Designing LLM-Powered Multimodal Instructions to Support Rich Hands-on Skills Remote Learning: A Case Study with Massage Instructors and Learners

Chutian Jiang*
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
cjiang893@connect.hkust-gz.edu.cn

Yinan Fan*
Division of Emerging Interdisciplinary Areas
The Hong Kong University of Science and Technology
Hong Kong, China
Smart Manufacturing Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
yfanaw@connect.ust.hk

Junan Xie
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
jxie622@connect.hkust-gz.edu.cn

Emily Kuang
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, USA
ek8093@rit.edu

Baichuan Feng
Smart Manufacturing Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
baichuanfeng@hkust-gz.edu.cn

Kaihao Zhang†
Smart Manufacturing Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
Department of Mechanical and Aerospace Engineering
The Hong Kong University of Science and Technology
Hong Kong, China
kaihaozhang@hkust-gz.edu.cn

Mingming Fan†
Computational Media and Arts Thrust & Internet of Things Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
Division of Integrative Systems and Design
The Hong Kong University of Science and Technology
Hong Kong, China
mingmingfan@ust.hk

Abstract

Although remote learning is widely used for delivering and capturing knowledge, it has limitations in teaching hands-on skills that require nuanced instructions and demonstrations of precise actions, such as massage. Furthermore, scheduling conflicts between instructors and learners often limit the availability of real-time feedback,

*Both authors contributed equally to this research.
†Corresponding authors

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713677

reducing learning efficiency. To address these challenges, we developed a synthesis tool utilizing an LLM-powered Virtual Teaching Assistant (VTA). This tool integrates multimodal instructions that convey precise data, such as stroke patterns and pressure control, while providing real-time feedback for learners and summarizing their performance for instructors. Our case study with instructors and learners demonstrated the effectiveness of these multimodal instructions and the VTA in enhancing massage teaching and learning. We then discuss the tools' use in other hands-on skills instruction and cognitive process differences in various courses.

CCS Concepts

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

Keywords

Remote Massage Learning; Multimodal Teaching and Learning; Hands-on Training.

ACM Reference Format:

Chutian Jiang, Yinan Fan, Junan Xie, Emily Kuang, Baichuan Feng, Kaihao Zhang, and Mingming Fan. 2025. Designing LLM-Powered Multimodal Instructions to Support Rich Hands-on Skills Remote Learning: A Case Study with Massage Instructors and Learners. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3706598.3713677

1 Introduction

Remote learning is effective for teaching a wide range of skills, from language acquisition [19, 80], STEM education [72, 78] to health and medical training [18, 21]. However, remote instruction faces some challenges when it comes to teaching hands-on skills—those that require precise techniques, physical practice, and tactile feedback. Traditional visual-audio methods, such as video lessons and online calls, fall short in replicating the nuances needed for effective training in tasks like mechanical assembly, cooking, or massage [18]. Massage is one such hands-on skill that is particularly important, as it serves both medical and everyday relaxation purposes, making proper instruction vital for professionals and enthusiasts alike [23, 57]. In massage instruction, instructors not only demonstrate the correct strokes but also guide students in applying the right amount of pressure. The preciseness of strokes and pressure control impact the massage's effectiveness. In these cases, inadequate instruction may lead to improper techniques, reducing therapeutic benefits or even causing harm [23]. Some researchers investigated the solutions to these challenges and developed various VR/AR-based remote learning systems [10, 18, 27, 71]. For example, Faridan et al.'s AR-based system used a surrogate to teach students by demonstrating gestures via AR [18]. However, though improved learning immersion, most existing solutions focus primarily on gesture recognition and neglect the crucial aspect of precise pressure control, which was particularly important for massage. This gap highlighted the need for a more advanced approach that integrates gesture detection, pressure control, and video instructions to enhance remote learning for hands-on skills.

In addition, according to the deliberate learning theory, real-time feedback from the instructors is necessary for hands-on skills instruction because the learners can immediately correct any mistakes [7]. The lack of timely feedback can reduce learners' efficiency in acquiring hands-on skills. However, due to time constraints, teachers are not always available to provide prompt instructions, which may hamper the learners' progress. LLMs, such as ChatGPT are widely used in educational scenarios, including maths [48, 81], programming [50, 58], language [17], and medicine [45, 61]. For learners, prior works demonstrated that LLMs could provide real-time feedback and customized learning experiences by quickly responding to questions [58, 70]. For instructors, LLMs could provide updates on the learners' progress, which may help them adjust their course and understand teaching effectiveness [75]. This inspired us to leverage LLMs as a bridge between instructors and learners—offering real-time feedback to support learners' progress in massage learning and aiding instructors in monitoring and guiding massage teaching.

