IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/47EDU/2026/4733526/
Vol. 2026 (3), Article ID 4733526

A Spatially Anchored Augmented Reality Framework for Remote Guidance of Hands-On Tasks in Bilingual Settings*

Ibrahim DAGHRIRI[1,2]*, Jie LIANG[1], Jianlong ZHOU[3], Changbo WANG[4],

Yi CHEN[5] and Quang Vinh NGUYEN[6]

[1] School of Computer Science, Faculty of Engineering & IT, University of Technology Sydney, Ultimo, Australia

[2] College of Engineering & Computer Science, Department of Computer Science, Jazan University, Jazan, Saudi Arabia

[3] Data Science Institute, University of Technology Sydney, Ultimo, Australia

[4] School of Data Science and Engineering, East China Normal University, Shanghai, China

[5] School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing, China

[6] School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, Australia

Correspondence should be addressed to: Ibrahim DAGHRIRI, idaghriri@jazanu.edu.sa

*Presented at the 47th IBIMA International Conference, 29-30 June 2026, Madrid, Spain

Abstract

In vocational and technical training, remote instruction involves a trainer guiding a learner through a hands-on procedure from another location. Video conferencing is often used for this purpose, but it provides limited support for linking the trainer's verbal guidance to the relevant equipment component, workspace location, or procedural step. The guidance situation becomes more demanding when explanations are given in one language while technical terms and component labels appear in another. The learner must then connect the instruction, the term, and the component at the same point in the task. This added coordination can make the procedure harder to follow and increase the mental effort required. Human-computer interaction research has examined remote guidance for physical tasks, yet bilingual vocational instruction remains underexplored. This paper presents a cloud-based, spatially anchored augmented reality framework built around trainer-cloud-learner architecture. It links bilingual guidance to the relevant physical component and active procedural step, while the trainer controls progression and the cloud coordinates the shared procedural state and delivers the corresponding guidance. By keeping the instruction, terminology, and point of action aligned within the task view, the framework is intended to support task fluency by reducing avoidable sources of mental effort.

Keywords: Augmented reality; remote hands-on instruction; vocational education and training; spatial guidance; bilingual technical instruction; cloud-based guidance

Introduction

Hands-on vocational training depends on a close connection between what a trainer says and what a learner does. A learner must identify the correct component, understand the instruction, keep the steps in order, and act on the equipment in front of them. Inside a shared workshop, the trainer can support the learner as each step unfolds. Standing beside the learner, the trainer points to the part in question, slows down when the learner hesitates, repeats a direction, and intervenes before an incorrect action goes further. Remote delivery reduces access to most of these immediate forms of support. A video call keeps trainer and learner in contact. Camera angle, viewing

Cite this Article as: Ibrahim DAGHRIRI, Jie LIANG, Jianlong ZHOU, Changbo WANG, Yi CHEN and Quang Vinh NGUYEN Vol. 2026 (3) "A Spatially Anchored Augmented Reality Framework for Remote Guidance of Hands-On Tasks in Bilingual Settings " Communications of International Proceedings, Vol. 2026 (3), Article ID 4733526, https://doi.org/10.5171/2026.4733526

distance, screen size, and obstruction can make the component under discussion difficult to see. Studies of remote vocational education report a reduction in interaction between trainers and learners and difficulties in assessing learners' practical work remotely (Shdaifat et al. 2020; Yeap et al. 2021). Research conducted during the move to online delivery also reports limited task guidance and platforms poorly suited to practical instruction (Magetos et al. 2021; Seyffer et al. 2022).

The task may become more demanding when instruction and technical terminology use different languages. Within Saudi Arabia's Technical and Vocational Training Corporation (TVTC), learners may receive general instruction in Arabic while working with English technical terminology during vocational tasks (Alhomidan 2025; Alzahrani and Saleh 2025). Saudi research has examined digital tools and augmented reality for English-language learning, but its focus is language learning rather than the use of English-labeled components during remote physical procedures (Binhomran and Altalhab 2021; Alzahrani and Saleh 2025). A learner may understand the Arabic explanation but pause when identifying the component named by an English label. During the same step, the learner must locate the component, keep the instruction in memory, read the technical term, and perform the action. When the explanation, terminology, and physical component appear in separate places, attention shifts repeatedly among them, which may raise mental effort and interrupt the flow of the task.

