection prompting. The reflection prompts and frequency should also be tailored to the target learner population. How and how often an instructor or AI agent may interject should depend on the age and education level of the learners, as their characteristics will determine how much support they may need, what vocabulary is appropriate, and even what format the reflection should take depending on their reading and writing abilities.

With advancements in AI [70] and immersive technologies to sense affective states like frustration through biometric data, it is additionally possible XRLEs could include automated reflective prompting based on learners' emotions. In fact, because XRLEs permit learners to express and experience a wider range of emotional states, such an approach would be especially beneficial to XRLEs relative to other digital learning environments [52]. For example, emotional gestures such as those described by our participants would be observable in an XRLE. This does, of course, raise concerns regarding the safe and ethical use of such data, which would likely need to be discarded after each individual session, raising further questions about efficient data pipelines.

Another important experiential dimension for escape rooms is how the GMs' support is embedded within the room's theme or narrative. In other words, it was *diegetic* to the world to ensure immersion is not broken when players interact with the GM's prompts and hints. While it may sound obvious that an immersive environment should have similarly immersive reflection opportunities, surprisingly, many XRLEs do not present their reflective prompts in a diegetic (originating from within the story world) manner. Instead, reflection is prompted between tasks [76] or after students complete the environment entirely [66] and are characterized as completely separated from the XRLE, its setting, and its story. These reflective activities that take the learner outside of the XRLE have been shown to increase learning from these environments; however, integrating practices that enhance the learner's experience is likely to amplify the impact of reflection on learning. This is a very informative design implication for XRLEs. Essentially, the reflection prompts should lie within the XR world, potentially interacting with a character in the story who asks the learner to reflect or integrate the reflections within the tasks they are completing in the XR world. This will help learners maintain their immersion in the world while engaging in reflection, which can be orchestrated by an instructor in a manner in which an escape room GM would provide hints or via an automated system within the XRLE. By prioritizing the act of reflection as part of the experience and presenting it in a way that adapts to the affective state of the learner, we may enhance the experience of the learner, intrinsically motivating them to engage in reflection and learning activities.

5.3 Delivery

Finally, the category of "delivery" encompassed four themes related to how aid was delivered or how players were meant to perceive or understand the aid they received. These themes emphasize the need to structure what hints and nudges are given incrementally so that they support players to come up with answers on their own and give the player agency in how and whether they are nudged. These themes have several implications for XRLEs as well. The way GMs described wanting their hints and nudges to support the player in uncovering answers to the escape room themselves aligns with the approach to prompt reflection rather than give hints or answers. Reflective activities encourage learners to think deeply and perhaps take a different perspective on the information at hand in order to help them solve problems themselves, rather than relying on tools that provide the answers for them.

What we see emerging from the themes in this category is a *scaffolding* [7] process in which GMs begin with vague aid geared towards reflection and self-discovery, but then iterate based on the players' reactions. While scaffolding learning is not a new concept, it does not often see use in the design of reflecting prompting, which typically does not adjust its phrasing based on how learners respond. This means that, in some cases, reflections may be insufficient for learning outcomes [45]. Our results suggest that scaffolded reflection could benefit XRLEs, and likely digital learning as a whole. Based on these themes, reflective prompting in XRLEs should start with broad and perhaps vague reflective questions, followed by more specific and directed probes for the learners who need to be directed to specific information. These prompts should help build a sense of competency among learners as they support the feeling that they came up with the solutions themselves rather than being told the answer or having their errors emphasized. This would, in turn, enhance both immersive learning and self-discovery [3, 40]. Along with giving the learner agency over how they are supported, this sense of competence will foster intrinsically motivated learning and ensure learners are not simply focused on gaining extrinsic awards like points [51]. In addition to human instructors scaffolding their support, intelligent technology could be built to evaluate the quality

of reflection responses and, instead of just moving on, prompt the student differently to try and elicit a more appropriate or beneficial response, similar to how the GMs start with one form of aid and then repeat and iterate until they see an appropriate reaction.