Therefore, we identified two main challenges of current remote learning systems: (C1) Difficulty in conveying nuanced hands-on operations with video demonstration and audio expressions; (C2) Lack of real-time feedback for remote learning due to instructors' and learners' schedule conflicts. To address C1, we designed multimodal instructions consisting of video demonstrations, gesture, and pressure data to teach learners. To address C2, we designed an LLM-powered virtual teaching assistant (VTA) to provide the learners with real-time feedback when the instructors were absent and provide the instructors with the learners' performance summaries to support their instructions. Using the multimodal instructions and VTA, we investigated the following research question (RQ):

How might *multimodal instructions* and the LLM-powered *virtual teaching assistant (VTA)* support hands-on skills teaching and learning in the context of massage?

We conducted a case study with four massage instructors and twelve learners, who used multimodal instructions and VTA over five sessions. Traditional video instruction was also used as a baseline for comparison. We found that the multimodal instructions and VTA could resolve both challenges effectively. To resolve C1, we found that multimodal instructions could convey nuanced hands-on operations to improve the teaching and learning experience. For instructors, the multimodal instructions ensured detailed instruction content and ease of creation. For learners, the multimodal instructions ensured learning massage via clearer and standardized quantitative data and reduced their mental stress. For resolving C2, VTA's real-time feedback and learning performance summary further enhanced the teaching and learning process. For instructors, VTA improved their efficiency in analyzing performance records, provided grounded suggestions and comments, and made them focus more on critical aspects. For learners, VTA enhanced their learning confidence and could efficiently bridge the gap between theory and practice. This case study focused on massage, while we propose other applications of this system in the discussion, such as sign language and cooking instruction. In addition, we discussed the cognitive process differences in various courses that were also supported by LLM-powered VTA. Our contributions include:

- A synthesis tool using LLM-based VTA to integrate multimodal data for remote massage instruction and learning

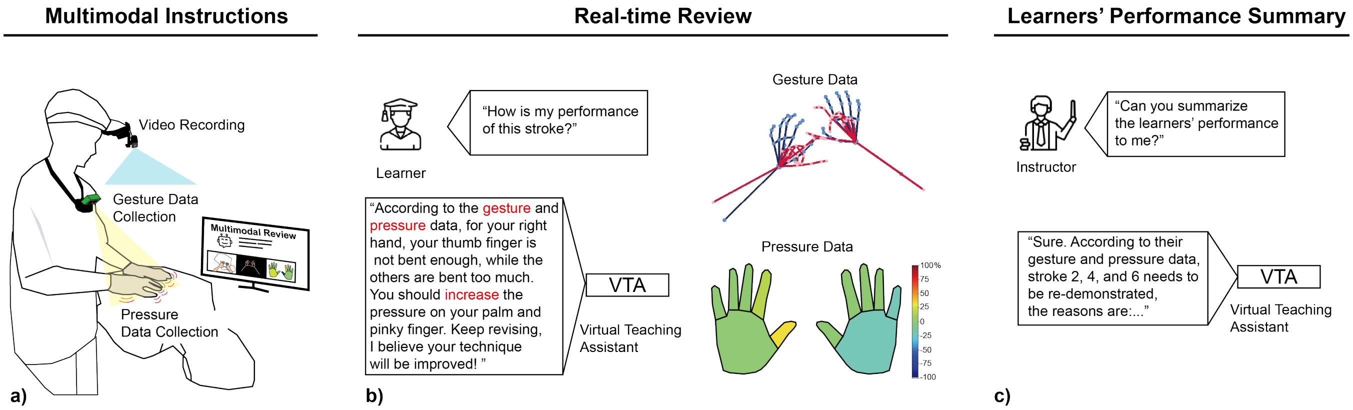

Figure 1: Overview of the Multimodal Instructions and the LLM-Powered Virtual Teaching Assistant (VTA) for hands-on skills training in the context of massage: (a) Multimodal Instructions: the learner wears the web camera, Leap Motion hand-tracking camera, and the pressure sensor while watching the multimodal review of their stroke performance; (b) Real-time review: the learner asks the VTA to review their massage performance. The VTA provides a review based on the gesture and pressure difference data. For the gesture difference data, the red hands belong to the learner while the blue hands are from the instructor's prerecorded instructions. For the pressure difference data, the color demonstrates the pressure difference between the learner's real-time pressure and the instructor's prerecorded pressure; c) Learners' performance summary: the instructor asks the VTA to summarize the learners' performance and recommends which massage strokes need to be redemonstrated.