Augmented reality can place guidance within the learner's view of the physical task. Augmented reality and other immersive technologies have been applied across education and training, with documented benefits for engagement and learning alongside practical implementation challenges (Daghriri et al. 2024). A label, arrow, highlight, or short note can appear over or beside the relevant component or action point rather than being presented separately from the task. Studies in maintenance, assembly, and industrial training have examined task-linked guidance for component identification and procedural work (Henderson and Feiner 2011; Chiang et al. 2022). Remote collaboration research has also examined how an expert can provide spatial guidance to a learner at another location (Wang et al. 2025). Many existing applications, however, emphasize learner-operated guidance or use within a single location (Chiang et al. 2022). Trainer-led remote vocational instruction requires an architecture that keeps pacing, correction, and step order under trainer control while linking spatial presentation to procedural step management, learner feedback, and a shared session record. Trainer commands, a cloud-held procedural state, bilingual content, learner responses, and session records are assigned across the trainer, cloud, and learner components and connected through defined guidance and feedback flows. This paper makes four contributions:

- It identifies the coordination difficulty that arises when first-language instruction, second-language terminology, and physical task execution meet at the same point of action.
- It introduces bilingual procedural element interactivity, a construct that names this difficulty and applies to other cross-language training settings.
- It proposes spatially anchored remote scaffolding as the design response, combining trainer control with spatially placed guidance.
- It presents an architecture that organizes these design responses for trainer-led remote vocational instruction, together with five evaluation dimensions that define how the framework will be assessed.

Related Work

Vocational competence requires more than recalling information. A learner may be able to name a component or describe a procedure but still have difficulty locating and orienting the component, making the correct connection, or following the required sequence during task performance. Common remote teaching tools provide different forms of support, but each has limits during physical tasks. Video conferencing allows direct conversation, although a trainer may need to rely on descriptions such as "the connector on the left" or "the second port from the edge." Such descriptions depend on a shared view that is difficult to maintain when the trainer and learner see the equipment from different angles. Pre-recorded video provides a stable demonstration and allows repeated viewing, but it cannot respond to a learner's error or uncertainty. Virtual reality provides a controlled setting for practice, although the simulation does not always reproduce weight, resistance, fit, or tool handling. Immersive vocational training may support instruction but has limits when the skill depends on physical contact with equipment (Alhajri et al. 2026).

Spatial presentation, however, is not sufficient on its own. An overlay may indicate where the learner should look, but deciding whether to proceed, repeat a step, or correct an action requires the visual guidance to be linked to procedural control and the trainer's judgment. Remote practical instruction may instead require the trainer to pause progression at a step, provide further explanation, or wait for the learner to indicate readiness. The framework therefore includes five requirements: spatially positioned instructions, an ordered sequence of procedural states,

trainer control over progression, a feedback channel for the learner, and bilingual technical support at the point of use. The architecture adds two coordinating elements that deliver these requirements across locations: a cloud layer that holds the shared procedural state, and a session record available to the trainer for review.

Remote guidance of physical tasks has been studied in human-computer interaction for nearly three decades, where a remote helper directs a local user through spatial cues, annotations, and a shared visual space (Huang et al. 2022). Research in this area has examined how experts guide learners through physical procedures using mixed-reality telepresence and visual communication cues (Thoravi Kumaravel et al. 2019; Kim et al. 2019). Augmented reality research in vocational and industrial training has also examined spatial overlays, procedural instruction, learner engagement, and technology acceptance (Chiang et al. 2022; Hamid et al. 2024; Mustapha et al. 2024). This body of work, however, generally assumes that instruction and the learner share a single language, and it often treats the remote party as a worker being assisted. Remote collaboration systems also tend to support learner-autonomous execution rather than trainer-controlled progression (Wang et al. 2025). Studies of remote vocational education and training report weak interaction, limited practical guidance, and difficulty correcting physical work through conventional platforms (Shdaifat et al. 2020; Magetos et al. 2021; Yeap et al. 2021). Research on language technology in Saudi education supports presenting English terms in context but gives limited attention to their use during remote physical procedures (Binhomran and Altalhab 2021; Alzahrani and Saleh 2025).

