Complete but Not Practising: Examining Agency and Professional Realism in Skill-Based Job Simulations

ANONYMOUS AUTHOR(S)

Skill-based job simulations can be easy to complete yet provide limited agency or sense of professional practice. This study examined how four early-career participants experienced scaffolding, agency, and professional realism in one selected platform case. A comparative heuristic inspection of Forage and Springpod informed case selection. Participants then completed remote moderated think-aloud usability sessions, System Usability Scale questionnaires, and debrief interviews using the selected Forage case. Reflexive thematic analysis identified four themes: professional realism varied with expectations and task content; scaffolding supported clarity but could limit participation; contextual support was difficult to discover; and transactional value remained when practice felt shallow. The study contributes a bounded account of how completion, agency, and professional realism can separate. Productive friction offers a tentative interpretation, translated into three findings-informed, unevaluated design directions. The phases examined different modules, the inspection involved one evaluator, and neither the lens nor the prototype was tested with participants.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *Usability testing*; *Interaction design*; *HCI theory, concepts and models.*

Additional Key Words and Phrases: job simulations, scaffolding, user experience, professional realism, productive friction

1 Introduction

Skill-based job simulation platforms occupy an ambiguous space between learning systems and employability interfaces. Digital recruitment literature describes candidate experience as a sequence of designed touchpoints, while workplace-simulation and gameful-assessment research shows how digital tasks can combine practice, signalling, and evaluation [21, 23, 27, 29]. Early-career users may therefore approach the same simulation as a learning activity, a preview of professional work, or a recruitment signal. Although these purposes are related, they create different expectations for a successful experience. User experience (UX) also extends beyond efficient task completion to pragmatic, experiential, and contextual qualities [12, 15, 24]. Completion alone therefore does not show whether a job simulation provides enough agency, judgement, and role exploration for users to feel meaningfully connected to the work it represents.

The central design tension concerns instructional scaffolding and professional realism. Guidance can help users navigate complex activities and reduce demands unrelated to the target task [25, 32], but it can also reduce the judgement and reasoning required for completion. This raises a motivating design question: when does support enable meaningful engagement with professional practice, and when does it begin to replace that practice? Simulation fidelity depends on the task-relevant processes and forms of action represented rather than surface resemblance alone [2, 3, 29]. The challenge is therefore not simply to make a simulation easier or harder, but to determine which forms of ambiguity, judgement, and consequence belong to the represented work and which are avoidable interface failures.

This paper addresses the following research question: How did early-career participants experience scaffolding, agency, and professional realism in one selected skill-based job simulation? Phase 1 used a researcher-led comparative heuristic inspection of one module on each of two browser-based job-simulation platforms to identify recurring interface patterns and select a case with fewer basic usability obstacles. Phase 2 followed with a user study comprising remote moderated think-aloud usability-testing sessions, SUS questionnaires, and debrief interviews with four early-career participants using the registered participant-study scenario, Forage's *Bloomberg Essentials*. Because the comparative heuristic inspection and user study examined different modules, they were integrated only at the level of recurring

interface patterns across the overall study, not as direct task replication. Scaffolding, task content, and prior expectations also varied together, so the observed patterns cannot be attributed to scaffolding alone. Participants are described through coarse prior-expectation contexts to support interpretation without needing to disclose identifying backgrounds. The deliberately selected case remains exploratory and non-generalisable.

The primary contribution is a bounded empirical account of how smooth completion, agency, and professional realism can separate in one selected job-simulation case. Productive friction offers a tentative interpretation of this tension, not a claim to have coined the term. Earlier scholarship uses productive friction for resolving socio-cognitive conflict in collaborative knowledge construction [13], while more recent design literature frames it as calibrated challenge in digital interaction [10] or applies it to effortful augmented-reading interactions [16]. Here, the lens organises the completion–agency–professional-realism tension observed in the study. Three proposed design directions—reversible user-controlled scaffolding, layered contextual support, and consequential feedback and recovery—form an interaction loop illustrated through the post-study Project Room prototype. The design material is findings-informed and unevaluated; it does not establish improved user outcomes.

2 Related Work

2.1 Professional Realism in Simulation

Simulation-based learning represents selected aspects of practice rather than reproducing professional work in full. Bauer et al. describe this selection as representational scaffolding: designers can modify the information, agency, typicality, and dynamics available within a simulation [2]. Bell et al. distinguish surface realism from psychological fidelity, where a simulation prompts task-relevant processes without recreating the entire performance environment [3]. Reviews similarly describe workplace simulation through combinations of scenario, learner control, feedback, and fidelity [17, 29]. These accounts locate realism in what users are asked to interpret and do, not only in how an interface looks.

In this paper, professional realism concerns whether an interaction preserves credible opportunities for interpretation, choice, communication, and consequence. It is an experiential construct rather than a claim that the simulation reproduces professional work in full. A clear route to completion may therefore coexist with low professional realism when users have little scope to exercise the judgement implied by the represented role.

2.2 Scaffolding and User Experience

Scaffolding shapes access to the forms of action a simulation makes available. Chernikova et al.'s meta-analysis shows that simulation outcomes vary with task conditions, prior knowledge, authenticity, and instructional support [5]. Technology-mediated scaffolds can be static or responsive, while predetermined fading may not match an individual's emerging need [31]. Park consequently recommends support that is visible, usable, selectable, and differentiated in level [25]. These sources do not identify a universal optimum. Instead, they describe a design space in which guidance can enable access to a task, perform part of the reasoning on the user's behalf, or introduce additional interface work.

