Multimodal Technologies and Interaction

Article

A Conceptual Framework for Mobile Augmented-Reality Storytelling to Support Collaborative Language Learning in Vocational Education and Training

Eirini Maria Paraskevioti [1], Athanasios Christopoulos [1,*], Stylianos Mystakidis [2], Mikko-Jussi Laakso [1] and Tapio Salakoski [1]

[1] Faculty of Science, Turku Research Institute for Learning Analytics, University of Turku, 20500 Turku, Finland; eirini.m.paraskevioti@utu.fi (E.M.P.); milaak@utu.fi (M.-J.L.); tapio.salakoski@utu.fi (T.S.)
[2] School of Natural Sciences, University of Patras, 26504 Patras, Greece; smyst@upatras.gr
[*] Correspondence: atchri@utu.fi

Abstract

Augmented Reality (AR) has been found to produce significant effects on individual learning outcomes but its impact on collaborative applications remains moderate. Existing AR frameworks emphasize individual instructional design, whereas frameworks for collaborative learning rarely engage with the spatial and device-mediated affordances of mobile AR. In response to this inadequacy in the literature, we introduce the Mobile Augmented-Reality Storytelling for Vocational Education and Training (MARS-VET) framework, a four-dimensional conceptual architecture that integrates Computer-Supported Collaborative Learning (CSCL) scripting principles with mobile AR affordances for collaborative English as a Foreign Language (EFL) writing in Vocational Education and Training (VET) settings. MARS-VET synthesizes theoretical perspectives across four dimensions: contextual anchoring, which embeds activities within authentic workplace scenarios; collaborative orchestration, which structures group interaction through macro- and micro-level scripts; competency cultivation, which sequences writing progression from model-based reproduction toward autonomous professional text production; and capacity building, which addresses the professional-development requirements of implementing educators. Content validity was established through expert panel evaluation involving international specialists (N = 11) who rated the framework against 36 items using a four-point relevance scale and provided additional qualitative feedback. The Scale-level Content Validity Index (S-CVI/Ave = 0.91) exceeded established thresholds, with all four dimensions achieving satisfactory item-level indices. Experts reached unanimous agreement on items addressing workplace scenario identification and co-located access to linguistic resources. Qualitative feedback led to terminology refinements and clarification of orchestration mechanisms. The framework offers VET institutions and educators a reference for the design and evaluation of collaborative AR experiences in an area where integrative frameworks have so far been lacking.

Keywords: augmented reality; mobile learning; collaborative learning; vocational education and training; CSCL; EFL/ESL; content validity; conceptual framework

Academic Editors: Nikleia Eteokleous and Rita Panaoura

Received: 9 April 2026
Revised: 30 April 2026
Accepted: 7 May 2026
Published: 11 May 2026

1. Introduction

The demand for workplace communication competencies in globalized economies has placed pressure on Vocational Education and Training (VET) systems to prepare learners for professional environments where both technical expertise and collaborative skills are expected [1]. Writing technical reports, handling customer correspondence, and documenting incidents are communicative tasks that VET graduates must be able to perform competently and often in a second language. In view of these needs, educational technologists have turned to immersive technologies as a means to bridge the gap between classroom instruction and authentic workplace practice [2]. The challenge, however, is not purely technological, as it concerns how such technologies are integrated with the pedagogical principles that can make them effective for collaborative and communicative skill development.

Evidence from recent meta-analyses confirms that AR has clear educational value. Howard and Davis [3] documented moderate positive effects across instructional contexts while Li et al. [4], who analyzed experimental studies in Higher Education, reported large effects. Amongst the key findings of both studies is the fact that mobile devices produced larger effect sizes than studies which involved specialized equipment (e.g., Head-Mounted Displays), an observation which supports cost-effective implementation in resource-constrained VET institutions. Notwithstanding the foregoing considerations, when collaborative applications were examined separately, the effect was more moderate than that observed for direct instruction—a difference which should be interpreted with care since direct instruction optimizes individual knowledge acquisition whereas collaborative learning targets social and communicative competencies which are substantially harder to assess.

The aforementioned inadequacy of the literature points to an implementation problem rather than a technological one. For instance, the study of Li et al. [4] found that content delivery dominated AR instructional functions at 51%, with collaborative applications being underrepresented. In the same vein, Buchner and Kerres [5] observed that 80% of AR studies employ media-comparison designs instead of investigating the specific pedagogical conditions under which collaboration thrives. Kazlaris et al. [6] identified further methodological gaps including inconsistent evaluation approaches for collaborative outcomes. Across the abovementioned findings the same problem recurs—when collaboration occurs in AR environments it is seldom designed with the deliberateness that decades of collaborative learning research recommend.

A proposed solution to address this challenge comes from Computer-Supported Collaborative Learning (CSCL) research wherein structured collaboration—implemented through interaction scripts with explicit role assignments, turn-taking sequences, and prompts for substantive discourse—has been found to outperform unstructured group work [7]. Fischer et al. [8] developed a script theory of guidance that articulates how scripts can scaffold productive interaction without over-constraining learner autonomy. However, AR research has not incorporated—at least systematically—these principles, which thus leaves educators who seek to implement collaborative AR without design guidance [1,6].

Mobile AR has also been found to be effective for language learning. Wu et al. [9] identified positive effects, with mobile devices being used in nearly 9 out of 10 of the reviewed studies. Vocabulary acquisition, however, dominates the field—as opposed to writing skills and genre-based professional communication which have received limited attention—a shortcoming which negatively affects VET contexts where learners need competence in profession-specific written genres such as technical reports, business correspondence, and incident documentation. Carrión-Robles et al. [10] found that AR enhanced paragraph organization, structure, grammar, and vocabulary which led them to suggest that mobile AR storytelling (i.e., the combination of spatial narratives with collaborative writing tasks) could address gaps in both collaborative AR research and AR-supported writing instruction. Digital storytelling has already shown its potential as a pedagogical approach for language learning as it offers opportunities for learners to engage with narrative structures while developing communicative competencies [11]. When combined with AR, storytelling gains a spatial dimension that allows narratives to unfold within and respond to physical environments—a quality which is very relevant for VET contexts where the physical workplace itself constitutes part of the learning content.

Motivated by this convergence of opportunities, in the present work, we introduce the Mobile Augmented-Reality Storytelling for Vocational Education and Training (MARS-VET) framework, a conceptual architecture that integrates CSCL script principles with mobile AR affordances for collaborative VET language learning. The overarching goal comprises the objectives of the present work, which are fourfold: (1) to synthesize the theoretical foundations from AR affordance research, CSCL script theory, embodied cognition, and situated learning for collaborative mobile AR in vocational language learning; (2) to propose a conceptual framework that organizes design considerations across four dimensions—contextual anchoring, collaborative orchestration, competency cultivation, and capacity building; (3) to gather expert feedback; and (4) to articulate implementation considerations and feasibility constraints that practitioners and institutional leaders should account for during deployment planning.

The selection of the aforementioned dimensions—as opposed to three or five, or some alternative configuration—reflects a deliberate mapping of the principal design concerns identified across the convergent literature. AR research has highlighted contextuality as the affordance most directly relevant to situated learning [5,12,13]. CSCL research has demonstrated that productive group interaction does not emerge automatically and, accordingly, calls for deliberate orchestration through scripted collaboration [7,8]. Genre-based pedagogy and writing development research have identified staged competency progression as essential for professional language acquisition [10,11]. Implementation research has, in turn, recognized educator capacity as a critical determinant of sustained adoption [14]. Alternative configurations were considered in the course of the framework's development—the merging of capacity building into collaborative orchestration, for instance, or the treatment of competency cultivation as a sub-component of contextual anchoring—but were ultimately rejected on the grounds that they would conflate design considerations that practitioners commonly address as separate.

MARS-VET occupies a distinctive position in the broader landscape of relevant frameworks. The AR affordances framework [5] is technology-centered whereas the ARLEAN framework [15] prioritizes Learning Analytics; in either case, collaborative learning is treated as a peripheral concern. By comparison, MARS-VET places collaboration at the center of the design space and treats it as a primary affordance. The framework introduced by Upadhyay et al. [1] provides a broad mapping of collaborative AR across disciplines, whereas MARS-VET narrows the focus to the specific intersection of mobile AR, VET, and professional language development—an intersection that has received unsystematic attention in the pertinent literature. Closer to the VET context, Bacca et al. [16] proposed a framework for the design of motivational AR applications in VET, drawing on the ARCS model of motivation and on universal design for learning. The aforementioned framework, however, addresses individual learner motivation and does not articulate guidance for the structuring of collaborative interactions. CSCL script theory [7,8] articulates the principles of structured collaboration in a generic sense but has rarely engaged with the spatial, attentional, and device-mediated properties of mobile AR. The collaborative orchestration dimension of MARS-VET adapts these principles to that medium.

In view of the foregoing considerations, the contribution of the present work is threefold. First, AR affordance research, CSCL script theory, situated learning, and embodied cognition are integrated into a coherent conceptual structure for the specific intersection of mobile AR, collaborative learning, VET, and professional language development—an intersection that has so far received only peripheral attention. Second, concrete design guidance—organized across four dimensions—is provided to VET educators and instructional designers for an area where design frameworks for collaborative mobile AR have been notably lacking. Third, initial evidence of the framework's content validity is established through expert consultation, with the components that call for further empirical refinement being identified and proposed as future work. The practical implication of the foregoing design is that institutions can rely on the four-dimensional structure as a basis for the planning, sequencing, and evaluation of collaborative AR initiatives instead of resorting to ad hoc adaptations of frameworks designed for other contexts. It should be noted, though, that the framework has been evaluated for content validity through expert consultation but has not been empirically tested as an intervention. Therefore, the design considerations which are presented herein should be understood as theoretically grounded proposals at this stage.

2. Theoretical Foundations

None of the theoretical perspectives that inform MARS-VET (i.e., AR affordances, CSCL scripts, embodied cognition, situated learning) are themselves new. CSCL scripts have been examined across technology contexts [7,8], situated learning has been integrated into technology-enhanced instructional frameworks [12], embodied cognition has been studied in immersive technology environments [17,18], and AR affordances have been mapped onto general learning theories [15]. The contribution of MARS-VET, therefore, lies in the extension of these perspectives to an intersection—mobile AR, collaborative learning, vocational education, and professional language development—that, despite its practical relevance, has received limited and unsystematic attention in the pertinent literature.