To the best of our knowledge, prior research has not integrated bilingual technical instruction with remote spatial guidance for physical procedures, where first-language instruction, second-language terminology, and physical action meet at the same point of action. This paper identifies that difficulty as bilingual procedural element interactivity and proposes spatially anchored remote scaffolding as a design response: trainer-directed guidance placed at the learner's point of action and linked to the active procedural state.

Design Foundations

The proposed framework draws on three connected foundations. Augmented reality places guidance within the learner’s view of the physical task, while cloud coordination maintains the procedural state shared by trainer and learner. Bilingual support and trainer-controlled scaffolding shape how that guidance is presented and adjusted during each step. These foundations link spatial presentation, procedural control, and language support within one remote guidance process, each grounded in established learning theory.

Augmented Reality and Spatial Guidance

Augmented reality places digital information in relation to physical tasks. In procedural training, an overlay can link an instruction to a component, tool, or action point. Henderson and Feiner (2011) showed that augmented reality documentation can keep information within the task view during maintenance and repair. Chiang et al. (2022) also identified recurring uses for component identification, sequenced instruction, assembly, maintenance, and demonstration.

In the framework, augmented reality serves as a guidance medium rather than a self-directed teaching system. The trainer decides what appears, when it appears, and whether the learner advances. Spatial anchoring links the current instruction to the relevant part, adopting the use of anchored visual cues to guide viewer attention (Chen et al. 2026). Guidance may appear as a component label, an arrow at an action point, a highlight around the target, or a short note. The Guidance Content Repository stores each item and links it to the active procedural state.

Cloud Coordination and Procedural Step Management

A remote session requires a shared layer that connects trainer and learner across locations. In the architecture, the Cloud Session Manager holds a shared procedural state that both interfaces can read, while instructional judgment remains with the trainer. Both sides can therefore work from the same procedural information, and neither interface has to manage the full procedure.

Multi-step work depends on the correct sequence. A skipped, repeated, or reordered action can result in incorrect assembly, create an unsafe state, or disrupt task continuity. Step Management Logic represents the procedure as an ordered set of states. The learner receives only the guidance linked to the active state. The Segmenting Principle holds that learners handle complex material more readily when it is divided into manageable parts (Mayer 2001, 2009). Evidence from data videos likewise suggests that segmented presentation supports viewers only when pacing matches their processing needs (Chen et al. 2025), a condition the framework addresses through trainer-controlled progression. In the framework, segmentation governs the procedure rather than serving only as a

display choice. Using the same name for the current step gives trainer and learner a common reference and reduces reliance on informal descriptions of progress.

Bilingual Support and Trainer-Controlled Scaffolding

Cummins (2000) distinguishes everyday conversational language from the language required for academic and technical work. Krashen's account of comprehensible input also stresses the need for language that learners can understand in context (Krashen 1982). The framework places Arabic instructions and English technical terms within the same task view. A label can show both languages, and a note can explain a term for the current step. A learner may know the English term in isolation but pause when matching it to the relevant component. The design keeps the term, object, and instruction in the same view rather than separating them across different sources.

Scaffolding provides temporary expert support matched to the learner's current ability (Vygotsky 1978; Wood et al. 1976). In a workshop, the trainer observes the learner and adjusts support by adding a cue, correcting an action, or allowing the learner to continue. The Trainer Control Panel gives the trainer remote control over progression, a channel for guidance input, and access to the session record. Learner feedback lets the trainer adjust support without relying on a fixed automated sequence to control progression. Bandura (1977, 1997) links confidence to successful task experiences and the belief that the next action is manageable. Step confirmation and controlled advancement may help maintain confidence when the learner is unsure.

Theoretical Mapping

The framework draws on six theoretical traditions, and each informs a particular design choice rather than serving as background. Table 1 sets out the link from source to function. Cognitive Load Theory explains why separated sources of information can raise mental effort, and the multimedia learning principles guide where guidance sits and how it is divided. Scaffolding theory places instructional control with the trainer. Self-Efficacy Theory supports paced advancement and step confirmation. Second Language Acquisition Theory informs bilingual terminology support at the point of use, and task fluency research ties the design to the quality of procedural movement rather than to task completion alone. Together, these theoretical links explain why the design keeps language, guidance, and task as close together as possible. The two design concepts in the next section give this reasoning a specific form.