- Empirical findings from a case study demonstrating the multimodal instructions and VTA's effectiveness in helping learners obtain real-time feedback and aiding instructors to analyze learners' performance;

2 Related Work

We describe prior remote learning systems and how LLMs are used for remote learning.

2.1 Remote Learning Systems

Numerous remote learning systems have been developed across various fields, such as language instruction [19, 80], STEM education [20, 72, 78], and health training [18, 21, 25, 57, 66]. These systems designed for theoretical or linguistic instructions primarily rely on visual and auditory channels for teacher demonstrations and instructions. However, teaching and learning hands-on skills, such as mechanical tasks, cooking, and massage, required precise operations, which were hard to convey through video demonstrations and audio descriptions. For example, massage involves precise stroke techniques and hand pressure control, requiring learners to manage pressure in both fingers and palms simultaneously or separately [23]. To address these challenges, researchers have explored VR/AR-based approaches for remote learning and collaboration [10, 14, 18, 27, 29, 30, 33, 59, 69, 71]. For instance, Faridan et al. developed an AR-based system where a surrogate mimics the instructor's gestures to teach skills like cooking, massage, and mechanical tasks [18]. However, despite their enhanced learning immersion, these systems often focused on gesture recognition and overlooked the crucial aspect of pressure control, which was particularly important for certain hands-on skills such as massage [18, 32]. To address these limitations, there is a need to design multimodal remote instructions that integrate video, stroke demonstrations, and pressure control, offering a more comprehensive teaching experience.

2.2 LLMs for Remote Learning

Previous research emphasized the importance of deliberate learning and the critical role of real-time feedback from instructors in correcting errors during hands-on skills acquisition [7]. Current remote learning systems for hands-on skills often invite an instructor or collaborator to provide real-time guidance [18, 27, 59, 71]. However, most of these works fail to address how to provide such support when instructors or collaborators are unavailable due to constraints like tight schedules, real-time network or other technical issues, and unavailability due to reasons like time zone differences or appointments forgetting. The absence of timely feedback may hinder learners' efficiency in acquiring hands-on skills, particularly when instructors are unavailable for extended periods.

LLMs, such as ChatGPT, are increasingly being adopted in educational settings across diverse domains, including STEM education [22, 40, 48, 49, 81], programming [11, 38, 39, 50, 58], language learning [17], and medical education [45, 61, 70]. These applications highlight the versatility of LLMs in facilitating learning and problem-solving. Research has identified several benefits of utilizing LLMs in education, such as personalized tutoring [9], automated essay grading [53, 79], language translation [6, 77, 82], interactive learning [16, 37], and adaptive learning [41, 65]. Building on these advancements, researchers have begun exploring the potential of LLMs in massage therapy and instruction. For example, Shen et al. reviewed the progress of machine learning techniques and highlighted the potential benefits of LLMs in traditional Chinese massage [63]. However, they also emphasized three key challenges:

difficulties in data access and labeling, issues with model training and optimization, and the necessity for clinical validation to establish credibility. Similarly, Baskwill et al. discussed the opportunities for incorporating LLMs into massage research, practice, and education, suggesting that these models could help instructors design personalized and interactive learning experiences or generate prompts and ideas for learning activities and assessments, as proposed by Baidoo-Anu and Ahsah [2, 3]. LLMs offer promising applications for both learners and instructors. For learners, they provide real-time feedback and personalized learning experiences, enabling efficient self-study even without direct instructor supervision. For instance, LLMs can deliver tailored answers, solutions, and suggestions, catering to individual needs throughout the learning process [58, 60, 70]. For instructors, LLMs enhance teaching by offering insights into student progress, refining course content, and improving teaching strategies [75]. Inspired by these needs and applications, we leveraged LLMs to bridge the gap between instructors and learners in massage education. By providing real-time feedback, our approach supports learners in mastering massage techniques while assisting instructors in delivering more effective teaching.