UX research broadens this issue beyond instrumental efficiency. Hassenzahl and Tractinsky describe experience as pragmatic, affective, and contextual rather than solely task-oriented [12], while Ntoa argues that observed and perceived evidence provide complementary views of usability and UX [24]. Simulation-specific research adds fidelity, psychological safety, reflection, and debriefing as relevant experiential dimensions [8, 15]. In this study, structural usability concerns navigation, feedback, terminology, help, and ease of completion. Professional realism concerns

whether users can interpret a situation, exercise agency, and encounter credible consequences. Smooth progression can coexist with limited professional practice, while unclear navigation remains an interface failure rather than a productive challenge.

2.3 Employability Context

Recruitment literature provides an important contextual boundary and suggests that users often interpret digital and gameful assessments not only as tasks but also as signals about employability, organisations, and selection decisions. Empirical studies consequently show that applicant reactions vary with task design, procedural transparency, explanation quality, and perceived assessment validity [18, 27, 28, 30]. The findings are mixed: one vendor-linked study reports favourable test-taking experiences alongside continuing face-validity concerns [19], while broader hiring research identifies construct, fairness, and access problems in apparently efficient digital processes [6, 7]. Vendor-linked work involving autistic job seekers likewise reinforces the need for job relevance and group-level evaluation [35].

The module examined with participants was a voluntary career-exploration activity rather than a selection instrument. Recruitment research is therefore used to explain why completion, certificates, and organisational signalling may contribute to perceived value even without formal decision stakes. Together, the simulation, UX, and employability literatures provide the concepts needed to examine how scaffolding, agency, professional realism, and transactional value meet in one browser-based case. Realist evaluation further supports treating these relationships as context-dependent rather than uniformly transferable [11].

3 Methodology

The study examined how four early-career participants experienced scaffolding, agency, and professional realism in one selected browser-based skill simulation. It combined a comparative heuristic inspection with remote moderated usability testing supported by participants' think-aloud, descriptive SUS questionnaires, semi-structured debrief interviews, and qualitative analysis. The study received institutional ethics approval. Participants provided written consent and verbally reconfirmed recording consent; audio and screen activity were recorded, faces were not, and research records were de-identified.

Table 1 summarises the two phases, cases, purposes, and involvement.

Table 1. Study phases, cases, purposes, and involvement. Platform names are shown in plain text and module titles in italics.

Phase	Platform and module	Purpose	Method and involvement
1	Forage, *Lloyds Banking Group UX Design Introduction*; Springpod, *Amazon Digital Careers*	Compare recurring interface patterns and select the participant-study platform	Comparative heuristic inspection by one researcher
2	Forage, *Bloomberg Essentials*	Examine perceived usability, expectations, professional realism, and career value in the selected case	Remote moderated sessions with four early-career participants; think-aloud, SUS, and debrief

3.1 Study Design and Case Selection

Phase 1 compared navigation, progress feedback, error prevention, terminology, and help in two browser-based job-simulation platforms, Forage and Springpod. The project records identify the two modules in Table 1 as the cases used

for this comparison but do not document a narrower sampling rationale. Jakob Nielsen's original ten heuristics supplied a common comparison frame [22]; the researcher interpreted them in context rather than constructing a specialised educational or gameful instrument [33]. The researcher also assigned study-defined severity categories: problems that blocked completion, caused delay or confusion, or were minor.

The inspection indicated that the Forage module had the strongest structural-usability baseline and was therefore selected as the platform for Phase 2 to reduce the likelihood that basic interface defects could dominate the sessions. This decision also conditions the participant findings because the tested platform was not selected independently. Within the Forage platform, Phase 1 inspected *Lloyds Banking Group UX Design Introduction*, whereas Phase 2 used the registered participant-study task, *Bloomberg Essentials*. The phases were therefore integrated sequentially and only through recurring interface patterns, not as direct task replication, and neither phase provided an independent validation of the other. Inspection, observed behaviour, and self-report provided complementary views of the interface [24, 26, 36].

3.2 Participants and Procedure

Four early-career participants were purposively recruited for the Phase 2 examination of the *Bloomberg Essentials* module: penultimate- or final-year university students and people who had graduated within the previous two years. From stated backgrounds and prior familiarity with the represented activities, the researcher classified P1 and P4 as domain-agnostic and P2 and P3 as domain-relevant. These coarse labels contextualised prior expectations without needing to disclose detailed backgrounds; they were not user types or inferential comparison groups. Participants received no payment or other incentives.

Each participant completed a single remotely moderated 60–75-minute session on their own computer. During the session, participants created a free Forage account, imagined seeking practical career experience, found and enrolled in *Bloomberg Essentials*, completed its opening content and activities, and responded to a scenario prompted by the researcher, acting as the moderator, in which a video would not load and in-site help had to be found. The last task examined where support was sought during the activity. Participants used the think-aloud protocol throughout; because moderator intervention can affect this evidence [1], the researcher used neutral reminders only after extended silence [20, 36]. The role-play prompts and debrief questions elicited participants' prior expectations.

After the tasks, participants completed the ten-item SUS as a subjective perceived-usability measure [34]. A semi-structured debrief interview then examined expectations, notable moments, perceived professional realism, and career value. Professional realism was derived from these qualitative data, not SUS: it concerned whether the activity preserved credible interpretation, choice, communication, and consequence. The two measures therefore addressed separate but complementary evidence needs. The introductory and debrief interview questions are provided in Appendix A.

3.3 Analysis and Cross-Phase Interpretation

SUS totals were calculated on the 0–100 scale and used descriptively; the limited sample of four scores cannot support inference or stable benchmark comparison. The complementary qualitative data comprised think-aloud verbalisations, observed interactions, screen recordings, and debrief interview transcripts. The researcher used reflexive thematic analysis to examine both explicit accounts and underlying meanings [4, 9]. Repeated review was followed by semantic and latent coding on a digital whiteboard; six initial patterns were checked against the data and refined into four themes from Phase 2: professional realism varied with prior expectations and task content; scaffolding supported clarity but could limit active participation; contextual support was hard to discover in the prompted task; and transactional value

could remain when practice felt shallow. The researcher's interaction-design background shaped what was noticed, while observation logs, codes, grouping decisions, and revisions documented the reflexive process.