Table 1: Theoretical foundations and framework design functions

Theoretical source	Applied principle	Framework function
Cognitive Load Theory (Sweller 1988, 1994)	Reduce avoidable mental effort caused by separated information	Current-step display and spatially co-located guidance
Cognitive Theory of Multimedia Learning (Mayer 2001, 2009)	Spatial contiguity, segmenting, and signaling	Component labels, visual cues, and step-based content
Scaffolding theory (Vygotsky 1978; Wood et al. 1976)	Adjust expert support during task performance	Trainer-controlled progression and remote guidance input
Self-Efficacy Theory (Bandura 1977, 1997)	Build certainty through managed task progress	Step confirmation and controlled advancement
Second Language Acquisition Theory (Krashen 1982; Cummins 2000)	Present technical language in context at the time of need	Arabic-English labels and guidance notes
Task fluency research (Van Merriënboer et al. 2002)	Support procedural movement with limited hesitation and backtracking	Step-state management, progress feedback, and corrective guidance

Key Design Concepts

This section sets out two terms used in this paper to describe the design. Bilingual procedural element interactivity describes the coordination problem the design addresses, and spatially anchored remote scaffolding describes the proposed response.

Bilingual Procedural Element Interactivity

Element interactivity refers to the extent to which several elements must be processed together because they cannot be understood or performed independently (Sweller 1988, 1994; Sweller et al. 1998). During bilingual practical instruction, a learner may need to coordinate the physical component, the Arabic explanation, and the English technical term at the same time. This paper uses the term bilingual procedural element interactivity to describe this condition. It describes the coordination required when first-language instruction, second-language technical terminology, and physical task execution occur at the point of action. This differs from the split-attention effect in cognitive load theory, where separated sources of information must be integrated to be understood (Sweller et al. 1998). Here the difficulty is coordination during action rather than integration for understanding. The term does not assume weak English ability. A learner may understand each element separately but still experience disruption when the component, explanation, and term are separated in space or time. When information is separated, the learner must search for, retain, interpret, and connect it before acting, which may increase mental effort. Placing bilingual guidance within the same task view may reduce shifts among sources and position the technical term beside the component it identifies.

Spatially Anchored Remote Scaffolding

Spatially anchored remote scaffolding describes the design response. Trainer-directed guidance is placed at the learner's point of action and linked to the active procedural state. The trainer controls progression, the cloud coordinates the session, and the learner receives guidance for the current step. Spatial anchoring alone does not provide scaffolding when guidance appears without trainer control. Trainer control alone does not resolve spatial separation when guidance remains verbal or outside the learner's task view. The term combines expert direction with spatial placement. The design is intended to reduce the separation among language, guidance, and task while keeping progression under trainer control. The trainer-cloud-learner architecture presented next realizes this response as a set of connected components. Task fluency, the intended outcome of the framework, refers to the smooth, continuous, and low-disruption execution of procedural tasks with minimal hesitation, sequencing breakdown, or interruption (Van Merriënboer et al. 2002). By co-locating the bilingual term, the component, and the instruction within a single view and presenting one step at a time, the design is intended to reduce the extraneous mental effort that arises from concurrent demands, which may in turn support more fluent task execution.

Proposed Framework Architecture

This section presents the trainer-cloud-learner architecture.

Architecture Overview

The architecture follows a trainer-cloud-learner structure. The trainer uses the Trainer Control Panel, while the Cloud Session Manager maintains the procedural state, retrieves guidance content, and records session activity. Guidance is presented through the Learner Task Guidance Interface, and the Devices and Access Layer manages session entry for trainer and learner. Figure 1 shows the four components and the connections among them. Instructional delivery moves from the trainer through the cloud to the learner, while progress feedback travels in the opposite direction through the same cloud layer. The framework treats trainer control as a design principle: the trainer directs progression and guidance, while the learner performs the physical task. Within a trainer-cloud-learner structure, the framework assigns each function to a defined component and flow. The architectural contribution comes from how these functions are combined and distributed across the trainer, cloud, and learner components. Together, the trainer's control over progression and the spatial placement of guidance realize spatially anchored remote scaffolding within the architecture.