3 Multimodal Instructions and Virtual Teaching Assistant Design for Massage Teaching and Learning

We begin by outlining the challenges associated with remote learning and teaching hands-on skills, followed by presenting potential solutions. Next, we demonstrate the implementation of precise remote instructions for hands-on skills. Finally, we showcase the remote learning functionalities that support these processes.

3.1 Current Challenges and Potential Solutions of Hands-on Skills Remote Teaching and Learning

We identified two main challenges of current remote hands-on skills teaching and learning. From Section 2.1, we identified (C1) the difficulty in conveying nuanced hands-on operations with video demonstration and audio expressions. From Section 2.2, we identified (C2) the lack of real-time feedback for remote learning due to instructors' and learners' schedule conflicts.

We also derived design considerations around these challenges. To address C1: in addition to video instructions, the system should also capture data that convey hand and finger positions, which could inform learners whether their hands are in the right shape. The system should capture pressure data on the fingers and hands, which could allow learners to understand how much force they should apply when performing a stroke.

To address C2: the system should integrate an LLM-powered teaching assistant to provide the learners with real-time feedback when the learners were not available. Additionally, to ensure that the instructors understand their learning progress, the LLM-powered teaching assistant should provide the instructors with the learners' performance summary.

3.2 Implementation of precise remote hands-on skills instructions for massage

According to the design considerations, we developed a multi-modal remote learning system for massage, consisting of **point-of-view (POV) video recording**, **hand gesture acquisition**, **hand pressure collection**, and **Virtual Teaching Assistant (VTA)**, shown in Figure 2.

POV video recording. We used a head-mounted web camera to record the POV video instructions for massage strokes. The instructors' audio instructions for each stroke were also recorded.

Hand gesture acquisition. We used a Leap Motion hand-tracking camera worn on the neck to recognize the hands' locations and gestures of each stroke. The Leap Motion hand-tracking camera is a commercially available high-precision hand-tracking module equipped with two cameras and multiple infrared LEDs. It collects gesture data within the near-infrared spectrum using optical sensing technology [51]. Utilizing the embedded algorithm, the device could scan up to 200 frames per second with a wide field of view of 150°, which is sufficient for our massage instructions [24]. The devices could convert gesture information into each finger-joints' spatial coordinates. We obtained the positions of the joints and their flexion angles, demonstrated in Figure 3 a) and b).

Hand pressure collection. Based on prior works, we developed piezoresistive pressure sensors attached to the gloves that could be worn on the hands to measure the different strokes' pressure application on hand's five fingers and palm [36, 46]. The piezoresistive pressure sensors' resistance varies with applied pressure, allowing for the quantification of force through resistance measurements [76, 83]. We placed the sensors under the five fingers and palms individually, facilitating a detailed analysis of force dynamics during various massage techniques.

Virtual teaching assistant (VTA). We developed the VTA using OpenAI's GPT-4 model, which enabled quick responses throughout the massage instruction process [55]. The VTA generated personalized feedback for learners and instructors by analyzing differences in stroke gestures and pressure, as shown in Figure 2.

Specifically, the differences in gestures and pressure between the learners' performance and the instructors' recorded data were first computed via a program. For stroke and pressure data, we calculated each finger's flexion angle differences and each hand area's pressure differences between the learners' performance and the instructors' recorded data. The results of the gesture and pressure differences were then formatted in JSON and integrated into our prompts. To provide user-friendly feedback, the VTA identified the most significant portions of the difference data and generated corresponding reviews or summaries using the Chain-of-Thought prompting technique [74]. The detailed prompts are in Section 8.

Learners interact with the VTA to ask detailed questions after watching the instructor's demonstration videos, such as "what are the target muscles of this stroke?" and "what are the essences to perform this stroke?". A massage instruction database was incorporated into our prompts to provide professional massage domain knowledge [4]. After each stroke, learners could request real-time feedback in natural language, allowing them to refine their performance by comparing it with the instructor's standards.

Figure 2: The Multi-modal Remote Learning System for Massage Consists of Four Major Components: 1) POV Video Recording: records the instructions and learners' performance; 2) Hand gesture acquisition: collects the gesture data (joints' coordinates); 3) Hand pressure collection: collects hand pressure data from six areas of each hand (five fingers and palm); 4) VTA: provides real-time review to the learners and learners' summaries to the instructions.