A synthesis matrix compared the themes and observed behaviours with recurring Phase 1 interface findings. Agreement or difference was recorded only at this interface-pattern level because the phases used different modules. This comparison did not validate either phase. Only P4's session record includes additional framing that described the module as short, introductory, and not role-specific; comparable wording is not recorded for the other sessions. Whether administration or recording differed cannot be determined, so possible priming is retained as a limitation. Tables 1 and 3 report anonymised study structure and findings-to-design traceability; raw samples remain excluded because the source records contain identifying information.

4 Results

4.1 Descriptive Usability Context

The four SUS totals were 82.5, 47.5, 80.0, and 85.0. These values describe perceived usability in four sessions; they do not support a population pattern or a comparison between prior-expectation contexts. Participants with similarly favourable totals nevertheless described agency and professional realism differently.

Table 2 reports participant-level prior-expectation contexts, SUS totals, and qualitative session context.

Table 2. Participant-level descriptive SUS totals with qualitative session context; contextual descriptions are not score interpretations.

Participant	Prior-expectation context	SUS total	Qualitative session context
P1	Domain-agnostic	82.5	Clear completion but limited active participation
P2	Domain-relevant	47.5	Onboarding and professional-realism concerns
P3	Domain-relevant	80.0	Reclassified the experience as generic training
P4	Domain-agnostic	85.0	More favourable orientation and flow

The qualitative analysis generated four themes: professional realism varied with prior expectations and task content; scaffolding supported clarity but could limit active participation; contextual support was hard to discover in the prompted task; and transactional value could remain when practice felt shallow. These are interpretive patterns from four sessions, not user types.

4.2 Theme 1: Professional Realism Varied with Prior Expectations and Task Content

Participants evaluated professional realism against what they expected the represented work to involve. The two domain-relevant participants entered with comparatively specific expectations about activities and responsibilities in their fields. When the module instead emphasised company information, simple choices, and a pre-written communication task, both reclassified it in educational terms. P2 said, "It feels more like a course", while P3 compared it with mandatory university training. In these sessions, the activities provided less judgement than those participants expected; this pattern does not establish a general effect of domain familiarity.

The domain-agnostic participants did not respond uniformly. P1 expected practical job preparation and criticised the limited opportunity to perform work-like activity. P4 began without detailed industry knowledge and treated the

module as useful orientation once the flow became familiar, describing it as “crisp” and “expected”. These accounts show variation within as well as between the two contexts.

4.3 Theme 2: Scaffolding Supported Clarity but Could Limit Active Participation

Across the sessions, participants recognised the clarity of the platform’s linear structure. Information appeared in small segments, progress was visible, and the sequence was easy to follow. Participants described definitions as concise, while the module’s quiz and navigation patterns provided a stable route. This recurring pattern matched the Phase 1 observation of visible progress and linear guidance; because the phases used different modules, the match concerns the interface pattern only.

The same scaffolding narrowed what participants could decide or produce. P1 reported mainly reading information, while observation showed that a communication task supplied the message, leaving P1 to “just click Send”. Participants also described quiz options as obvious or too easy to provide meaningful progress, and one characterised the interaction as a click-through task. Easy progression therefore coincided with limited active judgement.

Not every difficulty came from excessive guidance. A card appeared to invite a choice but only revealed information when flipped, and a circular team diagram did not appear interactive until hover exposed it. These observed ambiguities created interface work rather than professional practice.

4.4 Theme 3: Contextual Support Was Hard to Discover in the Prompted Task

All four participants initially failed to locate in-site help during the prompted video-failure task described in Section 3.2. They searched the activity, top navigation, profile area, messages, and a tips control. Generic support routes existed elsewhere, but not where participants first looked. This finding applies only to the prompted scenario; it does not show that help was absent or that every support interaction would fail.

After the initial search, participants moved to a footer or contact route, proposed web or browser troubleshooting, or asked the researcher for a clue. P2 could not locate in-site support, while P3 shifted to browser inspection tools. The inspection had found site-level support routes, whereas participants first searched near the represented task.

4.5 Theme 4: Transactional Value Could Remain When Practice Felt Shallow

Several participants treated certificates, visible completion, or preparation for a company-specific application as reasons to finish or recommend the module. P1 valued that interviewers could see completion. P3 called the module “not useful” for learning the job itself but still valued its explanation of company structure. We use transactional value for these benefits of completion and signalling rather than skill practice.

The data show that participants cared about more than just credentials. P4 valued orientation to how the company worked, while P2 questioned whether requested onboarding information improved the experience or imposed a cost of entry. Participants thus weighed practice, company orientation, and recruitment signalling against participation costs. A favourable SUS total did not identify which form of value a participant recognised.

Together, the themes separate smooth completion from professional practice: clear sequencing supported navigation, but generic tasks limited judgement, and support was not discovered near the prompted activity. This leads to an additional design question: how might simulations preserve clarity while retaining meaningful decisions, contextual support, and consequences?

5 Discussion

5.1 Scaffolding, Agency, and Professional Realism

In the selected Forage *Bloomberg Essentials* case, participants experienced scaffolding as supporting clear progression while sometimes limiting the judgement and action available to them. Professional realism varied with prior expectations and task content, particularly for the domain-relevant participants, P2 and P3. Contextual support was hard to discover in the prompted task, while transactional value could remain when practice felt shallow. The RQ is therefore answered through these experiences: participants valued guidance that clarified progression, but described limited agency or professional realism when activities supplied expected judgement or redirected attention to website troubleshooting.