Section 2 proceeds in four steps. AR's technological affordances for learning are reviewed in Section 2.1 with particular attention to the underrepresentation of collaboration in the AR affordance literature. The discussion turns in Section 2.2 to CSCL scripting which constitutes a more deliberate response to the foregoing gap. Embodied cognition (Section 2.3) and situated learning (Section 2.4) are subsequently introduced as a means of clarifying the conditions under which AR's spatial properties can translate into learning gains. The section closes with a synthesis (Section 2.5) that articulates how the four perspectives jointly inform the four dimensions of the framework and that explains why alternative theoretical groundings—although available—were not adopted as primary lenses.

2.1. Technological Affordances of AR for Learning

The literature distinguishes three core AR affordances: (1) visualization, which refers to the capacity to render abstract concepts perceptible through digital overlays; (2) interactivity, which encompasses the manipulation possibilities that AR provides such as touch, gesture, or voice commands; and (3) contextuality, which represents the capacity to anchor digital content within physical space, thereby reducing cognitive load [5]. The aforementioned affordances are of particular interest for VET given that the competencies acquired must transfer to professional practice. Notwithstanding the foregoing affordances, a fourth candidate affordance—user-based social collaboration—has received limited attention in the AR affordance literature, an asymmetry whose theoretical implications are taken up in Section 2.5.

The aforementioned disparity is mostly descriptive, as AR affordance taxonomies that omit collaboration shape what designers attend to. When 'visualization, interactivity, and contextuality' constitute the canonical triad, social-collaboration design becomes an additional consideration instead of a primary affordance—an asymmetry that helps to explain why media-comparison studies dominate the AR literature [5,6] whereas the orchestration of group interaction has remained underspecified. The treatment of user-based social collaboration as a first-order affordance, as we propose here, is therefore both a theoretical adjustment and a design commitment, the consequences of which become apparent in the components introduced in Section 4.

In a related field, Van der Meer et al. [19] reviewed the literature on collaborative immersive technologies and reported several affordances, among them efficient engagement, remote collaboration support, learner control, and the development of social skills. Their focus, however, was on Virtual Reality (VR) instead of mobile AR which differs in ways that may constrain collaborative affordances (e.g., the small screen size limits shared viewing, device-holding restricts physical collaboration, the overlay of digital content on physical environments creates attentional demands that differ from those of full immersion). In the context of collaborative learning, such constraints are non-trivial since they mean that not all group members will have the same visual access to the augmented content at the same time—a limitation which has design implications for how roles and responsibilities are distributed within learner groups. Despite the constraints, Howard and Davis [3] found that mobile devices produced larger effect sizes than Head-Mounted Displays, a finding which supports cost-effective implementation in resource-constrained VET institutions.

2.2. Computer-Supported Collaborative Learning Scripts

The key findings of empirical studies suggest that productive collaborative learning does not emerge automatically when learners are placed in groups but rather requires deliberate structuring so that the interactions generate cognitive benefits [7,8]. In line with this evidence, scripted collaboration has been found to outperform unscripted group work with strong effects for domain-specific knowledge acquisition [7]. CSCL scripts operate at two complementary levels: (1) the macro-level scripts, which define the overall structure including phase sequences, group composition, and activity transitions and (2) the micro-level scripts, which specify the interaction patterns including role assignments, turn-taking rules, and prompts for transactive discourse. With regard to the former, Tahirbegi and Järvenoja [20] showed that macro-level scripts grounded in cyclical models of regulated learning enhanced socio-emotional awareness among learners, whereas Özbek et al. [21] examined how different goals affect students' internal collaboration script configurations and found that the goal structure influences how learners coordinate their collaborative behavior—a finding whose relevance for the staged scripting decisions of the present work is taken up in Section 4.2. Similarly, as far as collaborative writing is concerned, Ouyang et al. [22] found that idea-centered argumentation-oriented scaffolding showed the strongest effects.

Two empirical refinements deserve attention. To begin with, scripts that supply explicit transactivity prompts have been found to outperform scripts that supply role assignments alone, with the reason being that the cognitive work of 'building on a partner's contribution' does not emerge spontaneously from the assignment of distinct roles alone [7,22]. In addition, the time horizon of script effects appears to matter—short-term gains in coordination are observable within single sessions, whereas effects on collaborative competence as such call for sustained scripting across multiple sessions [23]. Both findings have informed the macro/micro and scaffolded-progression decisions presented in Section 4.2.

Nevertheless, the CSCL literature also documents concerns about over-scripting, as overly prescriptive scripts may constrain natural collaborative processes, reduce learner autonomy, and undermine motivation—a claim which is further supported from a self-determination perspective wherein excessive external structure can interfere with a learner's sense of autonomy and competence [8,24,25]. Script effectiveness also appears to be context-dependent, since scripts which benefit novice learners may constrain experts and scripts that are effective in one cultural setting may require adaptation elsewhere. To address these concerns, scaffolded fading (i.e., the gradual removal of script elements as learners internalize collaborative competencies) has been proposed [8] with evidence from comparison studies supporting this position. For instance, Strauß et al. [23] compared the effects of collaboration scripts with collaborative reflection and found that both approaches promoted knowledge about good collaboration though through different mechanisms. It can be thus suggested that the relationship between scripting and collaborative competence is more conditional than it first appears and that fading scripts does not necessarily mean losing their benefits if learners have internalized the underlying principles.

2.3. Embodied Cognition

Embodied cognition proposes that thinking is grounded in bodily experience and sensorimotor interaction—with physical movement, gesture, and spatial exploration constituting the mechanisms through which understanding develops [17,18]. Research supports this claim for particular task types; Liu et al. [17] found a moderate positive effect, with high-level embodiment that involves whole-body movement significantly outperforming low-level embodiment—such as mouse clicks—whereas Zhang et al. [18] reported similar effects and concluded that both AR and VR enable learners to interconnect physical and virtual spaces. Stalheim and Somby [26] further explored this connection in the context of AR in schools and found that bodily engagement with augmented content created meaningful learning experiences that went beyond what 'screen-based interaction' alone could offer.

The evidence, however, for embodied learning relates primarily to spatial and physical tasks—mental rotation, spatial reasoning, procedural skills—but not to writing or language production. The proposition that physical manipulation of text elements in AR environments (e.g., arranging, sequencing, spatially organizing ideas) might support writing development lacks direct empirical support. The implication for design is that embodied lenses can serve as a supplementary ground for writing-focused activities with genre pedagogy retained as the primary lens for writing development (Section 4.3).

2.4. Situated Learning and Contextual Anchoring

Situated-learning theory holds that knowledge acquisition is tied to the context in which learning occurs and that transfer is enhanced when instruction takes place in authentic environments. For vocational education—where the goal is to prepare learners for specific professional practices—this principle is especially relevant given that competencies developed in abstract settings may fail to transfer to workplace demands. In this regard, the contextuality affordance discussed in Section 2.1 provides the mechanism through which situated-learning principles can be operationalized. Schorr et al. [13] stressed contextual and situated learning as primary benefits of AR for language education, whereas Christou et al. [12], who reviewed a broader range of AR language learning studies, confirmed that situated learning was the most frequently adopted theoretical framework in the field. In the context of VET language learning, the added value of this approach lies in the fact that professional vocabulary, genre conventions, and communicative strategies can be presented within the same context they would be applied (e.g., at the hotel reception desk, in the logistics warehouse, within the healthcare facility) thereby allowing learners to encounter linguistic resources where they serve an actual communicative function—an arrangement which can support both comprehension and retention as the physical environment itself provides retrieval cues that anchor the learning experience.

Situated learning in simulated contexts, such as AR-enhanced classroom spaces, may not transfer identically to authentic workplace settings. AR can approximate workplace scenarios, but the social dynamics, time pressures, emotional stakes, and unpredictability of actual professional practice differ from educational simulations. Therefore, the assumption that AR-situated learning will transfer to workplace practice is theoretically grounded but remains to be verified.

2.5. Theoretical Synthesis

Taken together, the four perspectives reviewed above frame the design space that MARS-VET addresses. Situated learning and AR's contextuality affordance (Sections 2.1 and 2.4) account for where authentic vocational learning can take place; CSCL scripting (Section 2.2) accounts for how productive group interaction may be structured; and embodied cognition (Section 2.3) accounts for the bodily and spatial dimension that mobile AR introduces. Each perspective, however, has its limits. AR affordance research has paid limited attention to collaboration as such; CSCL theory has rarely engaged with mobile AR as a medium; situated learning does not specify the structuring of group interaction; and embodied cognition currently offers limited evidence in the context of writing or language production. The four perspectives are, therefore, complementary and no single perspective is, on its own, sufficient to meet the design challenge at hand.

Alternative theoretical groundings were considered in the course of the framework's development. Activity Theory could have provided a unifying lens but operates at a level of abstraction that is too high for direct guidance at the dimension and component level. Likewise, sociocultural theories of language learning would have offered a richer account of writing development; however, this strength would have come at the expense of the technological affordances that distinguish AR from other learning environments. Self-determination theory was retained as a corrective to the over-scripting concerns discussed in Section 2.2 but was not adopted as a primary lens. The four perspectives selected here were chosen on the grounds that, in combination, they map onto the four design concerns that practitioners must address: where the activity occurs, how interaction is structured, how competence develops, and how educators are supported. It can be argued that the foregoing mapping constitutes a design decision instead of a logical entailment. Alternative configurations remain possible with the framework being open to revision as empirical evidence accumulates.

3. Materials and Methods

The MARS-VET framework was developed and subsequently validated in two sequential phases. The first concerned the construction of the framework through structured synthesis of the pertinent literature; the second concerned content validation through expert consultation. The methodological choices that underpin both phases are summarized in the present section with the detailed validation results being reported in Section 5.

3.1. Framework Construction Methodology

The framework was developed through a four-step procedure. First, a scoping review of the literature on AR in education, CSCL, mobile language learning, and VET was conducted with the aim of identifying recurring design concerns and unresolved gaps. Second, the design concerns thus identified were grouped thematically through iterative clustering, an exercise that converged on four orthogonal categories—context, interaction structure, competency progression, and educator capacity—which subsequently became the candidate dimensions of the framework. Third, each dimension was elaborated into components and sub-components by mapping theoretical principles onto operational design choices. The construction process was guided by two principles: theoretical economy, in the sense that no design element was introduced without explicit theoretical or empirical justification;

and practitioner addressability, in the sense that every component had to be something a practitioner could act upon. Fourth, the resulting draft was internally reviewed by the authoring team—drawing on collective experience in immersive technology research and educational technology design [27,28]—and was iteratively refined.