Notably, the GMs also discussed situations in which they would simply give players the answer to keep them from failing. This raises questions about whether reflective systems in XRLEs should consider intervening to prevent failure. Similar to earlier work that explored failure in games [36], prior work in learning and education discussed failure as an opportunity for effective reflection [18] and, as a result, failure is not typically viewed as detrimental to learning; in fact, it is often viewed as necessary and referred to as "productive failure" [37, 106]. However, considerations need to be made, such as, for example, whether or not the students can ultimately reach the solution [59, 84]. This consideration echoes the GMs' sentiment that it is better to give the answer than sentence participants to failure. This may be especially important in immersive environments like escape rooms and XRLEs, as failure may disrupt immersion.

Our results also highlight the need to give agency to players to decide whether and when they want to be prompted and provide a sensation that they determined the solution on their own, almost as if they forget about the aid entirely. XRLE designers should consider ways learners can make decisions about when they want to pause and reflect, or whether they want the system to determine the prompting, an approach to agency that has not been extensively explored in the XRLE context [39, 104]. This would also align with how GMs foster players' sense of competence during problem-solving tasks by making hints into prompts that allow the player to arrive at the answer on their own rather than solving it for them. Together, agency and a sense of competence are powerful ways of increasing intrinsic motivation to learn [86]. In XRLEs learners should be given choices about when and whether to engage in the reflective activities compared to when they want to persist to find the solution on their own. This will be a delicate balance for XRLE designers as learners need to engage in some level of reflection to help them consolidate their learning and develop metacognitive strategies for future learning, and too much agency may allow them to opt out of reflection. Yet, as the GMs describe, they could be allowed to set their preferences for how often they would like intervention to have some level of agency in the process.

These themes around delivery highlight the need to rethink reflection in XRLEs, which currently feature prompts posed to learners in or after tasks with no feedback or interaction given on the reflection itself [76, 79, 80]. Taking a scaffolded and iterative approach requires a degree of back-and-forth between an instructor or an automated system to provide increasingly targeted prompts that help the learner arrive at solutions themselves. Future research should assess the impact of such a reflective method on learning outcomes, but also on the impact it has on learners' sense of competence. Further, research should assess the impact of providing choice on how or how often to reflect while still ensuring learners engage in adequate reflection. This agency over the delivery of reflection and the potential feelings of competence that result should be studied for their impact on the learner's intrinsic motivation.

5.4 Limitations

Here, we briefly acknowledge the limitations of this work. Primarily, we acknowledge that we interviewed 13 GMs for this study. While we did see saturation in the data, and this number is consistent with common HCI standards [27, 44], a larger sample size may yield additional insights. To better understand GMs' best practices more broadly, we aim, in the future, to expand this work with a survey study that can reach a wider audience and build a stronger foundation of knowledge. Through such work, we can also overcome the limitation born from our decision to only recruit from the United States. Problem solving practices may differ in different cultures, not only GM practices, and accordingly, our results may not generalize beyond North American contexts.

Going beyond the sample, future work can also better distinguish the types of escape rooms (e.g., live actor vs. not, mixed reality, puzzle-focused vs. physical-focused) to see how the style of delivery impacts GM practices. Additionally, GM training likely impacts their practices, with large, chain escape rooms generating consistent training materials for GMs, but independent or local escape rooms potentially allowing GMs more freedom of style or expression; thus, future work should further examine GM training and GM protocols.

Regarding the research scope, there are two important limitations. First, we focused on learning from GMs only and did not consider escape room players. As such, our work focuses on the "art of aid delivery" (or best practices) rather than inquiring directly with players about how they experience receiving aid. Second, we focused on inferring best practices on how GMs provide hints and nudges and did not validate their suggestions, specifically on how they stimulate reflection and, consequently, impact learning. Leveraging our exploratory work as a foundation, future work can examine the experiences of players (in escape rooms) and learners (in XRLEs), as well as conduct experiments to validate our ideas and their potential impacts on learning.