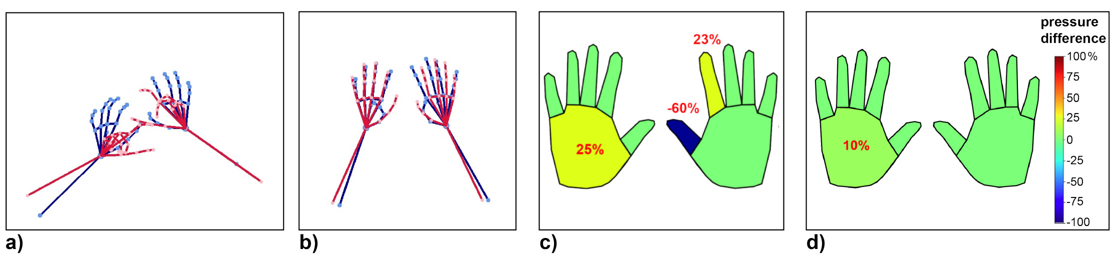

Figure 3: Improvements after VTA's Review and the Instructors' Feedback (Data from L3): a) Stroke before review/feedback; b) Stroke after review/feedback; c) Hands' pressure differences before review/feedback, shown in percentage relative to the instructors' recorded data; d) Hand's pressure differences after review/feedback. The figures demonstrate that after review/feedback, the learner's hands' locations and strokes were more similar to the instructors' recorded data, and the learner's pressure control became close to the instructors' data.

Instructors could request summaries of learners' performances from the VTA, either for all students or individual students. These summaries highlighted areas needing improvement, enabling the instructors to focus on specific strokes that require re-demonstration. The instructors could then make informed decisions based on both the VTA's suggestions and their own assessments.

3.3 Remote Learning Functions

Based on the prior works, such as the interactive online massage learning system [66] and the teleoperation system for remote instruction in hands-on skills [18], we designed seven functions classified into **Instructor Mode** and **Learner Mode** that are shown in Table 1.

The **instructor mode** consisted of three functions: *multimodal instruction record*, *video instruction record*, and *learner performance summary*. Instructors used *multimodal instruction record* function to record video instructions, along with gesture and pressure data. *Video instruction record* function allowed instructors to record only the video instructions. With *learner performance summary* function, instructors could request a performance summary of learners from the VTA, which provided an overview of all or individual learners' gesture and pressure data, highlighting areas for improvement.

The **learner mode** included four functions: *multimodal instruction learning*, *video instruction learning*, *learner performance record*, and *real-time review* The learners could use *multimdoal instruction learning* function to learn the massage strokes with multimodal instructions consisting of video instructions, gesture and pressure differences visualizations. The learners could use *video instruction learning* to learn the massage strokes through video demonstrations only. The learners could use *learner performance record* to record their performance for each massage stroke, capturing videos, photos, gestures, and pressure data of the current stroke. The learners

Table 1: Seven Functions in two modes: instructor mode, learner mode

Modes	Functions	Function Descriptions
Instructor Mode	Multimodal instruction record	Records the instructor's multimodal instructions
	Video instruction record	Records the instructor's video instructions
	Learner performance summary	Analyzes the learner's performance and provides a summary
Learner Mode	Multimodal instruction learning	Enables the learner to watch and learn from multimodal instructions
	Video instruction learning	Enables the learner to watch and learn from video instructions
	Learner performance record	Records the learner's performance during practice
	Real-time review	VTA provides real-time feedback and review to the learner

could use *real-time review* to request real-time feedback from the VTA, allowing for immediate review and correction of their current stroke performance.

4 Case Study

We conducted a within-subjects case study to understand how multimodal instructions and LLM-powered VTA's real-time feedback and learning performance summaries might help instructors and learners by comparing our multimodal instruction (MI) to traditional video instruction (VI). This research was approved by the university's ethics review board.

4.1 Participants and Apparatus

We recruited 12 massage learners (9 males, 3 females), denoted L1 to L12, and 4 massage instructors (1 male, 3 females), denoted I1 to I4, from the local community. All learners had no prior massage learning experience. The instructors had on average 9 (SD = 4.32) years of massage operation and instruction experience. Each instructor was randomly assigned three learners.

The experiment setup is illustrated in Figure 2. The user (instructor or learner) stood close to the human model with their hands and arms relaxed. The user wore the pressure sensor on the hands to record the pressure data, a web camera on the head to record stroke video, and a Leap Motion hand-tracking camera on the neck to recognize the hands' locations and gestures. A laptop to record the audio instruction or listen to the summary from the VTA was set adjacent to the human model. A display was set in front of the user for learning instructions and real-time review from the VTA.