3.2. Validation Approach

Content validity was selected as the appropriate form of evidence for a conceptual contribution because it establishes whether the proposed components are relevant, clearly articulated, and collectively coherent [29]. Empirical validation through implementation studies was, by contrast, deferred to subsequent work, on the grounds that intervention research presupposes a stable framework specification. The validation procedure followed the recommendations made by Polit and Beck [29]. The validation instrument was developed by the authoring team and was pilot-tested with two educational technology researchers prior to its distribution; it comprises 36 rating items and four open-ended questions across five sections (the full instrument is provided in Appendix A). The detailed procedures, panel composition, and results are reported in Section 5.

3.3. Ethical Considerations

The validation study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. In accordance with the Finnish national research ethics guidelines, the study did not require formal ethics committee review, as it consisted only of standard psychometric assessments with adult, professional participants. All invited experts were informed about the objectives of the study, the procedures, the data handling, and the confidentiality safeguards in place. Informed consent was obtained from all participants prior to their participation. Responses were stored on password-protected institutional servers and were analyzed in aggregate.

4. Framework Overview

The four dimensions of MARS-VET are not sequential stages of an instructional design process; they are mutually reinforcing design concerns that should be addressed jointly during planning, with the relative emphasis on each calibrated to local conditions (Figure 1). Contextual anchoring provides the authentic scenarios within which collaboration unfolds; collaborative orchestration structures the interactions through which language competencies develop; competency cultivation sequences the progression from model-based reproduction toward autonomous professional text production; and capacity building ensures that the educators who implement the framework have the technical and pedagogical expertise which sustained implementation requires.

Each dimension is anchored to a primary theoretical lens drawn from Section 2: contextual anchoring to situated-learning theory and AR contextuality, collaborative orchestration to CSCL script theory, competency cultivation to genre-based pedagogy, and capacity building to professional-development research. Each dimension is, in turn, operationalized through three components.

4.1. Dimension 1: Contextual Anchoring

Contextual anchoring is the dimension that—drawing on situated-learning theory [12,13] and AR's contextuality affordance [5]—specifies how collaborative language learning activities can be embedded within authentic or simulated workplace scenarios. The dimension comprises three interrelated components: workplace scenario identification, AR content development, and domain-specific activity design. As the pertinent literature suggests, knowledge developed in decontextualised settings often fails to transfer to real-world application [12,13].

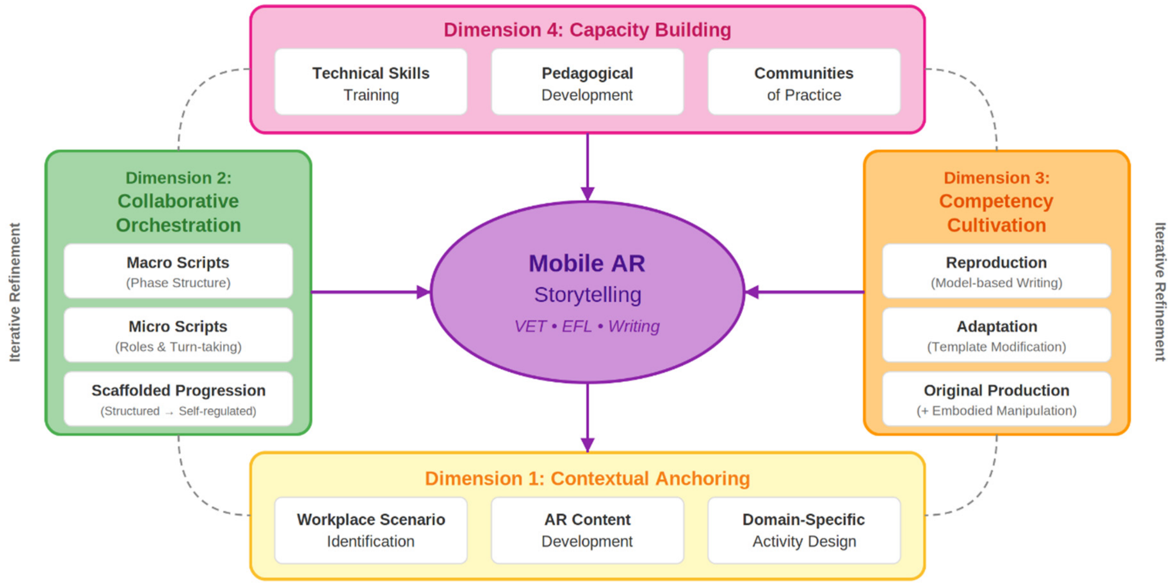

Figure 1. Overview of the MARS-VET framework.

Scenario selection should be guided by three primary criteria: (1) curricular relevance, which ensures alignment with VET learning outcomes; (2) spatial accessibility, which considers the availability of physical spaces for AR deployment; and (3) linguistic richness, which evaluates whether scenarios provide sufficient opportunities for meaningful language use across multiple genres and registers. In practice, the choice of scenario varies across VET fields; for instance, hospitality programs might use hotel reception areas for complaint handling correspondence, whereas healthcare programs could focus on patient intake areas for incident documentation and business programs on customer service centers for supplier correspondence. It is recommended that scenario selection prioritizes those contexts which learners are likely to encounter in their professional lives so as to maximize the potential for transfer. When access to an authentic workplace is not feasible—which is often the case in the VET sector due to logistical or safety constraints—educators can resort to alternatives such as classroom spaces that have been configured to approximate workplace layouts, institutional common areas, or simulation laboratories. The trade-off between authenticity and repeatability that these alternatives introduce should be weighed during the planning phase.

Drawing on the taxonomy of 'augmented narrative spaces' proposed by Shin et al. [30], we specify four primary types of AR content for use within the framework. The first, text annotations, involves the positioning of vocabulary support, genre conventions, and linguistic scaffolds at relevant locations such as professional terminology overlaid on equipment or register guidance placed alongside customer interaction points. The second, three-dimensional objects, enables learners to examine physical items that they will encounter in their profession and must learn to describe; the ability to view these objects from multiple perspectives supports vocabulary acquisition through visual–verbal association. The third, video demonstrations, provides modeled examples of professional communication practices which learners can observe and subsequently analyze. The fourth, interactive elements, allows learners to manipulate content by selecting options, sequencing steps, or assembling components. It should be acknowledged, however, that the development of such content requires technical expertise in 3D modeling, AR platforms, and UX design—which may not be available within VET institutions. As such, institutions should assess their capacity and, where necessary, seek partnerships with technical specialists.

Each session follows a three-phase activity sequence. First, there is an exploration phase wherein learner groups navigate the augmented environment, discover the available content, and—most importantly—build a shared understanding of the scenario context; without common ground, the collaborative language tasks that follow lose their foundation. Second, a language focus phase where the groups examine the linguistic features of professional texts that are relevant to the scenario at hand with attention to genre conventions, vocabulary, and register. Third, a production phase in which the groups collaboratively compose texts that respond to the demands of the scenario by drawing on the contextual knowledge and the linguistic resources encountered in the preceding phases.

4.2. Dimension 2: Collaborative Orchestration

Collaborative orchestration is the dimension that translates CSCL script theory [7,8] into guidance for the structuring of group interactions within mobile AR environments, with attention to the spatial and device-mediated properties that distinguish mobile AR from other learning media. The dimension comprises three components: macro-level scripts (which define phase structure), micro-level scripts (which specify roles and interaction patterns), and scaffolded progression (which guides the gradual release of collaborative responsibility). As discussed above, collaborative learning in AR contexts has shown moderate effects—an outcome that suggests that current implementations do not fully exploit AR's collaborative potential.

4.2.1. Macro Scripts: Phase Structure

The proposed three-phase implementation structure draws from the scaffolding literature which has demonstrated the effectiveness of gradually releasing responsibility to learners [8]. Its specific application to mobile AR collaborative learning has not been validated and the phase durations that follow should be understood as starting points—not prescriptions—that require adaptation based on learner needs and institutional context.

- Phase 1 (Introduction) is concerned with establishing foundational collaborative practices. At this stage, roles are explicitly assigned to learners, turn-taking sequences are prescribed, and the interaction protocols are specified. The educator takes an active facilitation role, modeling the expected behaviors and stepping in when groups struggle with the scripted patterns.
- Phase 2 (Development) transitions toward negotiated collaboration, wherein groups determine interaction patterns within educator-specified constraints; role assignments may rotate according to group preferences, turn-taking becomes more flexible while maintaining equitable participation requirements, and groups develop characteristic collaborative styles. The educator's role during this phase shifts from direct facilitation to monitoring and responsive intervention—a transition which is not trivial as it requires educators to develop the ability to identify when a group is struggling productively (which should be allowed to continue) versus when it is stuck unproductively (which requires intervention). In the collective experience with immersive educational technologies [27,28], this judgment is one of the most difficult aspects of facilitation and develops primarily through practice rather than training alone.
- Phase 3 (Consolidation) completes the transition toward self-regulated collaboration where external script elements are systematically faded with groups assuming full responsibility for organizing their interactions and are thus preparing for workplace collaboration contexts where external structuring is typically absent.

4.2.2. Micro Scripts: Roles and Interaction Patterns

Four complementary roles are proposed for AR storytelling activities, each representing one possible configuration that is designed to distribute collaborative responsibilities, with alternative role structures that may be equally effective. The 'navigator' controls the mobile device and directs movement through the augmented environment, deciding which content to access and in what sequence. The 'narrator' verbalizes observations, reads aloud text content, and articulates the emerging story so as to ensure that all group members share access to augmented information—a function which is especially important given the small screen size of mobile devices which limits shared viewing. The 'recorder' takes charge of documenting discoveries, noting linguistic features, and drafting text during production; in effect, the learner acts as the bridge between the group's verbal deliberations and the written output that results from them. The 'critic' has a different but complementary responsibility—to evaluate the proposed content, question assumptions, and check drafts against genre conventions and task requirements. We consider this role to be of great importance because it develops learners' capacity for peer assessment and self-regulation—competencies that are directly transferable to professional contexts where quality assurance is an expectation rather than an exception.