We additionally acknowledge the limitations of an interview methodology: what GMs state they are doing might differ from what they do. We suggest that future work consider observational work to address this. Importantly, such observational work can enable to identify the spatial-immersive properties better that correspond with the need for and success of reflection. Finally, some of our findings may generalize to XRLEs better than others due to the differences in design and intent between escape rooms and XRLEs. On the other hand, some findings may generalize to other domains beyond XRLEs, but that exploration is beyond the scope of this work.

6 Conclusion

Extended reality learning environments (XRLEs) have the potential to immerse students in experiences they would otherwise not be able to access. Prior work, however, found that, without reflection, these environments do not result in effective learning gains. Further, reflection is difficult to implement in XRLEs due to the immersive nature of the experiences. In this work, we argued that we may be able to glean insights into better implementation of reflection in XRLEs from the best practices of escape room game masters (GMs). From an interview study with 13 GMs, we identify 10 themes that capture their techniques for timing and structuring aid that

fall within three foci: progression, experience, and delivery, which inform their techniques. We discuss each of these foci and identify new ways to envision learning and reflection in XRLEs, thereby highlighting innovative design opportunities.

Acknowledgments

We would like to thank the escape rooms that shared our recruitment emails with their game masters and all the participants involved in this work. This work was supported by the National Science Foundation (NSF) under Grant #2418612. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Benjamin M Abdel-Karim, Nicolas Pfeuffer, K Valerie Carl, and Oliver Hinz. 2023. How AI-Based Systems Can Induce Reflections: The Case of AI-Augmented Diagnostic Work. *MIS Quarterly* 4 (2023).

[2] Vassilis Agouridas and Phil Race. 2007. Enhancing knowledge management in design education through systematic reflection practice. *Concurrent Engineering* 15, 1 (2007), 63–76.

[3] AA Akanbi and CB Kolawole. 2014. Effects of guided-discovery and self-learning strategies on senior secondary school students' achievement in biology. *Journal of Education and Leadership development* 6, 1 (2014), 19–42.

[4] Claudia Arcos, Walter Fuertes, César Villacís, Margarita Zambrano, Tatiana Noboa, Ana Tacuri, Hernán Aules, and Theofilos Toulkeridis. 2016. Playful and interactive environment-based augmented reality to stimulate learning of children. In *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 1–6.

[5] Masiar Babazadeh and Manrico Francesco Frigerio. 2021. Enhancing Problem-Solving Skills with Educational Escape Rooms: a Middle School Case Study. In *Proceedings of the European Conference on Games Based Learning*. 53–62.

[6] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 585–594.

[7] Joan Bliss, Mike Askew, and Sheila Macrae. 1996. Effective teaching and learning: Scaffolding revisited. *Oxford review of Education* 22, 1 (1996), 37–61.

[8] Robert Bodily, Judy Kay, Vincent Aleven, Ioana Jivet, Dan Davis, Franceska Xhakaj, and Katrien Verbert. 2018. Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 41–50.

[9] D Boud, R Keogh, and D Walker. 1985. Reflection: Turning Experience into Learning.

[10] Eelco Braad, Nick Degens, and Wijnand A IJsselsteijn. 2020. Designing for metacognition in game-based learning: A qualitative review. *Translational issues in psychological science* 6, 1 (2020), 53.

[11] Robert C Brusso, Karin A Orvis, Kristina N Bauer, and Amanuel G Tekleab. 2012. Interaction among self-efficacy, goal orientation, and unrealistic goal-setting on videogame-based training performance. *Military Psychology* 24, 1 (2012), 1–18.

[12] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research* 42, 3 (2013), 294–320.

[13] Daniel H Chang, Michael Pin-Chuan Lin, Shiva Hajian, and Quincy Q Wang. 2023. Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability* 15, 17 (2023), 12921.

[14] Chih-Yueh Chou, K Robert Lai, Po-Yao Chao, Chung Hsien Lan, and Tsung-Hsin Chen. 2015. Negotiation based adaptive learning sequences: Combining adaptivity and adaptability. *Computers & Education* 88 (2015), 215–226.

[15] Douglas B Clark, Emily E Tanner-Smith, and Stephen S Killingsworth. 2016. Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research* 86, 1 (2016), 79–122.