4.2 Selection of Strokes

We selected eight strokes for four body positions, including shoulders, back, arms, and neck. The eight strokes were chosen based on six fundamental techniques from Swedish massage—compression, petrissage, effleurage, stripping, cross-fiber friction, and trigger-point therapy (circular friction)—along with four basic hand positions: flat palms, fingers and thumbs, hand over hand, and fingertips [4, 54]. These selections were further refined based on suggestions from experienced massage therapists.

For each body position, we selected 2 massage strokes, which are shown in Figure 4.

Stroke 1, shown in Figure 4 a), requires working along the length of **splenius capitis** and **splenius cervicis** muscles using muscle **stripping** stroke followed by friction to address areas of adhesions.

Stroke 2 shown in Figure 4 b), requires positioning the recipient on their back and following the steps of **effleurage**, **petrissage**, **stripping**, and **cross-fiber friction** to address tension.

Stroke 3 shown in Figure 4 c), is conducted to address the middle **trapezius** and **rhomboids** by tracing the length of the muscles with a **stripping** stroke.

Stroke 4 shown in Figure 4 d), uses a **circular friction** stroke to work from the top of the shoulders along the length of the shoulder blade, between the spine and **levator scapulae**.

Stroke 5 shown in Figure 4 e), is done by placing the **fingers and thumbs** for greater depth, applying pressure with the fingertips while gliding deeply along the **biceps**, **extensors**, and **flexors**

Stroke 6 shown in Figure 4 f), is conducted by applying **cross-fiber friction** along the length of the arm. This involves gliding the fingers across the muscles and moving slowly back and forth from the outside of the arm to the inside.

Stroke 7 shown in Figure 4 g), is conducted for the kneading stroke of **petrissage**, focusing on the **thenar eminence**. This involves pinching the **thenar eminence** gently between the thumb and forefinger and applying pressure in a circular motion.

Stroke 8 shown in Figure 4 h), involves working gently on each finger separately. **Stripping** and **compression** strokes can be used to work around the palm from the **thenar eminence** to the **hypothenar eminence**.

4.3 Procedure

The user study consisted of five sequential sessions: 1) **Instruction Record**, 2) **Remote Learning**, 3) **Instructor Feedback**, 4) **Review and Adjustment**, and 5) **Instructor Summarization**, as illustrated in Figure 5.

In the **Instruction Record** session, four instructors recorded VI and MI for each massage stroke. VI only included the video demonstration (with audio instruction). MI included a video demonstration (with audio instruction) and multimodal data (stroke photograph, hand area pressure data, and gesture data). The video demonstrations were recorded continuously, while the multimodal data were recorded in the stroke's several keyframes based on the instructors' own decisions. For example, they could have more keyframes for more complicated strokes and fewer keyframes when the stroke