The four-role structure assumes groups of four learners; adaptations are needed for groups of three (combining roles) or five (adding a role or having rotating observers). Similarly, role rotation schedules should ensure that all learners experience each role across the implementation period. When only one mobile device is available per group, the 'navigator' role accumulates disproportionate influence—a dynamic which requires monitoring and which can be mitigated through explicit rotation at predetermined intervals.

4.2.3. Scaffolded Progression

What separates productive scaffolding from counterproductive over-scripting is the capacity to fade scripts: to remove script elements as learners develop the collaborative competencies that render them unnecessary [8,24]. The framework proposes four indicators of readiness for scaffold removal: (1) groups fulfill role responsibilities without prompting; (2) interactions demonstrate a transactive quality—that is, they are contributions that explicitly engage with, build upon, or critique what partners have previously said; (3) written outputs meet genre and accuracy requirements; and (4) group members report confidence in self-organizing collaborative work. Transactive interaction was, in fact, the single item in the expert validation to fall below the acceptability threshold (I-CVI = 0.73), a finding which suggests that operationalizing this construct for practitioners requires special attention. A practical definition was developed in response. Transactive interaction refers to contributions in which a learner explicitly references, extends, or challenges what a group member has previously said as opposed to simply adding unrelated ideas to the conversation. Educators should monitor the readiness indicators and adjust scaffolding levels accordingly while keeping in mind that different groups will progress at different rates.

4.3. Dimension 3: Competency Cultivation

Competency cultivation is the dimension that sequences writing development across three stages—reproduction, adaptation, and original production—each supported by an appropriate level of AR scaffolding and grounded in genre-based pedagogy.

VET language learning is, in many respects, a matter of genre acquisition rather than general communicative competence as learners need to master the standardized text types which are expected in their target professions— such as technical reports, business correspondence, incident documentation (c.f., creative or academic writing), etc. Genre-based pedagogy is well suited to this objective since it sequences writing development through structured engagement with model texts, guided practice in adapting conventions, and eventually autonomous production. The competency cultivation dimension translates this progression into the MARS-VET context using AR's interactivity affordance to provide the contextualized practice environments within which each stage unfolds.

Stage 1: Reproduction (Model-Based Writing)

The concern of Stage 1 is foundational competence in genre conventions and professional register. Learners examine model texts presented within AR environments, analyze structural features, identify key vocabulary, and discuss why specific textual choices serve particular communicative purposes. AR content at this stage provides extensive structural guidance in the form of sentence starters, paragraph templates, and vocabulary banks positioned within relevant scenario contexts. Assessment focuses on accuracy—conformity to expected genre structure, appropriate vocabulary and register, and correct grammar.

Stage 2: Adaptation (Template Modification)

Stage 2 develops the competence to modify templates as contexts, audiences, and purposes shift. Learners are confronted with challenge scenarios that require deliberate adaptation decisions—for instance, a customer complaint that warrants a more formal response than standard templates would provide, an incident report which addresses unusual circumstances, or a business email intended for recipients with different cultural expectations. The collaborative activities at this stage center on discussing modification options, evaluating alternatives, and reaching a consensus on the most appropriate adaptation. Meanwhile, AR content reduces scaffolding and introduces variability through multiple scenario variations that require different textual responses and feedback on appropriateness instead of accuracy alone. Assessment shifts toward appropriateness and flexibility—that is, suitability for the specific context and audience and awareness of how situational variables affect linguistic choices.

Stage 3: Original Production (Autonomous Creation)

In Stage 3, the scaffolding is largely withdrawn and learners must produce professional texts autonomously, supported by peer evaluation cycles that develop critical assessment competencies alongside production skills. Scenarios present novel situations which require original textual responses that cannot be addressed through template adaptation alone, whereas collaborative activities involve joint planning, distributed drafting, collective revision, and peer evaluation against professional standards. AR content at this stage supports the production process—spatial organization tools for collaborative composition, version tracking for iterative revision, and annotation features for peer feedback. Assessment emphasizes communicative effectiveness and professional quality—that is, whether the text achieves its communicative purpose and would be considered acceptable in an actual professional context.

4.4. Dimension 4: Capacity Building

Capacity building is the dimension that addresses the professional-development requirements of educators implementing the framework. Technology provision without educator expertise rarely translates into sustained instructional change [14]; as such, the dimension responds to the foregoing state of affairs through three components: technical-skills training, pedagogical development specific to collaborative AR integration, and the establishment of communities of practice across institutions.

The first component, technical skills training, is aimed at equipping educators with what they need to operate AR applications, troubleshoot device problems, and handle the logistics of mobile technology in a classroom setting. We stress that the goal here is practical competence and not comprehensive technical knowledge. Accordingly, the training should be organized in three stages—initial intensive workshops on device operation, application navigation, content triggering, and troubleshooting; supervised practice wherein educators rehearse facilitation with peer or student groups; and ongoing support that addresses problems as they arise in actual implementation.

The second, pedagogical development, is the more demanding component as it involves understanding how CSCL principles apply when learners are moving through physical space with mobile devices, monitoring multiple groups simultaneously, intervening when collaboration falters without undermining learner autonomy, and adapting activities in response to emergent needs—competencies which conventional teacher training rarely addresses. Collaborative AR facilitation, in particular, presents challenges which are distinct from traditional instruction, such as troubleshooting technical issues without disrupting collaborative flow and distinguishing productive from off-task interaction when AR environments complicate observation.

The third component concerns communities of practice—networks that connect educators who implement the framework across institutions and that enable resource sharing, experience exchange, and collective problem-solving. The caveat, however, is that communities of practice do not materialize on their own; they require critical mass, sustained commitment, and active coordination, all of which demand deliberate investment to establish and maintain. Things get worse in the VET sector where institutions may be geographically dispersed and operate under different governance structures, thus making cross-institutional collaboration difficult. Staff turnover compounds the problem. For instance, when a trained educator leaves, the institutional capacity that they built up over time departs with them—unless knowledge transfer mechanisms have been put in place.

5. Framework Validation

In the literature, content validity is defined as the degree to which a framework adequately represents the domain it purports to address [29]. Given that MARS-VET is a conceptual contribution, expert validation was deemed the appropriate form of evidence as it can establish whether the proposed components are relevant, clearly articulated, and collectively coherent.

5.1. Validation Methodology

The Content Validity Index (CVI) is a quantitative measure of expert agreement on the relevance of framework components, originally developed in nursing research [29] and increasingly applied in educational technology contexts (e.g., [31]). The CVI is computed at two complementary levels. The item-level CVI (I-CVI) is the proportion of experts who rated a given item as relevant—that is, who selected 3 (Quite Relevant) or 4 (Highly Relevant) on the four-point scale. As a worked example, when 9 out of 11 experts rate an item 3 or 4, the corresponding I-CVI is 9/11 = 0.82. The scale-level CVI, calculated as an average (S-CVI/Ave), is the arithmetic mean of all I-CVI values across the instrument; it summarizes the overall content validity of the instrument in a single index. The threshold values that guide interpretation are those recommended by [29]. At the item level, I-CVI values $\geq$ 0.78 indicate acceptable content validity (item retained); values in the range 0.60–0.77 indicate that the item should be revised; values below 0.60 indicate that the item should be removed. At the scale level, S-CVI/Ave values $\geq$ 0.90 are considered to indicate excellent content validity, whereas values $\geq$ 0.80 are considered acceptable. The 0.78 item-level threshold derives from the requirement that, with a panel of 6–10 experts, no more than one rater should consider an item irrelevant; the 0.90 scale-level threshold, in turn, constitutes the conventional benchmark for high-quality content validity studies [29]. It should be stressed that the CVI is informative about expert agreement on the relevance and clarity of framework components but not about the effectiveness of the framework's components in producing—for instance—learning outcomes; the latter is an empirical question, taken up in Section 6. The formulas, thresholds, and decision rules are summarized in Appendix B, Table A9.

The validation instrument comprised four sections—expert demographics, dimension-by-dimension component ratings (30 items across four dimensions), overall framework evaluation (6 items), and open-ended feedback—with the complete survey instrument provided in Appendix A (Tables A1–A5). Prior to distribution, the survey was pilot-tested with two educational technology researchers and subsequently administered online over a period of four weeks.

Expert panelists were recruited through purposive sampling that targeted individuals with demonstrated expertise in AR in education, VET, CSCL, EFL education, or mobile learning—a target of 10–15 experts being set to ensure adequate domain representation in line with the recommendation made by Nawagi et al. [32].

5.2. Validation Results

Expert Panel Characteristics

In total, 48 experts were invited and 11 completed the validation survey (23% response rate). All panelists were academics holding doctoral degrees with expertise in AR in education (n = 5), VET (n = 4), CSCL (n = 3), EFL education (n = 4), and mobile learning (n = 6); counts exceed N because several experts listed multiple domains. Seven (64%) reported over 15 years of experience in their primary field, three (27%) reported 11–15 years of experience, and one (9%) reported 5–10 years of experience. Geographically, the panel spanned Europe (n = 6), the Asia–Pacific region (n = 3), and North America (n = 2). A limitation worth noting is that the predominance of European academics (55%) may have embedded cultural assumptions into the validation. Moreover, all were researchers rather than VET frontline educators which means that practical knowledge of institutional constraints at the classroom level is not represented in the validation.

Content Validity Results

Across the 36 items (Appendix A, Table A7), I-CVI values ranged from 0.73 to 1.00 with a mean of 0.91 and an S-CVI/Ave of 0.91—a figure that exceeds the 0.90 threshold and can be conventionally interpreted as an indication of excellent content validity at the scale level [29] (dimension-level means are reported in Appendix A, Table A8). Contextual anchoring achieved the highest dimension-level mean (0.92), with items that address workplace scenario identification and co-location of resources receiving unanimous agreement (I-CVI = 1.00)—ratings that indicate strong expert agreement on the relevance—as opposed to the demonstrated effectiveness—of the contextualization of language learning within workplace scenarios as a framework element. Collaborative orchestration showed the widest range; eight of nine items exceeded the threshold, but one item, but one item, which addressed transactive interaction requirements, scored I-CVI = 0.73; an outcome which suggests expert uncertainty about whether such specific interaction requirements can be feasibly implemented or assessed in practice. Competency cultivation achieved consistent ratings across its eight items (0.82–0.91), a pattern which indicates agreement on the three-stage progression model. Capacity building received strong agreement (mean I-CVI = 0.92); a finding which we attribute, at least in part, to the widespread recognition in the educational technology literature that teacher professional development is a critical factor for successful technology integration. The overall framework evaluation section recorded the highest mean of all (0.95), with all experts agreeing that the framework addresses collaborative design challenges sufficiently.