[16] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.

[17] Nathaniel W Cradit, Jacob Aguinaga, and Caitlin Hayward. 2024. Surveying the (virtual) landscape: A scoping review of XR in postsecondary learning environments. *Education and Information Technologies* 29, 7 (2024), 8057–8077.

[18] Kristina B Dahlin, You-Ta Chuang, and Thomas J Roulet. 2018. Opportunity, motivation, and ability to learn from failures and errors: Review, synthesis, and ways to move forward. *Academy of Management Annals* 12, 1 (2018), 252–277.

[19] Elizabeth A Davis. 2003. Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences* 12, 1 (2003), 91–142.

[20] Daniel de la Flor, José Antonio Calles, Juan J Espada, and Rosalía Rodríguez. 2020. Application of escape lab-room to heat transfer evaluation for chemical engineers. *Education for chemical engineers* 33 (2020), 9–16.

[21] Chris Dede. 2009. Immersive interfaces for engagement and learning. *science* 323, 5910 (2009), 66–69.

[22] Christopher J Dede, Jeffrey Jacobson, and John Richards. 2017. *Introduction: Virtual, augmented, and mixed realities in education*. Springer.

[23] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. 2020. Human emotion recognition: Review of sensors and methods. *Sensors* 20, 3 (2020), 592.

[24] Jose Esteves, Konstantina Valogianni, and Anita Greenhill. 2021. Online social games: The effect of social comparison elements on continuance behaviour. *Information & Management* 58, 4 (2021), 103452.

[25] Panagiotis Fotaris and Theodoros Mastoras. 2019. Escape rooms for learning: A systematic review. In *Proceedings of the European Conference on Games Based Learning*. 235–243.

[26] Greg Guest, Emily Namey, and Mario Chen. 2020. A simple method to assess and report thematic saturation in qualitative research. *PloS one* 15, 5 (2020), e0232076.

[27] Nour Halabi, Günter Wallner, and Pejman Mirza-Babaei. 2019. Assessing the impact of visual design on the interpretation of aggregated playtesting data visualization. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 639–650.

[28] David Hamilton, Jim McKechnie, Edward Edgerton, and Claire Wilson. 2021. Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education* 8, 1 (2021), 1–32.

[29] Casper Harteveld, Erica Kleinman, Paola Rizzo, Dylan Schouten, Truong Huy Nguyen, Samuel Liberty, Wade Kimbrough, Paul Fombelle, and Magy Seif El-Nasr. 2019. Teamwork and adaptation in games (TAG) a survey to gauge teamwork. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–12.

[30] Danial Hooshyar, Margus Pedaste, Katrin Saks, Äli Leijen, Emanuele Bardone, and Minhong Wang. 2020. Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & Education* 154 (2020), 103878.

[31] I-Han Hsiao, Julio Guerra, Denis Parra, Fedor Bakalov, Birgitta König-Ries, and Peter Brusilovsky. 2012. Comparative social visualization for personalized e-learning. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 303–307.

[32] Shih-Yuan Huang, Yi-Han Kuo, and Hsueh-Chih Chen. 2020. Applying digital escape rooms infused with science teaching in elementary school: Learning performance, learning motivation, and problem-solving ability. *Thinking Skills and Creativity* 37 (2020), 100681.

[33] Ananya Ipsita, Levi Erickson, Yangzi Dong, Joey Huang, Alexa K Bushinski, Sraven Saradhi, Ana M Villanueva, Kylie A Peppler, Thomas S Redick, and Karthik Ramani. 2022. Towards modeling of virtual reality welding simulators to promote accessible and scalable training. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–21.

[34] Randolph L Jackson and Eileen Fagan. 2000. Collaboration and learning within immersive virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments*. 83–92.

[35] Mahdi Jelodari, Mohammad Hossein Amirhosseini, and Andrea Giraldez-Hayes. 2023. An AI powered system to enhance self-reflection practice in coaching. *Cognitive Computation and Systems* 5, 4 (2023), 243–254.