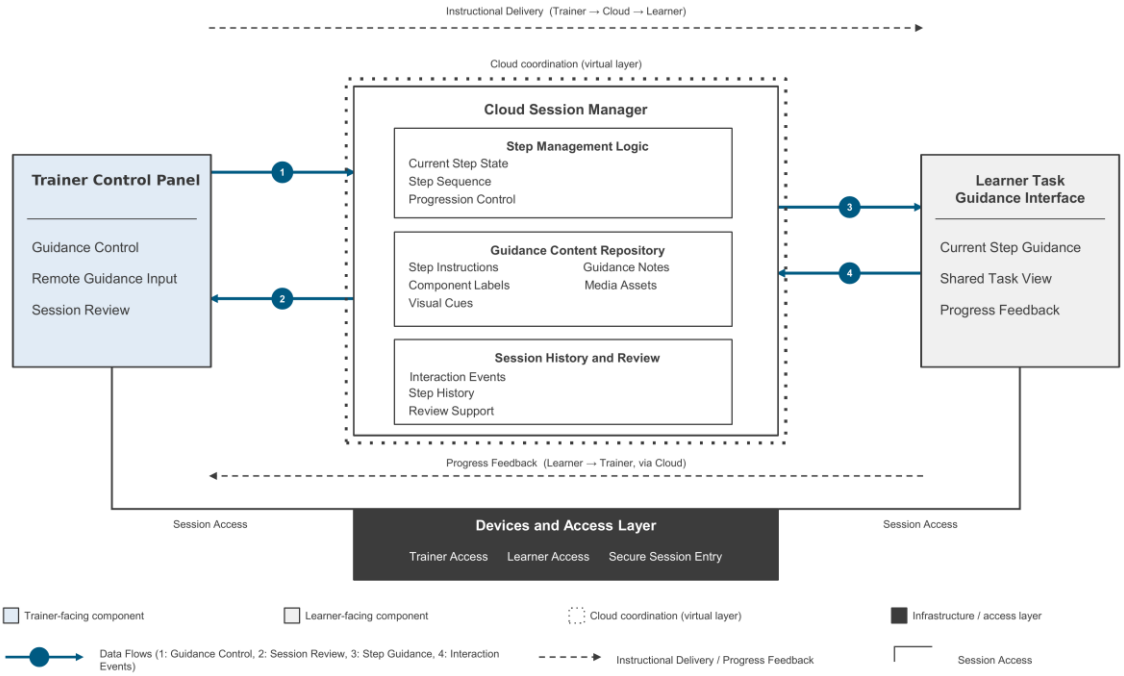

Fig 1. Proposed Framework Architecture

Trainer Control Panel

The Trainer Control Panel is the trainer-facing component. Guidance Control allows the trainer to select the current step, approve advancement, pause progression, or return to an earlier step. Remote Guidance Input supports intervention during the active step. The trainer can add a note, visual cue, annotation, or short clarification linked to the learner's current state. The Cloud Session Manager then makes that material available within the current-step view. Session Review gives the trainer access to the session record, including how the procedure progressed, where the learner requested support, and which steps involved correction.

Cloud Session Manager

The Cloud Session Manager is the central coordination layer and contains three functional groups. Step Management Logic represents the task as a sequence of procedural states. Current Step State identifies the active step, Step Sequence defines the order of actions, and Progression Control applies the trainer's decisions to movement through the sequence. The Guidance Content Repository stores the resources assigned to each step. When the active state changes, the Cloud Session Manager retrieves the content for the new step and prepares it for the learner interface. Session History and Review stores Interaction Events, Step History, and Review Support. Step History records movement through the procedure, while Review Support organizes the stored information for later examination by the trainer.

Learner Task Guidance Interface

The Learner Task Guidance Interface is the learner-facing component. Current Step Guidance presents the instruction and supporting content for the active state. The content may include a bilingual instruction, component label, spatial cue, note, or media asset retrieved from the repository. By presenting the Arabic instruction, the English technical label, and the spatial cue in one co-located view, this component directly addresses bilingual procedural element interactivity. Shared Task View presents a cloud-synchronized view of the active session state, including the current step, guidance state, trainer annotations, and learner progress. Through Progress Feedback, the learner can signal completion, request help, report difficulty, or indicate readiness. Learner feedback informs progression, but the decision to advance remains with the trainer.