Qualitative Feedback and Modifications

The open-ended responses revealed both endorsements and substantive concerns. On the positive side, several experts singled out the integration of CSCL principles with AR affordances as the framework's strongest element, with the attention to educator professional development also being commended. However, the experts raised a number of issues that warrant discussion. The term 'transactive interaction' was considered problematic by several panelists who pointed out that it may be unfamiliar to practitioners and difficult to operationalize in a classroom context. Questions were also raised about assessment—for instance, how collaborative writing would be evaluated and how individual contributions could be distinguished from the group product. Some experts expressed concern about whether VET institutions have the capacity to meet the implementation demands and others pointed to the need for cultural adaptation since scripts that assume a particular level of learner autonomy or openness to peer critique may not transfer well across educational cultures. At the individual level, certain experts questioned whether four dimensions are sufficient, expressed skepticism about the embodied cognition claims as they relate to writing, and suggested that the proposed role configurations lack sufficient theoretical justification.

In response to the abovementioned findings, we implemented targeted modifications. The transactive interaction item was supplemented with an operational definition which specifies that it refers to contributions in which a learner explicitly references, extends, or challenges what a group member has previously said along with practical examples. We also enhanced the competency cultivation dimension with evaluation criteria for each progression stage. It should be noted, however, that several of the concerns raised by the experts—cultural adaptation, dimensional sufficiency, and role configuration validity—are empirical questions that could not be resolved within the scope of a conceptual paper and will need to be addressed through future research. Figure 2 provides a consolidated implementation guide that maps educator actions, script configurations, writing progression, and transition indicators across the instructional phases.

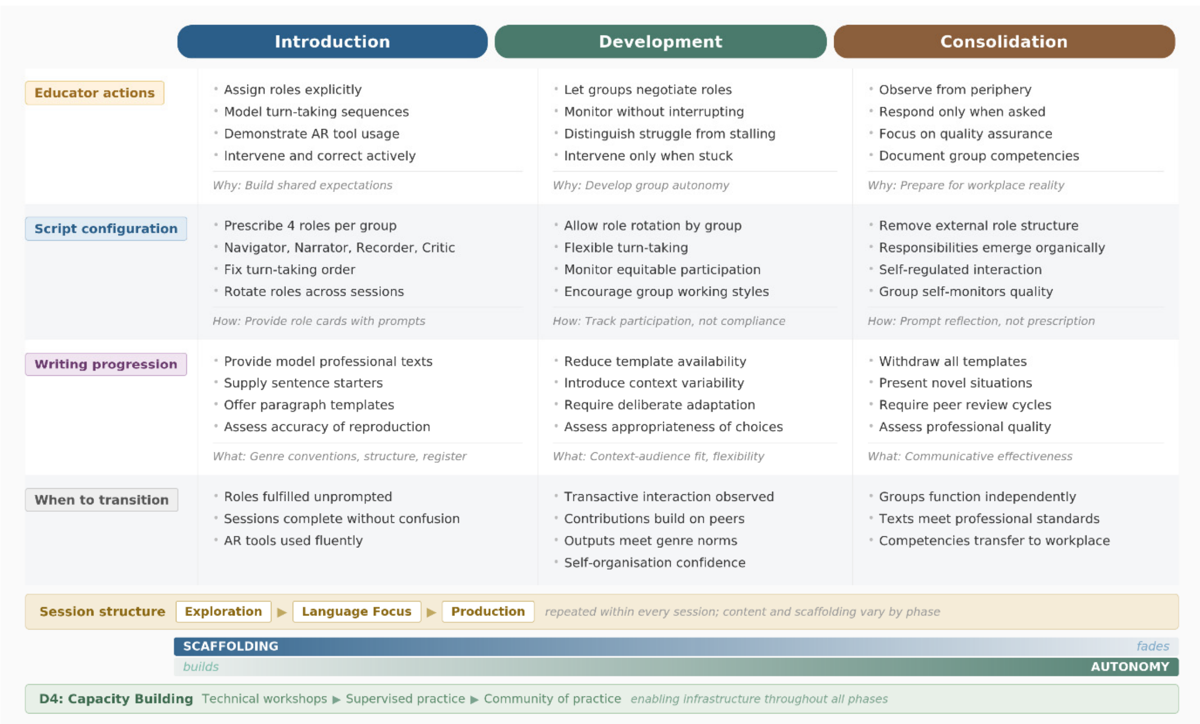

Figure 2. Implementation guide across the instructional phases.

Validation Limitations

There are several limitations to the validation process that should be acknowledged. The most important is that content validity, by definition, establishes expert agreement on relevance and clarity but cannot demonstrate that implementation will produce improved learning outcomes. Moreover, the single-round consultation limits the extent to which the framework could be refined on the basis of expert input. In addition, all panelists were academics, even though frontline VET educators were among those invited, none responded to the invitation, which means that practical knowledge of institutional constraints is not represented in the validation. Geographically, the panel was predominantly European, which may embed assumptions that do not generalize to other cultural or institutional contexts. Moreover, the sampling strategy introduces potential selection bias since experts who agreed to participate may have been more favorably disposed toward the framework's theoretical orientation.

The aforementioned limitations bound the conclusions that can be drawn from the validation conducted. The absence of frontline VET educators from the panel constrains the conclusions that can be drawn about practical implementability and points to the need for field-based stakeholder consultation in subsequent work. Likewise, the European predominance of the panel constrains the generalization of relevance judgements to other educational cultures, especially those with different traditions of teacher-led versus peer-led learning. In addition, content validity in itself does not justify claims about learning outcomes; the framework's effects on collaborative writing performance, on transfer to workplace communication, and on learner motivation remain empirical questions. Lastly, the single-round design adopted here did not allow for iterative refinement on the basis of expert feedback.

6. Discussion

In the present work we have introduced the MARS-VET framework and gathered initial evidence of its content validity. The validation confirmed that the proposed components are relevant and clearly articulated; what remains unknown, however, is whether the framework produces the learning outcomes that its grounding would predict.

6.1. Contribution and Positioning

CSCL research has established the benefits of scripted collaboration [7,8] and AR research has established the technology's efficacy for individual learning outcomes [3,33]. What the two streams have not done is converge; they have operated largely in parallel, with AR studies rarely drawing on CSCL design principles and CSCL studies rarely examining mobile AR as the medium. MARS-VET proposes how macro-level and micro-level scripts might structure collaborative interactions within AR-enhanced environments, thereby addressing a gap which has been identified in several reviews of collaborative AR [1,6].

Expert consensus confirms that the framework components are relevant to the intended domain and clearly articulated. The unanimous agreement that MARS-VET addresses collaborative design challenges is encouraging, though it should not be equated with evidence that the framework resolves them.

From a practical standpoint, VET educators who seek to implement collaborative AR currently lack systematic design guidance that is grounded in learning science. The four-dimensional structure—contextualizing learning in workplace scenarios, structuring collaboration through scripts, sequencing competency development, and building educator capacity—provides a principled starting point where none currently exists. In this regard, MARS-VET builds on and extends existing work including the AR affordances framework [5], the script theory [8], and the analysis of collaborative immersive environments [19] by applying them to the specific intersection of mobile AR, VET, and professional language development. Compared to the ARLEAN framework [15], which integrates situated, experiential, and transformative learning theories for AR applications with a focus on Learning Analytics, MARS-VET focuses on the orchestration of collaborative interaction and the development of professional writing competencies—a narrower but deeper scope. In a similar manner, while the framework of Upadhyay et al. [1] provides a broad mapping of collaborative AR applications across disciplines, MARS-VET provides more granular guidance for the specific design decisions which practitioners need to make (e.g., how to structure group roles, how to sequence the fading of scaffolds, how to progress from model-based writing to autonomous production). We acknowledge, however, that this specificity cuts both ways; on the one hand, practitioners receive concrete direction and, on the other, adaptation will be needed for contexts that differ from the VET language learning scenarios on which the framework was built.

6.2. Open Issues

The over-scripting concern is one which MARS-VET cannot sidestep. Dillenbourg [24] argued that excessive structure undermines learner autonomy and intrinsic motivation—and our framework does ask learners to work within assigned roles, prescribed turn-taking sequences, and interaction protocols that are only gradually relaxed. Our proposed response is scaffolded fading; script support is reduced as learners develop the collaborative competencies that render it unnecessary. Whether this constitutes an adequate solution, though, is far from settled. Strauß et al. [23] compared collaboration scripts with collaborative reflection and found that both approaches promoted knowledge about good collaboration—but through different mechanisms; a finding which suggests that the balance between structure and autonomy may depend on the specific learning goals being pursued. The picture is complicated further by Özbek et al. [21] who showed that goal structures influence how learners configure their own internal collaboration scripts, thus implying that external scripts may interact with learners' existing collaborative dispositions in ways that we cannot predict at this stage. Whether the framework gets this balance right is, in the end, a question that can only be answered through classroom implementation across different cultural and educational contexts.

There is also a cognitive load concern that deserves attention. Our framework requires learners to manage AR interaction, collaborative coordination, role responsibilities, and writing production at the same time. For novice learners who have not yet automated any of these processes, the compound demand may easily exceed the capacity of working memory. A related issue was encountered in a previous work with VR escape rooms [27] where learners who were using VR for the first time tended to become absorbed in the technological experience itself at the expense of content engagement. We anticipate that a similar novelty effect may occur with mobile AR, but the scaffolded release structure was designed to mitigate this by introducing elements progressively. Even so, the cumulative load during the autonomous phases may still be too much—and this is an issue that classroom trials with different learner populations and proficiency levels will need to resolve.