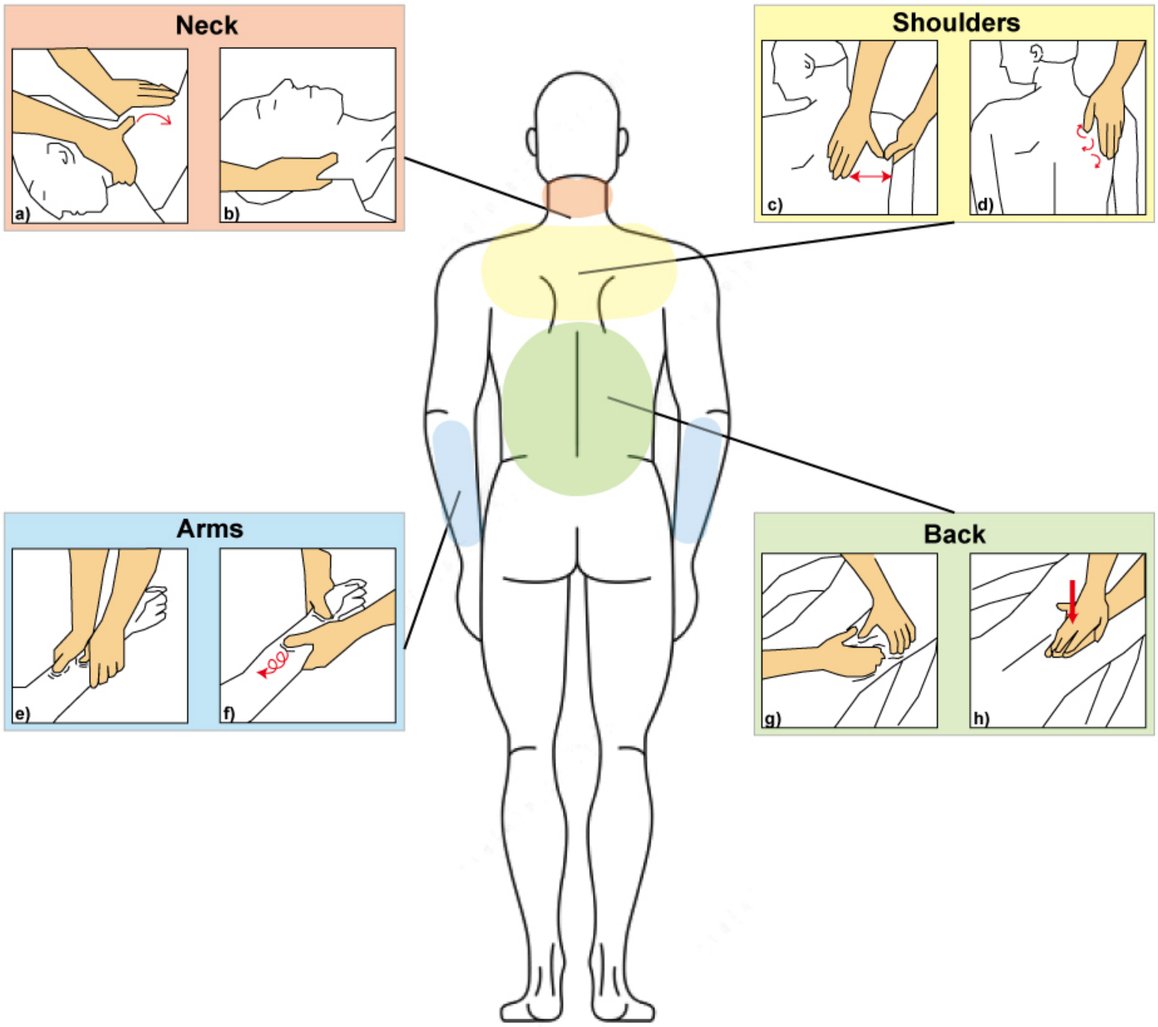

Figure 4: Massage Strokes: a) Stroke 1; b) Stroke 2; c) Stroke 3; d) Stroke 4; e) Stroke 5; f) Stroke 6; g) Stroke 7; h) Stroke 8.

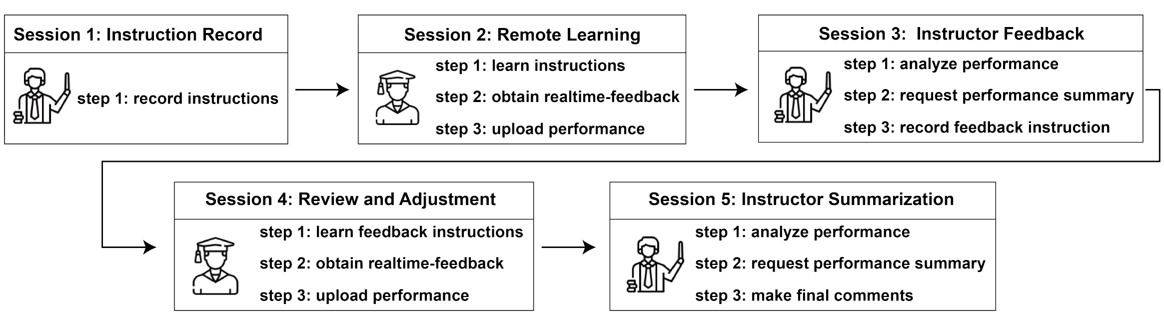

Figure 5: Five Sessions in User Study: a) Instruction record; b) Remote learning, c) Instructor feedback, d) Review and adjustment, and e) Instructor summarization.

could be easily conducted. Their keyframe selection stayed the same across the overall instruction process.