Devices and Access Layer

The Devices and Access Layer manages trainer and learner access to the session. Separating access functions from the instructional components allows authentication and session-entry procedures to change without altering the step logic. Trainer and learner enter the same cloud-coordinated session through interfaces suited to their

respective roles. The learner interface is intended to present augmented reality guidance on standard consumer devices, such as tablets or phones, without requiring specialist headsets.

Technical Operation and Data Flow

This section explains how information moves through the architecture during a guided session. It first identifies the main exchanges among the Trainer Control Panel, Cloud Session Manager, and Learner Task Guidance Interface. It then follows the session from entry and initial step selection through guidance delivery, learner feedback, progression, and later review.

Primary Data Flows

Figure 1 shows the broader relationships for instructional delivery, progress feedback, and session access. Table 2 lists the four numbered flows that carry the main exchanges. The walkthrough that follows traces these flows through one guided session.

Table 2: Primary framework data flows

Flow	From	To	Function
1. Guidance Control	Trainer Control Panel	Cloud Session Manager	Step selection, progression, and session commands
2. Session Review	Cloud Session Manager	Trainer Control Panel	Step history and interaction records for trainer review
3. Step Guidance	Cloud Session Manager	Learner Task Guidance Interface	Instruction, labels, cues, notes, and media for the active step
4. Interaction Events	Learner Task Guidance Interface	Cloud Session Manager	Completion signals, help requests, difficulty indicators, and learner actions

Guided Session Operation

The session uses a single procedural state held in the cloud, which both interfaces read so that trainer and learner remain on the same step. Through Secure Session Entry, both sides enter the same Cloud Session Manager and the interface assigned to their role. The trainer begins the procedure through Guidance Control, and the Cloud Session Manager sets the initial step, retrieves its assigned content, and sends it to the Learner Task Guidance Interface, which presents one action at a time. Spatial cues mark the relevant component or action point, while bilingual guidance pairs the Arabic explanation with the English technical term. Notes or annotations added through Remote Guidance Input are linked to the active step and sent to the learner. During the task, the learner returns Progress Feedback such as completion signals, help requests, or difficulty reports, which the trainer reviews before deciding whether to advance, pause, correct, or return to an earlier step. Progression Control then applies the decision and the cycle continues. Interaction Events and Step History remain stored in Session History and Review, available to the trainer through the Trainer Control Panel.

Framework Evaluation and Future Work

The framework is presented at the design stage, and a sequential mixed-methods evaluation is proposed. Parallel surveys of trainers and learners at TVTC colleges will first document current challenges in remote hands-on instruction across five dimensions: instructional clarity, language comprehension, mental effort, task confidence, and task fluency. The surveys separate comprehension of technical terms from coordination difficulty during the task, so that this difficulty can be examined independently of language ability. The framework has been implemented as a Unity-based prototype: a shared three-dimensional instructional environment in which the trainer and learner interfaces operate on the same session state. The prototype implements the framework's mechanisms directly. Bilingual labels anchored beside each component respond to bilingual procedural element interactivity, while trainer-controlled step progression, spatial highlighting, point-of-need support assets, and corrective guidance realize spatially anchored remote scaffolding. Expert trainers will then view a scenario video

recorded from this environment, demonstrating these mechanisms on a computer hardware maintenance task, and participate in semi-structured interviews to assess whether the design responds to those challenges, following practitioner-based validation approaches in visual design research (Errey et al. 2025). Each dimension is measured through five-point Likert subscales in the trainer and learner surveys, with internal consistency assessed using Cronbach's alpha, and mental effort measured as self-reported perceived demand. Interview data will be examined through thematic analysis coded against the same five dimensions, together with a comparative rating of the concept against the trainers' current tools. The framework will then be refined based on the trainer recommendations, providing the design basis for subsequent augmented reality delivery on the learner's physical equipment.