This finding complements Bauer et al.'s representational scaffolding and Bell et al.'s psychological fidelity: professional realism depended on the processes available to enact, not workplace appearance alone [2, 3]. It also qualifies scaffolding research. Visible, selectable support can aid access [25, 31], yet the pre-written message and obvious options illustrate how guidance can complete target reasoning, consistent with Chernikova et al.'s emphasis on task and prior-knowledge conditions [5]. The separation between favourable SUS totals and participants' accounts further supports multidimensional UX evaluation rather than treating efficient progression as sufficient [12, 15, 24].

The bounded empirical contribution is the completion–agency–professional-realism account: smooth completion, credible practice, and perceived value can separate. Because scaffolding, task content, and prior expectations varied together, this is an interpretive account rather than a causal or group effect. Productive friction provides a provisional lens for considering that account in design.

5.2 Productive Friction

We use productive friction for purposeful decisions, task-relevant uncertainty, and bounded consequences retained without compromising usability or safety. It excludes inaccessible interaction, generic difficulty, and interface failure: the aim is to preserve professional judgement, not make a simulation harder. Earlier uses concern collaborative knowledge construction, digital-media design, and augmented reading [10, 13, 16]. Here, the term organises the observed completion–agency–professional-realism tension for browser-based job simulations; it is not presented as a new term, a novel technique, or a general theory.

Following Höök and Löwgren's account of intermediate-level interaction-design knowledge [14], the lens connects this bounded case to questions that can travel without claiming a general theory. Does effort come from professional judgement or interface operation? Is support visible, optional, and restorable? Are consequences bounded and recoverable? Are access needs met without removing meaningful choice? The present findings offer tentative answers: pre-written communication removed judgement, misplaced help added interface work, and participants still valued orientation or signalling when practice felt shallow. These are proposed questions for future evaluation, not validated criteria.

5.3 Design Directions and Interaction Loop

The authors propose three design directions in an interaction loop, illustrated through the post-study Project Room prototype: Calibrate Support (Stage 1), Act with Contextual Support (Stage 2), and Experience Consequences and Recover (Stage 3), as presented in Figure 1. User choice and consequences can inform the next support level. The proposal is findings-informed but unevaluated.

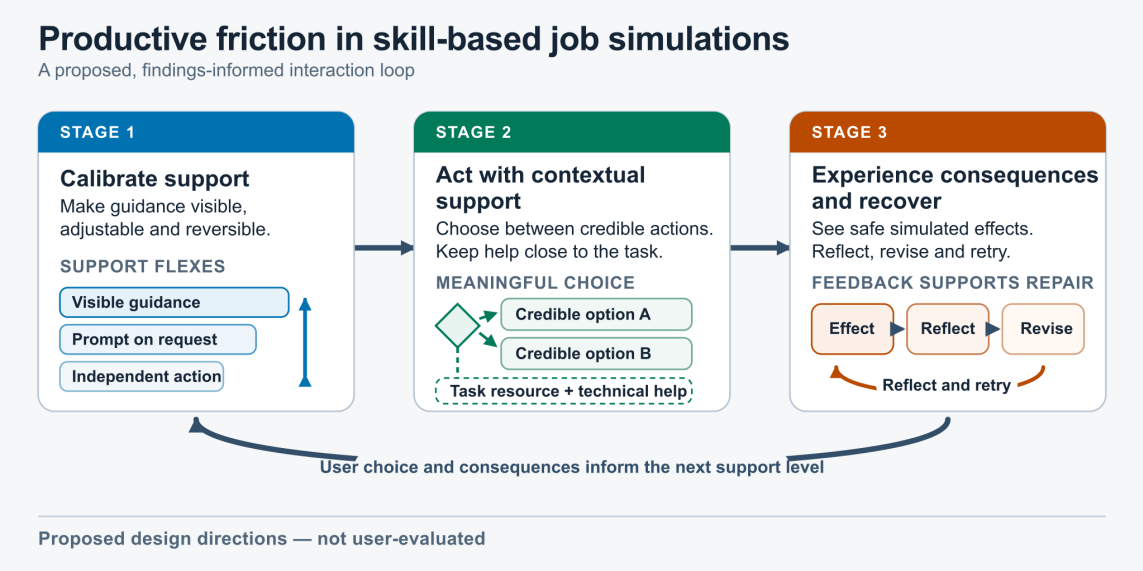

Fig. 1. Proposed interaction loop: Stage 1, Calibrate Support; Stage 2, Act with Contextual Support; and Stage 3, Experience Consequences and Recover. The directions have not been user-evaluated.

To illustrate how the proposed interaction loop could appear within a job-simulation interaction, a low/medium-fidelity Project Room prototype was developed after the empirical analysis. It represents a fictional team deciding how to handle a risky software launch. Stage 1 selects Guided, Balanced, or Independent support. Figure 2 shows the Balanced support level in Stage 2, Act and Decide, with a project brief, three deadline-versus-quality recommendations, optional project resources, and separate technical and accessibility help. Stage 3, Experience Consequences and Recover (Figure 3), shows what the selected limited pilot protected and where it transferred risk, then invites reflection, revision, or restart.

The prototype is presented for illustrative purposes only. Three author-led walkthroughs exercised restorable guidance, independent action, and help–consequence–recovery paths. They checked state coherence and traceability to the themes, but were not independent, involved no participants, and provided no usability or effectiveness evaluation as part of this study.

5.3.1 Reversible User-Controlled Scaffolding. The literature discussed and Themes 1 and 2 motivate support that users can adjust without being classified in advance. Project Room keeps Guided, Balanced, and Independent controls visible and permits change before confirmation. This extends calls for visible, selectable scaffolds [25, 31] while respecting evidence that support value varies with task and prior knowledge [5]. Support must remain restorable and accessible; the direction does not establish a novice–expert effect or improved learning.