[36] Jesper Juul. 2013. *The art of failure: An essay on the pain of playing video games*. MIT press.

[37] Manu Kapur and Katerine Bielaczyc. 2012. Designing for productive failure. *Journal of the Learning Sciences* 21, 1 (2012), 45–83.

[38] Sukran Karaosmanoglu, Sebastian Cmentowski, Lennart E. Nacke, and Frank Steinicke. 2024. Born to Run, Programmed to Play: Mapping the Extended Reality Exergames Landscape. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 309, 28 pages. https://doi.org/10.1145/3613904.3642124

[39] Danielle Keifert, Christine Lee, Maggie Dahn, Randy Illum, David DeLiema, Noel Enyedy, and Joshua Danish. 2017. Agency, embodiment, & affect during play in a mixed-reality learning environment. In *Proceedings of the 2017 conference on interaction design and children*. 268–277.

[40] Bert Y Kersh. 1962. The motivating effect of learning by directed discovery. *Journal of Educational Psychology* 53, 2 (1962), 65.

[41] Tuulikki Keskitalo. 2011. Teachers' conceptions and their approaches to teaching in virtual reality and simulation-based learning environments. *Teachers and Teaching: theory and practice* 17, 1 (2011), 131–147.

[42] Sharina Kimura, Michael Zintl, Claudius Hammann, and Florian Holzapfel. 2024. Simulator-based Mixed Reality eVTOL Pilot Training: The Instructor Operator

Station. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 662, 10 pages. https://doi.org/10.1145/3613904.3642060

[43] Chairi Kiourt, Dimitris Kalles, Aris S Lalos, Nikolaos Papastamatiou, Panayotis Silitziris, Evgenia Paxinou, Helena G Theodoropoulou, Vasilis Zafeiropoulos, Alexandros Papadopoulos, and George Pavlidis. 2020. XRLabs: Extended Reality Interactive Laboratories.. In *CSEDU (1)*. 601–608.

[44] Erica Kleinman, Reza Habibi, Garrett B Powell, Brent Reeves, James Prather, and Magy Seif El-Nasr. 2024. "Backseat Gaming" A Study of Co-Regulated Learning within a Collegiate Male Esports Community. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[45] Erica Kleinman, Jennifer Villareale, Murtuza N Shergadwala, Zhaoqing Teng, Andy Bryant, Jichen Zhu, and Magy Seif El-Nasr. 2023. "What else can I do?" Examining the Impact of Community Data on Adaptation and Quality of Reflection in an Educational Game. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–12.

[46] Sara Klingenberg, Maria LM Jørgensen, Gert Dandanell, Karen Skriver, Aske Mottelson, and Guido Makransky. 2020. Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology* 51, 6 (2020), 2115–2138.

[47] Karl W Kosko, Richard E Ferdig, and Lionel Roche. 2021. Conceptualizing a shared definition and future directions for extended reality (XR) in teacher education. *Journal of Technology and Teacher Education* 29, 3 (2021), 257–277.

[48] Tiffany A Koszalka, Mary K Wilhelm-Chapin, Christopher D Hromalik, Yuri Pavlov, and Lili Zhang. 2019. Prompting deep learning with interactive technologies: Theoretical perspectives in designing interactive learning resources and environments. *Learning in a digital world: Perspective on interactive technologies for formal and informal education* (2019), 13–36.

[49] Bracha Kramarski and Zehavit Kohen. 2017. Promoting preservice teachers' dual self-regulation roles as learners and as teachers: Effects of generic vs. specific prompts. *Metacognition and learning* 12 (2017), 157–191.

[50] Mohammad Amin Kuhail, Nazik Alturki, Salwa Alramlawi, and Kholood Alhejori. 2023. Interacting with educational chatbots: A systematic review. *Education and Information Technologies* 28, 1 (2023), 973–1018.

[51] Hee Yoon Kwon and Koray Özpolat. 2021. The dark side of narrow gamification: Negative impact of assessment gamification on student perceptions and content knowledge. *INFORMS Transactions on Education* 21, 2 (2021), 67–81.