Conclusion

This paper examines remote guidance for physical tasks in bilingual settings, where first-language instruction, second-language technical terminology, and physical action must be coordinated at the learner's point of action. Rather than treating remote practical guidance as a collection of separate communication and display tools, the paper frames it as a coordinated instructional process. It describes the difficulty of identifying components, following Arabic instructions, and interpreting English technical terms within the same procedural step as bilingual procedural element interactivity. In response, it proposes spatially anchored remote scaffolding, which connects trainer-controlled guidance with the learner's task view and the current procedural step. The proposed trainer–cloud–learner architecture brings together trainer decisions, step management, guidance content, learner feedback, and session records. The Trainer Control Panel supports progression and correction, the Cloud Session Manager maintains the current step and retrieves the relevant guidance, and the Learner Task Guidance Interface presents this guidance directly within the task context. Together, these components provide a structured approach to remote vocational tasks involving identifiable components and ordered procedures. The design aims to support task fluency while reducing avoidable sources of mental effort. In the context of TVTC colleges, the framework provides a basis for coordinating instructional explanations, English technical terminology, and the learner's physical actions during remote practical training.

References

- Alhajri, S., Yourston, D., Khassawneh, O., Mohammad, T. and Darwish, T. K. (2026) 'Blending realities: Enhancing vocational welding training through virtual reality-integrated models,' *Education + Training*, 68 (3), pp. 303–322. https://doi.org/10.1108/ET-05-2025-0354
- Alhomidan, A. M. A. (2025) 'Integrating English for Specific Purposes in technical and vocational education: A sustainable model for industry-education alignment,' *International Journal of Social Science and Humanities Research*, 13 (2), pp. 201–218. https://doi.org/10.5281/zenodo.15341890
- Alzahrani, A. and Saleh, A. M. (2025) 'The role of language learning technologies: Insights from the Saudi context,' in A. H. Al-Hoorie, C. Mitchell and T. Elyas (eds), *Language Education in Saudi Arabia: Integrating Technology in the Classroom*, Springer, pp. 129–145. https://doi.org/10.1007/978-3-031-84278-8_8
- Bandura, A. (1977) 'Self-efficacy: Toward a unifying theory of behavioral change,' *Psychological Review*, 84 (2), pp. 191–215. https://doi.org/10.1037/0033-295X.84.2.191
- Bandura, A. (1997) *Self-Efficacy: The Exercise of Control*, W. H. Freeman.
- Binhomran, K. and Altalhab, S. (2021) 'The impact of implementing augmented reality to enhance the vocabulary of young EFL learners,' *The JALT CALL Journal*, 17 (1), pp. 23–44. https://doi.org/10.29140/jaltcall.v17n1.304
- Chen, Y., Liang, J., Chan, K., Errey, N., Zhou, C. and Chen, Y. (2025) 'Enhancing cognitive clarity through drill-down structuring in data videos,' in *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (VINCI 2025)*, ACM. https://doi.org/10.1145/3769534.3769597
- Chen, Y., Liang, J., Chan, K. and Errey, N. (2026) 'Adaptive visual anchors in data videos: Guiding attention through visual persistence,' in Luo, Y. (ed.), *Cooperative Design, Visualization, and Engineering (CDVE 2025)*, Lecture Notes in Computer Science 16092, Springer, pp. 215–227. https://doi.org/10.1007/978-3-032-06952-8_22
- Chiang, F.-K., Shang, X. and Qiao, L. (2022) 'Augmented reality in vocational training: A systematic review of research and applications,' *Computers in Human Behavior*, 129, 107125. https://doi.org/10.1016/j.chb.2021.107125
- Cummins, J. (2000) *Language, Power and Pedagogy: Bilingual Children in the Crossfire*, Multilingual Matters.