A third concern relates to authenticity and transfer. We have stressed the value of situated learning through AR-simulated workplace scenarios, but an honest appraisal requires us to admit that the extent to which competencies developed in a simulation actually transfer to the workplace has not been established. AR simulations—however well-designed they are—cannot replicate the social dynamics, the stakes, and the unpredictability of authentic professional communication. Transfer from simulation-based training, as the literature documents, varies considerably [34] and should not be assumed without longitudinal evidence. Interviews with VET educators about AR and hypervideo technologies reported that practitioners see considerable promise in such tools but stressed—and we share their position—that evidence from actual implementation is still needed [2].

Finally, there is the question of sustainability. MARS-VET relies on mobile AR and, as with any technology-dependent framework, the implementation details will need to be revisited as AR platforms evolve in the context of region or country-wide implementations [35]. The underlying pedagogical principles—structured collaboration, scaffolded release, contextual anchoring—are, however, technology-agnostic and should retain their validity regardless of which specific tools are used to enact them. In other words, what will change is the how, not the why.

6.3. Operational and Feasibility Considerations

The fourth objective of the present work—the articulation of implementation considerations—calls for explicit treatment of the operational and feasibility constraints that the framework presupposes. Several classes of constraint, in our view, deserve attention from practitioners and institutional leaders.

Resources are the first concern. Mobile AR implementation requires devices (one device per learner is preferable but, where the foregoing option is not feasible, one device per group with explicit navigator rotation has proven workable in similar contexts), reliable network connectivity (especially when AR content is server-hosted), AR-content authoring or licensing—often the largest single budget item—and physical spaces that can accommodate learner movement.

Staffing is the second concern. The capacity-building dimension already addresses educator development; what we wish to flag here, however, is the staffing burden that the dimension imposes. Drawing on comparable technology-integration studies [14], the initial intensive training is estimated to require approximately 16–24 h per educator, in addition to 8–12 h of supervised practice, whereas communities of practice call for continuous coordination on a more sustained basis. The aforementioned figures should be treated as starting points for institutional planning and not as fixed prescriptions.

Time is the third concern. The framework's three-phase structure presupposes sustained implementation across multiple sessions; consequently, single-session demonstrations cannot operationalize scaffolded fading and should not be expected to produce the framework's intended effects.

Contextual constraints constitute the final class. The framework presupposes vocational learning environments in which authentic or simulated workplace scenarios are accessible, learners possess functional baseline literacy in the target language, and institutional leadership supports sustained implementation. Where the foregoing conditions are only partially met, partial implementation focused on the dimensions that are most amenable to local conditions remains a defensible option—provided that the partial nature of the implementation is reported as such.

6.4. Priority Directions for Empirical Validation

In view of the conceptual nature of the present contribution, the items that follow are framed as priority directions for empirical investigation.

Priority 1—Effect of structured scripting. CSCL evidence predicts that structured scripting will produce higher-quality writing outcomes than unstructured collaboration when mobile AR is the medium [7,8]. The first empirical priority is, therefore, to verify whether collaborative scripting in mobile AR-supported VET writing tasks indeed leads to higher-quality written outputs than unstructured group collaboration.

Priority 2—Relative contribution of dimensions. The framework treats the four dimensions as jointly necessary; nevertheless, the relative contributions to the overall effect remain unknown. Component studies are, accordingly, needed to establish which of the dimensions—contextual anchoring, collaborative orchestration, competency cultivation, or capacity building—contributes most to learning outcomes, and whether all four are indeed necessary for the framework's effects to materialize.

Priority 3—Scaffolded fading. The scaffolded-fading proposal is, in our view, the framework's most contested element because the over-scripting concerns documented in the CSCL literature [24] cast doubt on whether the progressive removal of script support is sufficient to mitigate them. The third empirical priority is, therefore, to determine whether gradual fading of script support indeed produces stronger long-term collaborative competencies than sustained scripting across the implementation period.

Priority 4—Transfer to professional practice. The most consequential—and perhaps the most difficult—question is one of transfer. The question of interest is whether the collaborative writing competencies developed in mobile AR-supported VET settings indeed transfer to authentic professional communication contexts. Addressing this question will require longitudinal designs that follow learners into early professional practice and that assess workplace communication outputs against professional benchmarks.

The four priorities are ordered by logical precedence: positive findings on Priority 1 would justify investigation of Priorities 2–4, whereas negative findings would prompt framework revision before further investment is in order.

7. Conclusions

7.1. Implications and Contexts of Application

The MARS-VET framework is most directly applicable to secondary and post-secondary VET programs that prepare learners for occupations with substantial written-communication demands—hospitality (e.g., correspondence and complaint handling), healthcare (e.g., incident documentation and patient communication), business administration (e.g., supplier and client correspondence), and customer-facing technical roles (e.g., incident reporting and service documentation). Within the foregoing contexts, the framework can serve four distinct functions: (1) as a design checklist for instructional designers planning collaborative AR experiences; (2) as a basis for staged staff development in institutions adopting mobile AR; (3) as a conceptual scaffold for institutional leaders deciding where to invest across infrastructure, content, and human capacity; and (4) as a structuring device for empirical studies investigating the effectiveness of collaborative AR. On the other hand, the framework is less directly applicable—and would call for adaptation—in contexts where the prerequisites of device access, network connectivity, scenario authenticity, and educator availability for sustained development cannot be met. Under such circumstances, partial implementation focused on the dimensions that are most amenable to local conditions remains a defensible option, provided that the partial nature of the implementation is reported as such.

From a practical standpoint, the four dimensions can serve instructional designers and VET educators as a guiding checklist when they develop collaborative AR experiences. Four questions, in particular, can help determine whether the design is adequately grounded: (1) Does the learning activity take place in an authentic or approximated workplace scenario? (2) Have the group interactions been structured through scripts? (3) Does competency progression move from reproduction toward autonomous production? (4) Have the educators themselves received the professional development they need? The collaborative orchestration dimension is worth singling out here as it provides concrete guidance on how facilitation should shift from direct structuring toward monitoring and responsive intervention across the implementation phases. It should be acknowledged, though, that none of this is straightforward to implement; it requires preparation, technological competence, and a willingness to adapt on the part of both the institution and the individual educator. For institutional leaders, the framework points to where investment matters most—across infrastructure, content, and human capacity—with the caveat being that purchasing devices and software without investing in sustained professional development will, in all likelihood, fail to produce the intended outcomes.

As far as technology development is concerned, we note that the mobile AR authoring tools currently available on the market do not adequately support what MARS-VET requires. The reason is that most of them have been designed with individual learning experiences in mind and provide limited or, in some cases, no support for structured group interaction. What would make a difference are features such as role-based access controls, collaborative annotation tools, interaction logging, and adaptive scaffolding mechanisms. Learning analytics integration constitutes a further opportunity [15] and demand for such features is likely to grow as AR gains traction in educational practice.

As far as the research agenda is concerned, the immediate priority is to establish the feasibility of implementation through pilot studies in authentic VET contexts, using mixed-method designs which combine quantitative learning outcomes with qualitative documentation of the implementation process. Medium-term priorities involve quasi-experimental comparisons at multiple sites, component studies that investigate whether simplified versions produce equivalent benefits, and learning analytics integration. Long-term priorities include transfer studies that examine workplace performance, extension to VET domains beyond EFL writing, and cross-cultural validation.

7.2. Limitations

The most fundamental limitation of the present work is that the framework has not been empirically tested. Whether MARS-VET implementation produces superior collaborative learning outcomes compared to unstructured AR collaboration is a question that only controlled studies can answer and, until such evidence is available, the framework should be treated as a theoretically grounded proposal rather than a validated intervention. The four proposed roles were theoretically motivated but remain untested and it is entirely possible that alternative configurations work equally well or even better; the optimal number of roles for mobile AR groups is a question to which the answer will likely differ depending on group size, task complexity, and learner proficiency. The same caveat applies to the phase durations as groups will inevitably develop collaborative competence at different rates. Lastly, the framework presupposes a level of technology access and infrastructure that is not universally available which constrains its applicability to contexts where these prerequisites can be met.

7.3. Concluding Remarks

The MARS-VET framework brings together four interconnected bodies of work—AR affordances, CSCL scripting, situated and embodied learning, and educator professional development—and translates them into operational design guidance for an intersection that has not previously received systematic treatment. The initial expert validation reported here supports the relevance and the coherence of the framework; what remains is to determine through empirical implementation in authentic VET settings whether the framework can indeed produce the learning outcomes that its theoretical grounding would predict. The convergence of mobile AR's pedagogical promise with the persistent challenge—faced by VET in particular—of preparing learners for collaborative professional communication renders such an investigation, we believe, worth pursuing.

https://doi.org/10.3390/mti10050053

Author Contributions: Conceptualization, E.M.P., A.C. and S.M.; methodology, E.M.P. and A.C.; validation, E.M.P., A.C. and S.M.; formal analysis, E.M.P.; investigation, E.M.P.; resources, M.-J.L. and T.S.; data curation, E.M.P.; writing—original draft preparation, E.M.P.; writing—review and editing, A.C., S.M., M.-J.L. and T.S.; visualization, E.M.P. and A.C.; supervision, A.C., M.-J.L. and T.S.; project administration, A.C.; funding acquisition, M.-J.L. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: The present study is part of the EDUCA Flagship project funded by the Research Council of Finland (#358924, #358947).

Institutional Review Board Statement: The study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. In accordance with the national research ethics guidelines, the conduct of this study did not require formal ethics committee approval as it consisted only of standard psychometric assessments. Before the commencement of the activities, information about the objectives, the procedures, the data collection methods, and the confidentiality measures was provided to all participants.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors wish to express their sincere gratitude to the international experts who dedicated their time to complete the validation survey and to provide detailed feedback. We also thank the two educational technology researchers who participated in the pilot testing of the instrument for their helpful comments on clarity and wording.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
CVI	Content Validity Index
I-CVI	Item-Level Content Validity Index
MARS-VET	Mobile Augmented-Reality Storytelling for Vocational Education and Training
S-CVI/Ave	Scale-Level Content Validity Index (Averaging Method)
VET	Vocational Education and Training
VR	Virtual Reality
CSCL	Computer-Supported Collaborative Learning
EFL	English as a Foreign Language

Appendix A. Expert Validation Survey Instrument

Appendix A.1. Expert Panel Demographics

Background information about the expert panelists.

Table A1. Expert panel demographic items.