5.3.2 Layered Contextual Support. In the study, the prompted task showed that participants expected technical help near the activity. Theme 3 therefore motivates distinguishing task uncertainty from interface failure. Figure 2's optional brief and teammate prompt preserve in-role reasoning, while technical and accessibility help stays explicit and outside the fiction. This applies fidelity to task-relevant processes [2, 3] and places support near need [25]. The study did not

test an in-role information need: that layer remains a proposal. Privacy, safeguarding, and accessibility routes must remain direct rather than hidden inside role-play [6, 7, 35].

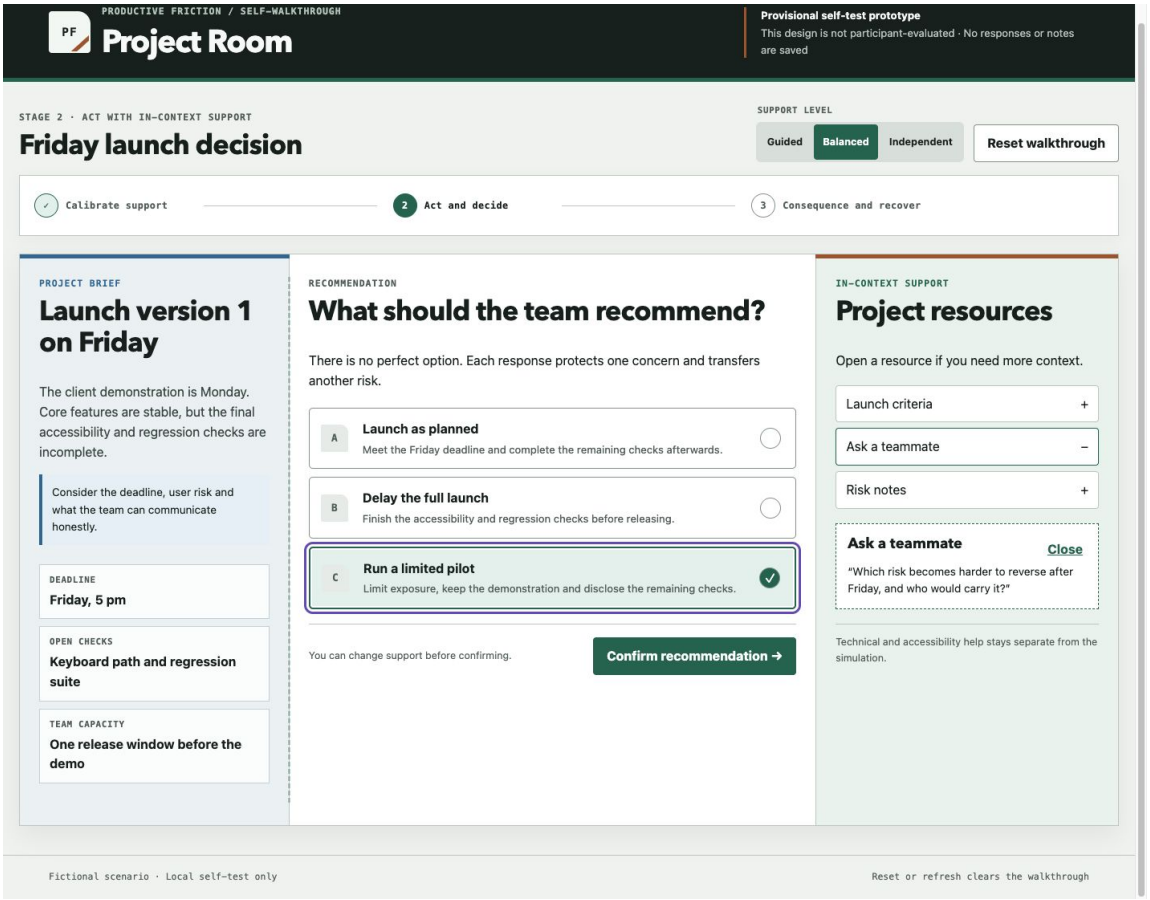

Fig. 2. Project Room Stage 2, Act and Decide. Three recommendations remain available while optional in-context support is open beside the task.

5.3.3 Consequential Feedback and Recovery. Theme 4 motivates making practice visible beyond completion signals. Figure 3 shows a fictional protected outcome and transferred risk, then permits reflection and revision. This translates simulation work linking feedback and debriefing [8, 15, 17, 29] into an untested interface direction. Recruitment research warns that experience, fairness, and face validity can diverge [18, 19, 27, 28, 30]; consequences must therefore be proportionate, recoverable, and explicitly simulated, never presented as predictions of employability or job performance.