- Daghriri, I., Liang, C. J. and Huang, W. (2024) 'Spatial immersive learning environments in primary education: a review of impacts and implementation challenges,' in *Proceedings of the International Conference on Information Technology, Data Science, and Optimization* (I-DO '24), Taipei, Taiwan. ACM, pp. 99–105. https://doi.org/10.1145/3658549.3658572
- Errey, N., Zowghi, D., Leong, T. W., Wu, Y., Yuan, X., Huang, W. and Liang, C. J. (2025) 'Heuristics for evaluating narrative visualization: A validation study,' *Journal of Visualization*, 28, pp. 645–659. https://doi.org/10.1007/s12650-025-01051-y
- Hamid, R. A., Ismail, I. and Yahaya, W. A. J. W. (2024) 'Augmented reality for skill training from TVET instructors' perspective,' *Interactive Learning Environments*, 32 (4), pp. 1291–1299. https://doi.org/10.1080/10494820.2022.2118788
- Henderson, S. J. and Feiner, S. (2011) 'Exploring the benefits of augmented reality documentation for maintenance and repair,' *IEEE Transactions on Visualization and Computer Graphics*, 17 (10), pp. 1355–1368. https://doi.org/10.1109/TVCG.2010.245
- Huang, W., Wakefield, M., Rasmussen, T. A., Kim, S. and Billinghurst, M. (2022) 'A review on communication cues for augmented reality based remote guidance,' *Journal on Multimodal User Interfaces*, 16 (2), pp. 239–256. https://doi.org/10.1007/s12193-022-00387-1
- Kim, S., Lee, G., Huang, W., Kim, H., Woo, W. and Billinghurst, M. (2019) 'Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration,' in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. https://doi.org/10.1145/3290605.3300403
- Krashen, S. D. (1982) *Principles and Practice in Second Language Acquisition*, Pergamon Press.
- Magetos, D., Sarlis, I., Kotsifakos, D. and Douligeris, C. (2021) 'Utilization of remote access and distance control technology for the management of virtual classrooms during COVID-19 in VET specialties' laboratories,' in *European Conference on e-Learning*, pp. 583–591. https://doi.org/10.34190/EEL.21.072
- Mayer, R. E. (2001) *Multimedia Learning*, Cambridge University Press.
- Mayer, R. E. (2009) *Multimedia Learning*, 2nd edn, Cambridge University Press.
- Mustapha, S. A. M., Zaharudin, R., Ariffin, H. F. and Abdullah, S. (2024) 'A systematic review exploring the impact of augmented reality on teaching modules in vocational education,' *Journal of Contemporary Social Science and Education Studies*, 4 (3 special issue), pp. 177–189. https://doi.org/10.5281/zenodo.14064593
- Seyffer, S., Hochmuth, M. and Frey, A. (2022) 'Challenges of the coronavirus pandemic as an opportunity for sustainable digital learning in VET,' *Sustainability*, 14 (13), 7692. https://doi.org/10.3390/su14137692
- Shdaifat, S. A. K., Shdaifat, N. A. and Khateeb, L. A. (2020) 'The reality of using e-learning applications in vocational education courses during COVID-19,' *International Education Studies*, 13 (10), pp. 105–112. https://doi.org/10.5539/ies.v13n10p105
- Sweller, J. (1988) 'Cognitive load during problem solving: Effects on learning,' *Cognitive Science*, 12 (2), pp. 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J. (1994) 'Cognitive load theory, learning difficulty, and instructional design,' *Learning and Instruction*, 4 (4), pp. 295–312. https://doi.org/10.1016/0959-4752(94)90003-5
- Sweller, J., van Merriënboer, J. J. G. and Paas, F. G. W. C. (1998) 'Cognitive architecture and instructional design,' *Educational Psychology Review*, 10 (3), pp. 251–296. https://doi.org/10.1023/A:1022193728205
- Thoravi Kumaravel, B., Anderson, F., Fitzmaurice, G., Hartmann, B. and Grossman, T. (2019) 'Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence,' in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174. https://doi.org/10.1145/3332165.3347872
- Van Merriënboer, J. J. G., Clark, R. E. and de Croock, M. B. M. (2002) 'Blueprints for complex learning: The 4C/ID-model,' *Educational Technology Research and Development*, 50 (2), pp. 39–61. https://doi.org/10.1007/BF02504993
- Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press.
- Wang, L., Li, X., Wu, J., Zhou, D., Im, S.-K. and Popescu, V. (2025) 'AVICol: Adaptive visual instruction for remote collaboration using mixed reality,' *International Journal of Human-Computer Interaction*, 41 (2), pp. 1260–1279. https://doi.org/10.1080/10447318.2024.2313920
- Wood, D., Bruner, J. S. and Ross, G. (1976) 'The role of tutoring in problem solving,' *Journal of Child Psychology and Psychiatry*, 17 (2), pp. 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

- Yeap, C. F., Suhaimi, N. and Nasir, M. K. M. (2021) 'Issues, challenges, and suggestions for empowering TVET education during COVID-19 in Malaysia,' *Creative Education*, 12 (8), pp. 1818–1839. https://doi.org/10.4236/ce.2021.128138