#	Item	Response Options
D.1	Primary and secondary domain(s) of expertise (select all that apply)	☐ AR in Education ☐ Vocational Education and Training ☐ Computer-Supported Collaborative Learning ☐ English as a Foreign Language Education ☐ Mobile Learning

Table A1. *Cont.*

#	Item	Response Options
D.2	Years of experience in your primary field of expertise	☐ 0–4 years ☐ 5–10 years ☐ 11–15 years ☐ More than 15 years
D.3	Geographic region of your current institutional affiliation	☐ Europe ☐ Asia-Pacific ☐ North America ☐ Latin America ☐ Africa ☐ Middle East
D.4	Primary professional role	☐ Academic researcher ☐ Teacher educator/VET trainer ☐ Frontline VET educator ☐ Instructional designer ☐ Other (please specify): ________________
D.5	Institutional affiliation (optional)	[free text]

Appendix A.2. Dimension 1: Contextual Anchoring

The following instrument was used to collect expert ratings on the relevance of each framework component. For each item, experts rated relevance on a four-point scale: 1 = Not Relevant, 2 = Somewhat Relevant, 3 = Quite Relevant, and 4 = Highly Relevant.

Table A2. Expert rating items for Dimension 1—contextual anchoring.

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
1.1	Identifying authentic workplace scenarios relevant to target occupational fields	☐	☐	☐	☐
1.2	Developing AR content that overlays professional contexts onto accessible physical spaces	☐	☐	☐	☐
1.3	Designing activities requiring engagement with contextualized content before producing written communication	☐	☐	☐	☐
1.4	Co-locating linguistic resources with situational contexts where such language naturally occurs	☐	☐	☐	☐
1.5	Reducing split attention by embedding vocabulary support within visual workplace contexts	☐	☐	☐	☐
1.6	Bridging formal classroom and informal workplace learning contexts	☐	☐	☐	☐
1.7	*What modifications would you suggest for this dimension? Are any components missing or redundant?*				

Appendix A.3. Dimension 2: Collaborative Orchestration

Table A3. Expert rating items for Dimension 2—collaborative orchestration.

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
2.1	[Macro] Structuring implementation phases with evolving collaborative expectations	☐	☐	☐	☐

Table A3. *Cont.*

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
2.2	[Macro] Initial phases: highly structured collaboration with assigned roles and explicit turn-taking	☐	☐	☐	☐
2.3	[Macro] Middle phases: negotiated collaboration where groups determine interaction patterns within constraints	☐	☐	☐	☐
2.4	[Macro] Final phases: scaffolded transition toward self-regulated collaboration transferable to workplace	☐	☐	☐	☐
2.5	[Micro] Specifying interaction patterns within collaborative activities	☐	☐	☐	☐
2.6	[Micro] Designating rotating roles (e.g., narrator, recorder, critic) with explicit responsibilities	☐	☐	☐	☐
2.7	[Micro] Implementing turn-taking sequences ensuring equitable participation	☐	☐	☐	☐
2.8	[Micro] Requiring transactive interaction (building on others' contributions) before proceeding	☐	☐	☐	☐
2.9	[Micro] Gradual fading of scripting supports as learners develop collaborative competence	☐	☐	☐	☐
2.10	*What modifications would you suggest for this dimension? Are the macro/micro distinctions clear and useful?*				

Appendix A.4. Dimension 3: Competency Cultivation

Table A4. Expert rating items for Dimension 3—competency cultivation.

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
3.1	[Sequencing] Initial level: accurate reproduction of professional text types with AR models and structural guidance	☐	☐	☐	☐
3.2	[Sequencing] Intermediate level: adaptation challenges requiring template modification for varying contexts/audiences	☐	☐	☐	☐
3.3	[Sequencing] Advanced level: original production with peer evaluation cycles	☐	☐	☐	☐
3.4	[Sequencing] Integration of domain-specific language competencies with transversal skills (collaboration, critical evaluation)	☐	☐	☐	☐
3.5	[Embodied] Physical manipulation of AR elements representing document sections	☐	☐	☐	☐
3.6	[Embodied] Gesturing to indicate relationships between content elements	☐	☐	☐	☐
3.7	[Embodied] Navigating physical space to access different stakeholder perspectives	☐	☐	☐	☐
3.8	[Embodied] Using embodied actions to interact with AR overlays representing workplace objects	☐	☐	☐	☐
3.9	*What modifications would you suggest? Is the competency progression appropriate for vocational learners?*				

Appendix A.5. Dimension 4: Capacity Building

Table A5. Expert rating items for Dimension 4—capacity building.

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
4.1	Technical skills development (operating AR applications, troubleshooting, device management)	☐	☐	☐	☐
4.2	Pedagogical understanding specific to collaborative AR integration	☐	☐	☐	☐
4.3	Training educators to monitor group interactions effectively	☐	☐	☐	☐
4.4	Developing educator capacity to intervene productively when collaboration falters	☐	☐	☐	☐
4.5	Training educators to adapt scripted activities in response to emergent learner needs	☐	☐	☐	☐
4.6	Establishing communities of practice connecting educators implementing MARS-VET	☐	☐	☐	☐
4.7	Ongoing support structures extending beyond initial training	☐	☐	☐	☐
4.8	*What modifications would you suggest? Are there critical educator support needs not addressed?*				

Appendix A.6. Overall Framework Evaluation

Table A6. Expert rating items for the overall MARS-VET framework evaluation.

Item	Component Description	1—Not	2—Somewhat	3—Quite	4—Highly
O.1	The four dimensions collectively address the collaboration deficit in AR-enhanced learning	☐	☐	☐	☐
O.2	The theoretical foundations adequately support the framework design	☐	☐	☐	☐
O.3	The framework provides actionable guidance for practitioners	☐	☐	☐	☐
O.4	The framework is appropriate for secondary vocational education contexts	☐	☐	☐	☐
O.5	The framework addresses the integration of language competencies with collaborative skills	☐	☐	☐	☐
O.6	The relationships between the four dimensions are logical and coherent	☐	☐	☐	☐
O.7	*What do you consider the main strengths of the MARS-VET framework?*				
O.8	*What limitations or gaps should be addressed?*				
O.9	*Please provide any additional comments, suggestions, or recommendations for improving the framework.*				

Appendix A.7. Content Validity Results

Table A7 reports item-level content validity (I-CVI) for each component, computed from the ratings of 11 experts. Table A8 summarizes dimension-level validity.

Table A7. Item-level content validity index (I-CVI) results and decision rules.

Item	Component Description	Experts Rating 3 or 4	I-CVI	Decision
1.1	Identifying authentic workplace scenarios relevant to target occupational fields	11	1.00	Retain
1.2	Developing AR content that overlays professional contexts onto accessible physical spaces	10	0.91	Retain

Table A7. *Cont.*

Item	Component Description	Experts Rating 3 or 4	I-CVI	Decision
1.3	Designing activities requiring engagement with contextualized content before producing written communication	10	0.91	Retain
1.4	Co-locating linguistic resources with situational contexts where such language naturally occurs	11	1.00	Retain
1.5	Reducing split attention by embedding vocabulary support within visual workplace contexts	10	0.91	Retain
1.6	Bridging formal classroom and informal workplace learning contexts	9	0.82	Retain
2.1	Structuring implementation phases with evolving collaborative expectations	11	1.00	Retain
2.2	Initial phases: highly structured collaboration with assigned roles and explicit turn-taking	11	1.00	Retain
2.3	Middle phases: negotiated collaboration where groups determine interaction patterns within constraints	10	0.91	Retain
2.4	Final phases: scaffolded transition toward self-regulated collaboration transferable to workplace	11	1.00	Retain
2.5	Specifying interaction patterns within collaborative activities	10	0.91	Retain
2.6	Designating rotating roles (e.g., narrator, recorder, critic) with explicit responsibilities	10	0.91	Retain
2.7	Implementing turn-taking sequences ensuring equitable participation	9	0.82	Retain
2.8	Requiring transactive interaction (building on others' contributions) before proceeding	8	0.73	Revise
2.9	Gradual fading of scripting supports as learners develop collaborative competence	9	0.82	Retain
3.1	Initial level: accurate reproduction of professional text types with AR models and structural guidance	10	0.91	Retain
3.2	Intermediate level: adaptation challenges requiring template modification for varying contexts/audiences	10	0.91	Retain
3.3	Advanced level: original production with peer evaluation cycles	10	0.91	Retain
3.4	Integration of domain-specific language competencies with transversal skills (collaboration, critical evaluation)	10	0.91	Retain
3.5	Physical manipulation of AR elements representing document sections	9	0.82	Retain
3.6	Gesturing to indicate relationships between content elements	10	0.91	Retain
3.7	Navigating physical space to access different stakeholder perspectives	9	0.82	Retain
3.8	Using embodied actions to interact with AR overlays representing workplace objects	9	0.82	Retain
4.1	Technical skills development (operating AR applications, troubleshooting, device management)	9	0.82	Retain
4.2	Pedagogical understanding specific to collaborative AR integration	10	0.91	Retain
4.3	Training educators to monitor group interactions effectively	10	0.91	Retain
4.4	Developing educator capacity to intervene productively when collaboration falters	9	0.82	Retain
4.5	Training educators to adapt scripted activities in response to emergent learner needs	11	1.00	Retain
4.6	Establishing communities of practice connecting educators implementing MARS-VET	11	1.00	Retain
4.7	Ongoing support structures extending beyond initial training	11	1.00	Retain
O.1	The four dimensions collectively address the collaboration deficit in AR-enhanced learning	11	1.00	Retain
O.2	The theoretical foundations adequately support the framework design	11	1.00	Retain
O.3	The framework provides actionable guidance for practitioners	10	0.91	Retain
O.4	The framework is appropriate for secondary vocational education contexts	10	0.91	Retain
O.5	The framework addresses the integration of language competencies with collaborative skills	11	1.00	Retain
O.6	The relationships between the four dimensions are logical and coherent	10	0.91	Retain

Table A8. Dimension-level content validity summary.

Dimension	Items	Mean I-CVI	Assessment
D1: Contextual Anchoring	6	0.92	Excellent
D2: Collaborative Orchestration	9	0.90	Good
D3: Competency Cultivation	8	0.88	Good
D4: Capacity Building	7	0.92	Excellent
Overall Framework Evaluation	6	0.95	Excellent
S-CVI/Ave	36	0.91	Excellent

Appendix B. CVI Calculation Procedures

Item-level CVI (I-CVI) was calculated as the proportion of experts who rated each item as relevant (i.e., three or four on the four-point scale) out of the total number of experts. Scale-level CVI (S-CVI/Ave) was computed as the arithmetic mean of all I-CVI values across items. Acceptability thresholds and decision rules follow the recommendations made by [29].