[52] Georgios Lampropoulos, Pablo Fernández-Arias, Álvaro Antón-Sancho, and Diego Vergara. 2024. Affective Computing in Augmented Reality, Virtual Reality, and Immersive Learning Environments. *Electronics* 13, 15 (2024), 2917.

[53] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[54] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.

[55] Check-Yee Law, John Grundy, Andrew Cain, Rajesh Vasa, and Alex Cummaudo. 2017. User Perceptions of using an open learner model visualisation tool for facilitating self-regulated learning. In *Proceedings of the Nineteenth Australasian Computing Education Conference*. 55–64.

[56] Eunbae Lee and Michael J Hannafin. 2016. A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. *Educational technology research and development* 64 (2016), 707–734.

[57] Äli Leijen, Kai Valtna, Djuddah AJ Leijen, and Margus Pedaste. 2012. How to determine the quality of students' reflections? *Studies in Higher Education* 37, 2 (2012), 203–217.

[58] Xiaodong Lin, Cindy Hmelo, Charles K Kinzer, and Teresa J Secules. 1999. Designing technology to support reflection. *Educational technology research and Development* 47 (1999), 43–62.

[59] Zhongxiu Liu, Visit Pataranutaporn, Jaclyn Ocumpaugh, and Ryan Baker. 2013. Sequences of frustration and confusion, and learning. In *Educational data mining 2013*.

[60] Guido Makransky and Gustav B Petersen. 2021. The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review* 33, 3 (2021), 937–958.

[61] Guido Makransky, Thomas S Terkildsen, and Richard E Mayer. 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and instruction* 60 (2019), 225–236.

[62] Karen Mann, Jill Gordon, and Anna MacLeod. 2009. Reflection and reflective practice in health professions education: a systematic review. *Advances in health sciences education* 14 (2009), 595–621.

[63] Sai Siddartha Maram, Erica Kleinman, Jennifer Villareale, Jichen Zhu, and Magy Seif El-Nasr. 2024. "Ah! I see"-Facilitating Process Reflection in Gameplay through a Novel Spatio-Temporal Visualization System. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[64] Catherine Marechal, Dariusz Mikolajewski, Krzysztof Tyburek, Piotr Prokopowicz, Lamine Bougueroua, Corinne Ancourt, and Katarzyna Wegrzyn-Wolska. 2019. Survey on AI-Based Multimodal Methods for Emotion Detection. *High-performance modelling and simulation for big data applications* 11400 (2019), 307–324.

[65] Richard E Mayer. 2005. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning* 41, 1 (2005), 31–48.

[66] Eileen McGivney. 2023. *Promoting Learning, Agency, and Motivation in STEM Classrooms with Virtual Reality Field Trips*. Ph. D. Dissertation.

[67] Miriah Meyer and Jason Dykes. 2018. Reflection on reflection in applied visualization research. *IEEE computer graphics and applications* 38, 6 (2018), 9–16.

[68] Oliver A Meyer, Magnus K Omdahl, and Guido Makransky. 2019. Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education* 140 (2019), 103603.

[69] Sam Morris. 2019. The frustration regulation journal: A reflective framework for educators. *Relay Journal* 2, 2 (2019), 294–305.

[70] Atharva Naik, Jessica Ruhan Yin, Anusha Kamath, Qianou Ma, Sherry Tongshuang Wu, Charles Murray, Christopher Bogart, Majd Sakr, and Carolyn P Rose. 2024. Generating Situated Reflection Triggers About Alternative Solution Paths: A Case Study of Generative AI for Computer-Supported Collaborative Learning. In *International Conference on Artificial Intelligence in Education*. Springer, 46–59.

[71] Scott Nicholson. 2015. Peeking behind the locked door: A survey of escape room facilities. (2015).

[72] Scott Nicholson. 2018. Creating engaging escape rooms for the classroom. *Childhood Education* 94, 1 (2018), 44–49.

[73] Santiago Ojeda-Ramirez, Sina Rismanchian, and Shayan Doroudi. 2023. Learning about AI to learn about learning: Artificial intelligence as a tool for metacognitive reflection. (2023).