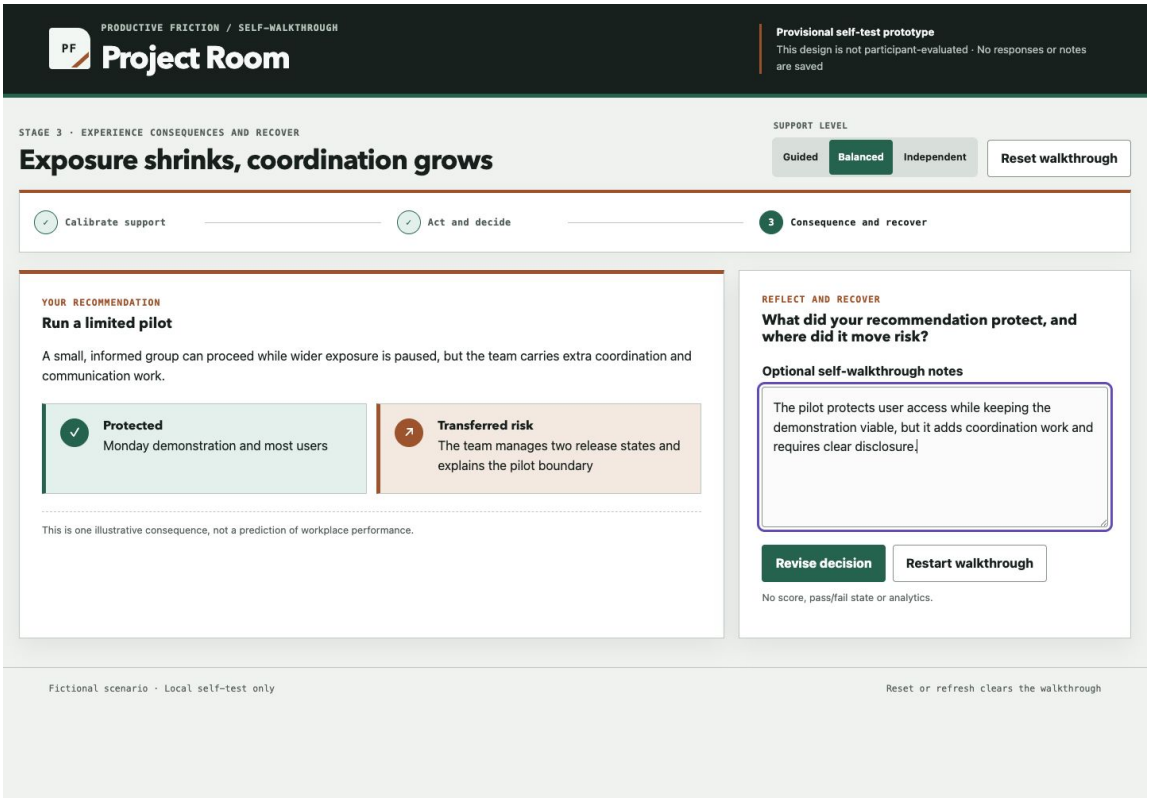

Fig. 3. Project Room Stage 3, Experience Consequences and Recover. The selected limited pilot produces a bounded fictional consequence, reflection prompt, and recovery routes.

Table 3 shows how the four themes feed the three directions and the guardrails that keep the proposal within the evidence.

Table 3. Findings-to-design-direction traceability.

Finding	Design tension	Design direction	Illustrative mechanism	Guardrail
Professional realism varied with prior expectations and task content	Identical support may not fit different expectations	Reversible user-controlled scaffolding	Guidance can recede and be restored	Do not infer fixed user types
Scaffolding supported clarity but could limit active participation	Guidance may enable completion while removing agency	Reversible user-controlled scaffolding	Credible options precede optional prompts	Do not claim improved learning
Contextual support was hard to discover in the prompted task	Site-level help may be distant from activity	Layered contextual support	Separate in-role resources from explicit technical help near the task	In-role support was not participant-tested
Transactional value could remain when practice felt shallow	Completion signals may overshadow practice	Consequential feedback and recovery	Show a simulated effect, reflection, and revision route	Do not claim employability outcomes

6 Limitations and Future Work

This bounded account concerns four early-career participants completing a single 60–75-minute session with the deliberately selected Forage *Bloomberg Essentials* case, which contains broad introductory content. Because Forage was chosen after showing the strongest structural-usability baseline, the SUS totals are not an unconditioned platform comparison. Scaffolding, task content, and prior expectations were not varied independently, so causal attribution is not possible. The study examined immediate UX, not learning, employability, or job performance.

Phase 1 used one researcher and inspected Forage *Lloyds Banking Group UX Design Introduction* and Springpod *Amazon Digital Careers*, not the Phase 2 module. Cross-phase synthesis therefore concerns recurring interface patterns only. Graham and McAleer's realist account supports treating transfer across contexts as conditional [11], not assuming that the observed relationships recur unchanged.

Remote moderation, concurrent think-aloud, and the short sessions may have influenced participant behaviour. Only P4's record contains extra framing that the module was short, introductory, and not role-specific; comparable wording is absent from the other records. Administration and recording cannot be disentangled, so priming cannot be ruled out [1]. The themes also reflect one researcher's reflexive interpretation.

The three author-led prototype walkthroughs were non-independent, and Project Room was not participant-tested. Future work should compare matched tasks, add evaluators, standardise moderator records, and recruit varied expectations and access needs. Participant studies should evaluate each direction separately alongside accessibility, fairness, privacy, and role relevance [6, 7, 35].

Declaration of Interests

The authors declare no competing interests. The platforms examined in this study were selected independently for research purposes. The study was not commissioned, funded, or supported by the platform providers. The authors have no commercial relationship with, financial interest in, or affiliation with the organisations that developed or operate these platforms.

7 Conclusion

This study examined how four early-career participants experienced scaffolding, agency, and professional realism in one selected Forage case. Professional realism varied with prior expectations and task content; scaffolding supported clarity but could limit active participation; contextual support was hard to discover in the prompted task; and transactional value could remain when practice felt shallow. The RQ is answered through these experiences: participants valued clear progression, but reported limited agency or professional realism when activities left little to interpret, decide, or produce.

The paper's primary contribution is this bounded completion–agency–professional-realism account. Productive friction provides a tentative interpretive lens, and the proposed interaction loop translates the findings into three unevaluated design directions: reversible user-controlled scaffolding, layered contextual support, and consequential feedback and recovery. Project Room illustrates these directions; it does not validate them. With four participants, one selected platform case, different modules across phases, and no participant evaluation of the design work, future studies must test whether the directions preserve judgement while remaining accessible, fair, and safe.