Table A9. CVI indices, formulas, acceptability thresholds, and decision rules.

Index	Formula	Threshold	Decision
I-CVI	(Experts rating $\geq$ 3) $\div$ Total experts	$\geq$0.78	Retain
		0.60–0.77	Revise
		<0.60	Remove
S-CVI/Ave	Sum of I-CVIs $\div$ Number of items	$\geq$0.90	Excellent validity

Items with I-CVI $\geq$ 0.78 were retained without modification. Item 2.8 (transactive interaction) fell within the revision range (I-CVI = 0.73) and was reworded based on expert feedback before inclusion in the final instrument. No items fell below the 0.60 removal threshold.

References

1. Upadhyay, B.; Brady, C.; Chalil Madathil, K.; Bertrand, J.; McNeese, N.J.; Gramopadhye, A. Collaborative Augmented Reality in Higher Education: A Systematic Review of Effectiveness, Outcomes, and Challenges. *Appl. Ergon.* **2024**, *121*, 104360. [CrossRef] [PubMed]
2. Candido, V.; Raemy, P.; Amenduni, F.; Cattaneo, A. Could Vocational Education Benefit From Augmented Reality and Hypervideo Technologies? An Exploratory Interview Study. *Int. J. Res. Vocat. Educ. Train.* **2023**, *10*, 138–167. [CrossRef]
3. Howard, M.C.; Davis, M.M. A Meta-Analysis of Augmented Reality Programs for Education and Training. *Virtual Real.* **2023**, *27*, 2871–2894. [CrossRef]
4. Li, G.; Luo, H.; Chen, D.; Wang, P.; Yin, X.; Zhang, J. Augmented Reality in Higher Education: A Systematic Review and Meta-Analysis of the Literature from 2000 to 2023. *Educ. Sci.* **2025**, *15*, 678. [CrossRef]
5. Buchner, J.; Kerres, M. Media Comparison Studies Dominate Comparative Research on Augmented Reality in Education. *Comput. Educ.* **2022**, *195*, 104711. [CrossRef]
6. Kazlaris, G.C.; Keramopoulos, E.; Bratsas, C.; Kokkonis, G. Augmented Reality in Education Through Collaborative Learning: A Systematic Literature Review. *Multimodal Technol. Interact.* **2025**, *9*, 94. [CrossRef]
7. Vogel, F.; Wecker, C.; Kollar, I.; Fischer, F. Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: A Meta-Analysis. *Educ. Psychol. Rev.* **2017**, *29*, 477–511. [CrossRef]
8. Fischer, F.; Kollar, I.; Stegmann, K.; Wecker, C. Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educ. Psychol.* **2013**, *48*, 56–66. [CrossRef]
9. Wu, J.; Jiang, H.; Chen, S. Augmented Reality Technology in Language Learning: A Meta-Analysis. *Lang. Learn. Technol.* **2024**, *28*, 1–23. [CrossRef]

10. Carrión-Robles, F.; Espinoza-Celi, V.; Vargas-Saritama, A. The Use of Augmented Reality through Assemblr Edu to Inspire Writing in an Ecuadorian EFL Distance Program. *Int. J. Eng. Pedagog.* **2023**, *13*, 121–141. [CrossRef]
11. Bai, Y.; Xian, H. Exploring the Interplay of Digital Storytelling, L2 Speaking Skills, Self-Regulation, and Anxiety in an IELTS Preparation Course. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1584. [CrossRef]
12. Christou, E.; Vassiliou, P.; Parmaxi, A. Augmented Reality in Language Learning: A Systematic Literature Review of the State-of-the-Art and Task Design Considerations. *Innov. Lang. Learn. Teach.* **2025**, 1–28. [CrossRef]
13. Schorr, I.; Plecher, D.A.; Eichhorn, C.; Klinker, G. Foreign Language Learning Using Augmented Reality Environments: A Systematic Review. *Front. Virtual Real.* **2024**, *5*, 1288824. [CrossRef]
14. Siliņa-Jasjukeviča, G.; Lūsēna-Ezera, I.; Iliško, D.; Surikova, S. Promoting Effective Vocational Education and Training Teacher's Professional Development and Its Transfer to Practice: A Systematic Review. *Educ. Sci.* **2025**, *15*, 596. [CrossRef]
15. Christopoulos, A.; Mystakidis, S.; Pellas, N.; Laakso, M.-J. ARLEAN: An Augmented Reality Learning Analytics Ethical Framework. *Computers* **2021**, *10*, 92. [CrossRef]
16. Acosta, J.L.B.; Navarro, S.M.B.; Gesa, R.F.; Kinshuk, K. Framework for designing motivational augmented reality applications in vocational education and training. *Australas. J. Educ. Technol.* **2019**, *35*, 102–117. [CrossRef]
17. Liu, Z.; Zuo, H.; Zhao, Y.; Lu, Y. The Effect of Embodied Learning on Students' Learning Performance: A Meta-Analysis. *Front. Psychol.* **2025**, *16*, 1658797. [CrossRef]
18. Zhang, R.-L.; Yu, S.; Gao, R.; Wang, L.; He, X. How Does Technology-Based Embodied Learning Affect Learning Effectiveness?—Based on a Systematic Literature Review and Meta-Analytic Approach. *Interact. Learn. Environ.* **2025**, *33*, 4135–4158. [CrossRef]
19. van der Meer, N.; van der Werf, V.; Brinkman, W.-P.; Specht, M. Virtual Reality and Collaborative Learning: A Systematic Literature Review. *Front. Virtual Real.* **2023**, *4*, 16. [CrossRef]
20. Tahirbegi, D.; Järvenoja, H. Collaboration Scripts as Scaffolds: Enhancing Regulatory Processes and Socio-Emotional Dynamics in Ensemble Rehearsals. *Music Sci.* **2025**, *8*, 20592043251342674. [CrossRef]
21. Özbek, T.; Greisel, M.; Wekerle, C.; Gegenfurtner, A.; Kollar, I. How Do Different Goals Affect Students' Internal Collaboration Script Configurations? Results of an Epistemic Network Analysis Study. *Front. Psychol.* **2024**, *15*, 1410152. [CrossRef]
22. Ouyang, F.; Zhang, N.; Chen, Y.; Shao, X. The Use and Effect of Cognitive Scaffoldings on Facilitating Collaborative Discussion-Supported Writing. *Technol. Pedagog. Educ.* **2025**, *34*, 275–295. [CrossRef]
23. Strauß, S.; Tunnigkeit, I.; Eberle, J.; Avdullahu, A.; Rummel, N. Comparing the Effects of a Collaboration Script and Collaborative Reflection on Promoting Knowledge about Good Collaboration and Effective Interaction. *Int. J. Comput.-Support. Collab. Learn.* **2025**, *20*, 121–159. [CrossRef]
24. Dillenbourg, P. Over-Scripting CSCL: The Risks of Blending Collaborative Learning with Instructional Design. In *Three Worlds CSCL. Can We Support CSCL?* Open Universiteit Nederland: Heerlen, The Netherlands, 2002; pp. 61–91.
25. Deci, E.L.; Ryan, R.M. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychol. Inq.* **2000**, *11*, 227–268. [CrossRef]
26. Stalheim, O.R.; Somby, H.M. An Embodied Perspective on an Augmented Reality Game in School: Pupil's Bodily Experience toward Learning. *Smart Learn. Environ.* **2024**, *11*, 24. [CrossRef]
27. Christopoulos, A.; Mystakidis, S.; Cachafeiro, E.; Laakso, M.-J. Escaping the Cell: Virtual Reality Escape Rooms in Biology Education. *Behav. Inf. Technol.* **2023**, *42*, 1434–1451. [CrossRef]
28. Pazakou, A.; Mystakidis, S.; Kazanidis, I. Collaboration in a Virtual Reality Serious Escape Room in the Metaverse Improves Academic Performance and Learners' Experience. *Future Internet* **2025**, *17*, 21. [CrossRef]
29. Polit, D.F.; Beck, C.T. The Content Validity Index: Are You Sure You Know What's Being Reported? Critique and Recommendations. *Res. Nurs. Health* **2006**, *29*, 489–497. [CrossRef]
30. Shin, J.-E.; Kim, H.; Park, H.; Woo, W. Investigating the Design of Augmented Narrative Spaces Through Virtual-Real Connections: A Systematic Literature Review. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024*; ACM: Honolulu, HI, USA, 2024; pp. 1–18.
31. Masuwai, A.; Zulkifli, H.; Hamzah, M.I. Self-assessment for teaching and learning in higher education: Instrument development and content validation. *Cogent Educ.* **2024**, *11*, 2308410. [CrossRef]
32. Nawagi, F.; Munabi, I.G.; Vyt, A.; Kiguli, S.; Rabin, T.; Waggie, F.; Mubuuke, A.G. Using the Modified Delphi Technique to Develop a Framework for Interprofessional Education during International Electives in Health Professions Training Institutions in Sub-Saharan Africa. *Front. Med.* **2023**, *10*, 1225475. [CrossRef] [PubMed]
33. Chang, H.-Y.; Binali, T.; Liang, J.-C.; Chiou, G.-L.; Cheng, K.-H.; Lee, S.W.-Y.; Tsai, C.-C. Ten Years of Augmented Reality in Education: A Meta-Analysis of (Quasi-) Experimental Studies to Investigate the Impact. *Comput. Educ.* **2022**, *191*, 104641. [CrossRef]

34. Nyström, S.; Ahn, S. Teaching with Simulators in Vocational Education and Training—From a Storing Place to a New Colleague. *Teach. Teach. Educ.* **2024**, *138*, 104409. [CrossRef]
35. Boel, C.; Dekeyser, K.; Lemal, M.; Rotsaert, T.; Valcke, M.; Schellens, T.; Struyf, D. Towards a Framework for a Nation-Wide Implementation of Augmented, Virtual and Mixed Reality in K-12 Technical and Vocational Education. *TechTrends* **2023**, *68*, 295–306. [CrossRef]