[74] Harold F O'Neil, Gregory KWK Chung, Deirdre Kerr, Terry P Vendlinski, Rebecca E Buschang, and Richard E Mayer. 2014. Adding self-explanation prompts to an educational computer game. *Computers in Human Behavior* 30 (2014), 23–28.

[75] Rui Pan, Henry Lo, and Carman Neustaedter. 2017. Collaboration, awareness, and communication in real-life escape rooms. In *Proceedings of the 2017 conference on designing interactive systems*. 1353–1364.

[76] Jocelyn Parong and Richard E Mayer. 2018. Learning science in immersive virtual reality. *Journal of Educational Psychology* 110, 6 (2018), 785.

[77] Tanner M Phillips, Asmalina Saleh, and Gamze Ozogul. 2023. An AI toolkit to support teacher reflection. *International Journal of Artificial Intelligence in Education* 33, 3 (2023), 635–658.

[78] Daniel Pimentel, Géraldine Fauville, Kai Frazier, Eileen McGivney, Sergio Rosas, and Erika Woolsey. 2022. An introduction to learning in the metaverse. *Meridian Treehouse* 3, 4 (2022), 348–352.

[79] Anna CM Queiroz, Géraldine Fauville, Adina T Abeles, Aaron Levett, and Jeremy N Bailenson. 2023. The efficacy of virtual reality in climate change education increases with amount of body movement and message specificity. *Sustainability* 15, 7 (2023), 5814.

[80] Anna C M Queiroz, Eileen McGivney, Sunny X Liu, Courtney Anderson, Brian Beams, Cyan DeVeaux, Kai Frazier, Kai Xr, Eugy Han, Mark Roman Miller, Xander S Peterson, Erika S Woolsey, Jeffrey T Hancock, and Jeremy N Bailenson. [n. d.]. Collaborative Tasks in Immersive Virtual Reality Increase Learning. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning* (2023). International Society of the Learning Sciences.

[81] Jaziar Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & education* 147 (2020), 103778.

[82] Mladen Raković, Matthew L Bernacki, Jeffrey A Greene, Robert D Plumley, Kelly A Hogan, Kathleen M Gates, and Abigail T Panter. 2022. Examining the critical role of evaluation and adaptation in self-regulated learning. *Contemporary Educational Psychology* 68 (2022), 102027.

[83] Abhishek Revadekar, Shreya Oak, Aumkar Gadekar, and Pramod Bide. 2020. Gauging attention of students in an e-learning environment. In *2020 IEEE 4th Conference on Information & Communication Technology (CICT)*. IEEE, 1–6.

[84] Ma Mercedes T Rodrigo, Ryan SJd Baker, and Julieta Q Nabos. 2010. The relationships between sequences of affective states and learner achievement. In *Proceedings of the 18th international conference on computers in education*. Universiti Putra Malaysia Malaysia, 56–60.

[85] María Jesús Rodríguez-Triana, Luis P Prieto, Andrii Vozniuk, Mina Shirvani Boroujeni, Beat A Schwendimann, Adrian Holzer, and Denis Gillet. 2017. Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. *International Journal of Technology Enhanced Learning* 9, 2-3 (2017), 126–150.

[86] Richard M Ryan and Edward L Deci. [n. d.]. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. ([n. d.]), 11.

[87] Johnny Saldaña. 2021. *The coding manual for qualitative researchers*. SAGE Publications Limited.

[88] Donald A Schön. 2017. *The reflective practitioner: How professionals think in action*. Routledge.

[89] Ximing Shen, Yun Suen Pai, Dai Kiuchi, Kehan Bao, Tomomi Aoki, Hikari Meguro, Kanoko Oishi, Ziyue Wang, Sohei Wakisaka, and Kouta Minamizawa. 2023. Dementia Eyes: Co-Design and Evaluation of a Dementia Education Augmented Reality Experience for Medical Workers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 778, 18 pages. https://doi.org/10.1145/3544548.3581009

[90] Jill Stefaniak, Tian Luo, and Meimei Xu. 2021. Fostering pedagogical reasoning and dynamic decision-making practices: A conceptual framework to support learning design in a digital age. *Educational Technology Research and Development* 69, 4 (2021), 2225–2241.