References

[1] Obead Alhadreti and Pam Mayhew. 2017. To Intervene or Not to Intervene: An Investigation of Three Think-Aloud Protocols in Usability Testing. *Journal of Usability Studies* 12, 3 (2017), 111–132. https://ueaeprints.uea.ac.uk/id/eprint/64914/

[2] Elisabeth Bauer, Nicole Heitzmann, and Frank Fischer. 2022. Simulation-Based Learning in Higher Education and Professional Training: Approximations of Practice through Representational Scaffolding. *Studies in Educational Evaluation* 75 (2022), 101213. doi:10.1016/j.stueduc.2022.101213

[3] Bradford S. Bell, Adam M. Kanar, and Steve W. J. Kozlowski. 2008. Current Issues and Future Directions in Simulation-Based Training in North America. *The International Journal of Human Resource Management* 19, 8 (2008), 1416–1434. doi:10.1080/09585190802200173

[4] Virginia Braun and Victoria Clarke. 2023. Toward Good Practice in Thematic Analysis: Avoiding Common Problems and Be(com)ing a Knowing Researcher. *International Journal of Transgender Health* 24, 1 (2023), 1–6. doi:10.1080/26895269.2022.2129597

[5] Olga Chernikova, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. 2020. Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research* 90, 4 (2020), 499–541. doi:10.3102/0034654320933544

[6] Jade Davies, Brett Heasman, Adam Livesey, Amy Walker, Elizabeth Pellicano, and Anna Remington. 2023. Access to Employment: A Comparison of Autistic, Neurodivergent and Neurotypical Adults' Experiences of Hiring Processes in the United Kingdom. *Autism* 27, 6 (2023), 1746–1763. doi:10.1177/13623613221145377

[7] Alessandro Fabris, Nina Baranowska, Matthew J. Dennis, David Graus, Philipp Hacker, Jorge Saldivar, Frederik Zuiderveen Borgesius, and Asia J. Biega. 2025. Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey. *ACM Transactions on Intelligent Systems and Technology* 16, 1, Article 16 (2025), 54 pages. doi:10.1145/3696457

[8] Ruth M. Fanning and David M. Gaba. 2007. The Role of Debriefing in Simulation-Based Learning. *Simulation in Healthcare* 2, 2 (2007), 115–125. doi:10.1097/SIH.0b013e3180315539

[9] Linda Finlay and Brendan Gough (Eds.). 2003. *Reflexivity: A Practical Guide for Researchers in Health and Social Sciences.* Blackwell Science. doi:10.1002/9780470776094

[10] Vincenzo Galatro, Ema Di Petrillo, and Alfonso J. López Rivero. 2025. Fricción Productiva: Un Marco Interdisciplinario para el Pensamiento Crítico en Medios Digitales [Productive Friction: An Interdisciplinary Framework for Critical Thinking in Digital Media]. *Journal of Digital Media & Interaction* 8, 20 (2025), 139–147. doi:10.34624/jdmi.v8i20.40139

[11] Alastair C. Graham and Sean McAleer. 2018. An Overview of Realist Evaluation for Simulation-Based Education. *Advances in Simulation* 3, 1 (2018), 13. doi:10.1186/s41077-018-0073-6

[12] Marc Hassenzahl and Noam Tractinsky. 2006. User Experience—A Research Agenda. *Behaviour & Information Technology* 25, 2 (2006), 91–97. doi:10.1080/01449290500330331

[13] Peter Holtz, Joachim Kimmerle, and Ulrike Cress. 2018. Using Big Data Techniques for Measuring Productive Friction in Mass Collaboration Online Environments. *International Journal of Computer-Supported Collaborative Learning* 13 (2018), 439–456. doi:10.1007/s11412-018-9285-y

[14] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Transactions on Computer-Human Interaction* 19, 3, Article 23 (2012), 18 pages. doi:10.1145/2362364.2362371

[15] Chris Jacobs, Georgia Foote, and Michael Williams. 2023. Evaluating User Experience with Immersive Technology in Simulation-Based Education: A Modified Delphi Study with Qualitative Analysis. *PLOS ONE* 18, 8 (2023), e0275766. doi:10.1371/journal.pone.0275766

[16] Anukriti Kumar. 2026. Beyond Frictionless: Designing Productive Friction into Augmented Reading Interfaces. CHI '26 STAR Workshop position paper. Position paper; no empirical system evaluation reported. https://chi-star-workshop.github.io/src/assets/pdf/papers/BeyondFrictionless_Designing_Productive_Friction_into_Augmented_Reading_Interfaces%20-%20Anukriti%20Kumar.pdf

[17] Fatimah Lateef. 2010. Simulation-Based Learning: Just Like the Real Thing. *Journal of Emergencies, Trauma, and Shock* 3, 4 (2010), 348–352. doi:10.4103/0974-2700.70743

[18] Maude Lavanchy, Patrick Reichert, Jayanth Narayanan, and Krishna Savani. 2023. Applicants' Fairness Perceptions of Algorithm-Driven Hiring Procedures. *Journal of Business Ethics* 188 (2023), 125–150. doi:10.1007/s10551-022-05320-w

[19] Franziska Leutner, Sonia-Cristina Codreanu, Saskia Brink, and Theodoros Bitsakis. 2023. Game Based Assessments of Cognitive Ability in Recruitment: Validity, Fairness and Test-Taking Experience. *Frontiers in Psychology* 13 (2023), 942662. doi:10.3389/fpsyg.2022.942662

[20] Alice C. Li, Joseph L. Kannry, Andre Kushniruk, Dillon Chrimes, Thomas G. McGinn, Daniel Edonyabo, and Devin M. Mann. 2012. Integrating Usability Testing and Think-Aloud Protocol Analysis with "Near-Live" Clinical Simulations in Evaluating Clinical Decision Support. *International Journal of Medical Informatics* 81, 11 (2012), 761–772. doi:10.1016/j.ijmedinf.2012.02.009

[21] Danijela Magdalenić and Ljerka Luić. 2025. Assessing the Impact of Digital Tools on the Recruitment Process Using the Design Thinking Methodology. *Administrative Sciences* 15, 4 (2025), 139. doi:10.3390/admsci15040139