[91] Kachina Studer, Hing Lie, Zhen Zhao, Ben Thomson, Dishita G Turakhia, and John Liu. 2024. An Open-ended System in Virtual Reality for Training Machining Skills. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 392, 5 pages. https://doi.org/10.1145/3613905.3648666

[92] Raja M Suleman, Riichiro Mizoguchi, and Mitsuru Ikeda. 2016. A new perspective of negotiation-based dialog to enhance metacognitive skills in the context of open learner models. *International Journal of Artificial Intelligence in Education* 26, 4 (2016), 1069–1115.

[93] Zouheir Trabelsi, Fady Alnajjar, Medha Mohan Ambali Parambil, Munkhjargal Gochoo, and Luqman Ali. 2023. Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing* 7, 1 (2023), 48.

[94] Päivi Tynjälä. 1998. Writing as a tool for constructive learning: Students' learning experiences during an experiment. *Higher education* 36 (1998), 209–230.

[95] Alice Veldkamp, Liesbeth van de Grint, Marie-Christine PJ Knippels, and Wouter R van Joolingen. 2020. Escape education: A systematic review on escape rooms in education. *Educational Research Review* 31 (2020), 100364.

[96] Jennifer Villareale, Colan F. Biemer, Magy Seif El-Nasr, and Jichen Zhu. 2020. Reflection in Game-Based Learning: A Survey of Programming Games. In *International Conference on the Foundations of Digital Games*. 1–9.

[97] Günter Wallner, Nour Halabi, and Pejman Mirza-Babaei. 2019. Aggregated visualization of playtesting data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[98] Abigail Ford Winkel, Sandra Yingling, Aubrie-Ann Jones, and Joey Nicholson. 2017. Reflection as a learning tool in graduate medical education: a systematic review. *Journal of Graduate Medical Education* 9, 4 (2017), 430–439.

[99] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376781

[100] Irmtraud Wolfbauer, Mia Magdalena Bangerl, Katharina Maitz, and Viktoria Pammer-Schindler. 2023. Rebo at Work: Reflecting on Working, Learning, and Learning Goals with the Reflection Guidance Chatbot for Apprentices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[101] Sebastian Wollny, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler. 2021. Are we there yet?-a systematic literature review on chatbots in education. *Frontiers in artificial intelligence* 4 (2021), 654924.

[102] Bian Wu, Xiaoxue Yu, and Xiaoqing Gu. 2020. Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British journal of educational technology* 51, 6 (2020), 1991–2005.

[103] Chin-Lan Yang, Ching-Yi Chang, and Hsiu-Ju Jen. 2023. Facilitating undergraduate students' problem-solving and critical thinking competence via online escape room learning. *Nurse Education in Practice* 73 (2023), 103828.

[104] Kexin Yang, Xiaofei Zhou, and Iulian Radu. 2020. XR-ed framework: Designing instruction-driven andLearner-centered extended reality systems for education. *arXiv preprint arXiv:2010.13779* (2020).

[105] Bekir Yıldırım, Emine Sahin Topalcengiz, Gökhan Arıkan, and Serkan Timur. 2020. Using virtual reality in the classroom: Reflections of STEM teachers on the use of teaching and learning tools. *Journal of Education in Science Environment and Health* 6, 3 (2020), 231–245.

[106] Jamaal Rashad Young, Danielle Bevan, and Miriam Sanders. 2024. How productive is the productive struggle? Lessons learned from a scoping review. *International Journal of Education in Mathematics, Science and Technology* 12, 2 (2024), 470–495.

[107] Guoyang Zhou, Amy Nagle, George Takahashi, Tera Hornbeck, Ann Loomis, Beth Smith, Bradley Duerstock, and Denny Yu. 2022. Bringing Patient Mannequins to Life: 3D Projection Enhances Nursing Simulation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 565, 15 pages. https://doi.org/10.1145/3491102.3517562