[22] Jakob Nielsen. 2024. *10 Usability Heuristics for User Interface Design.* Originally published 24 April 1994. https://www.nngroup.com/articles/ten-usability-heuristics/

[23] Ioannis Nikolaou, Konstantina Georgiou, and Vasiliki Kotsasarlidou. 2019. Exploring the Relationship of a Gamified Assessment with Performance. *The Spanish Journal of Psychology* 22 (2019), e6. doi:10.1017/sjp.2019.5

[24] Stavroula Ntoa. 2025. Usability and User Experience Evaluation in Intelligent Environments: A Review and Reappraisal. *International Journal of Human–Computer Interaction* 41, 5 (2025), 2829–2858. doi:10.1080/10447318.2024.2394724

[25] Soyoung Park. 2022. A Study on Visual Scaffolding Design Principles in Web-Based Learning Environments. *The Electronic Journal of e-Learning* 20, 2 (2022), 180–200. doi:10.34190/ejel.20.2.2604

[26] Dijana Plantak Vukovac, Valentina Kirinic, and Bozidar Klicek. 2010. A Comparison of Usability Evaluation Methods for E-Learning Systems. In *DAAAM International Scientific Book 2010*, Branko Katalinic (Ed.). DAAAM International Vienna, 271–288. doi:10.2507/daaam.scibook.2010.27

[27] Pedro J. Ramos-Villagrasa and Elena Fernández-del Río. 2023. Predictive Validity, Applicant Reactions, and Influence of Personal Characteristics of a Gamefully Designed Assessment. *Journal of Work and Organizational Psychology* 39, 3 (2023), 169–178. doi:10.5093/jwop2023a18

[28] Pedro J. Ramos-Villagrasa, Elena Fernández-del Río, Ramón Hermoso, and Jorge Cebrián. 2024. Are Serious Games an Alternative to Traditional Personality Questionnaires? Initial Analysis of a Gamified Assessment. *PLOS ONE* 19, 5 (2024), e0302429. doi:10.1371/journal.pone.0302429

[29] Diana R. Sanchez, Amanda Rueda, Kentaro Kawasaki, Saar Van Lysebetten, and Daniel Diaz. 2023. Reviewing Simulation Technology: Implications for Workplace Training. *Multimodal Technologies and Interaction* 7, 5 (2023), 50. doi:10.3390/mti7050050

[30] Jessica Schick and Sebastian Fischer. 2021. Dear Computer on My Desk, Which Candidate Fits Best? An Assessment of Candidates' Perception of Assessment Quality When Using AI in Personnel Selection. *Frontiers in Psychology* 12 (2021), 739711. doi:10.3389/fpsyg.2021.739711

[31] Priya Sharma and Michael J. Hannafin. 2007. Scaffolding in Technology-Enhanced Learning Environments. *Interactive Learning Environments* 15, 1 (2007), 27–46. doi:10.1080/10494820600996972

[32] John Sweller. 1988. Cognitive Load during Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (1988), 257–285. doi:10.1207/s15516709cog1202_4

[33] Estela Aparecida Oliveira Vieira, Aleph Campos da Silveira, and Ronei Ximenes Martins. 2019. Heuristic Evaluation on Usability of Educational Games: A Systematic Review. *Informatics in Education* 18, 2 (2019), 427–442. doi:10.15388/infedu.2019.20

[34] Prokopia Vlachogianni and Nikolaos Tselios. 2022. Perceived Usability Evaluation of Educational Technology Using the System Usability Scale (SUS): A Systematic Review. *Journal of Research on Technology in Education* 54, 3 (2022), 392–409. doi:10.1080/15391523.2020.1867938

[35] Colin Willis, Tracy Powell-Rudy, Kelsie Colley, and Joshua Prasad. 2021. Examining the Use of Game-Based Assessments for Hiring Autistic Job Seekers. *Journal of Intelligence* 9, 4 (2021), 53. doi:10.3390/jintelligence9040053

[36] Po-Yin Yen and Suzanne Bakken. 2009. A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol—An Example from a Web-Based Communication Tool for Nurse Scheduling. In *AMIA Annual Symposium Proceedings*, Vol. 2009. 714–718.

A Appendices

A.1 Introductory Interview Questions

(1) Could you tell me a little bit about what you're studying and what year you're in?

(2) What are your general career aspirations after you graduate?
(3) Before today, have you ever heard of or used the Forage platform?
(4) Have you used any other online learning or career development platforms like Coursera, LinkedIn Learning, or anything similar?
(5) Before you started today, what were your initial thoughts or expectations for a "job simulation" on a platform like this?

A.2 Debrief Interview Questions

(1) You just completed the *Bloomberg Essentials* module, which the site stated takes about 13 minutes. How did this specific experience align with or differ from the initial expectations you had?
(2) Thinking back on the entire process, can you describe any particular high points or low points you encountered?
(3) What was your experience of the step where you had to create an account before you could access the content?
(4) A key challenge in designing these platforms is balancing realism with clarity. How would you describe that balance in the module you just completed? Was there any point where it felt too complex, or perhaps too simplified?
(5) This platform is designed to provide experiences that feel like real-world job tasks. In your opinion, did the Bloomberg module feel more like an authentic work task, a structured university lesson, or perhaps something else, like company marketing?
(6) Thinking about what you actually did, what specific skills, if any, do you feel you practised or learned today? How confident would you be in applying what you learned in a real-world interview or task?
(7) Regarding the platform design, beyond just being easy or difficult, what aspects made the experience more engaging or motivating? Were there any elements that made you lose interest?
(8) Based on your experience, would you recommend the Forage platform and this specific module to a friend or colleague?
(9) What do you think is the ideal user for the specific module that you did? Who would get the most value out of it?
(10) Is there anything else that you would like to share about your experience today?