During **Remote Learning**, each learner learned 4 strokes using MI and the other 4 using VI, with the order alternating between MI and VI. The sequence was counterbalanced across participants using a Latin Square design (e.g., half of the learners started with MI, while the other half started with VI). After watching the instructions, the learners practiced the strokes, and their performances were recorded with the web camera, the Leap Motion hand-tracking camera, and the pressure sensor with the same data as the MI. Learners could review their recordings and revise their performance until satisfied. Learners using VI could watch the recorded video instruction. We also provided a PDF version of the massage textbook on the laptop in the experiment room [4]. In contrast, for strokes with MI, learners could review their performance by watching the recorded video performance, the strokes, and pressure difference visualizations, and asking VTA to comment on their performance. The final performance recording was then sent to the instructor. Learners rated confidence in their performance using a 7-point Likert scale (1 being not confident at all, 7 being fully confident).

In the **Instructor Feedback** session, the instructor used the VTA to understand each learner's performance by requesting a detailed summary. This summary included deviations in hand location, gesture data (with angle differences between the same finger of the instructor and the learner.), and hand area pressure differences, both individually and averaged across the three learners assigned to that instructor. The deviations were calculated by measuring the distance between the hands in the instructional videos and the hands during the learners' practice sessions. The stroke differences were calculated by comparing the finger angle differences. The pressure differences were calculated by comparing the pressure differences in different hand areas. Based on the summary provided by the VTA and their review, the instructors redemonstrated several strokes to help the learners revise their strokes and pressure application. The feedback for each stroke maintained the same condition (MI or VI) as the initial instructions.

In the **Review and Adjustment** session, learners revised their performance based on the instructors' feedback instructions. They could record their revised performance until satisfied, and their performance was then sent back to the instructors. Learners rated their confidence levels after receiving feedback, overall satisfaction with the system, the system's ease of use, the instruction's clarity, and the instruction's ease of use on a 7-point Likert scale for the MI and VI conditions.

Finally, in the **Instructor Summarization** session, the instructor reviewed each learner's final performance and the VTA's summary, which included comparisons with previous performances. They commented on the learners' performance improvement and the overall instruction process. Instructors rated their overall satisfaction with the system, instruction effectiveness, and the system's ease of use on a 7-point Likert scale for the MI and VI conditions.

After the sessions, both instructors and learners participated in semi-structured interviews to review their teaching and learning experiences.

4.4 Data Analysis

The feedback in each session and interviews with instructors and learners was transcribed using the transcription service from Zoom. Then a researcher rewatched the recording to correct any transcription errors. Two researchers independently coded the feedback and interviews using an open coding approach [15]. They then met to resolve any disagreements and consolidate codes. Afterward, they grouped the codes to identify the main themes on the instructors' experience and the main themes on the learners' experience. For all Likert ratings, we report the median and IQR. In addition, we conducted paired sample t-tests to determine whether the type of instruction significantly impacted learners' ratings.

5 Results

We report the results according to the multimodal instructions' and VTA's effectiveness in addressing C1 and C2 from two perspectives, shown in Figure 6: 1) Benefits of Multimodal Instructions for Nuanced Hands-on Operations; 2) Benefits of VTA's Real-time Feedback and Learning Performance Summary.

The instructors' and learners' user experience ratings are shown in Figure 7. All instructors were satisfied with the efficacy of the multimodal instructions and VTA's performance summary in teaching Massage. As shown in Figure 7 (top), instructors were more satisfied with MI (Md = 6.5, IQR: 1) than VI (Md = 5, IQR: 0.2). In addition, all learners were also satisfied with the efficacy of the multimodal instructions and VTA's real-time feedback in learning Massage. Paired sample t-tests revealed that learners were significantly more satisfied with MI (Md = 6, IQR: 0) compared to VI (Md = 4, IQR: 2), $t(11) = -6.66, p < 0.05$. The ratings are shown in Figure 7 (bottom).

5.1 Benefits of Multimodal Instructions for Nuanced Hands-on Operations

From our coding results, we identified two themes regarding how multimodal instructions benefited instructors and two themes for learners. The instructors' two themes included: *detailed instruction content from multimodal instructions*, and *multimodal instructions' ease of creation*. The learners' two themes included: *clearer and standardized quantitative data using multimodal instructions*, and *mental stress reduction from multimodal instructions.*

5.1.1 Detailed Instruction Content from Multimodal Instructions (For Instructors). All instructors reported that MI (Md = 6.5, IQR: 1) was more effective in teaching various massage strokes compared to VI (Md = 5.5, IQR: 1). During VI, instructors demonstrated the strokes while providing verbal explanations on how to perform the movements and apply pressure. However, this method often required vague descriptions, such as "apply more pressure here" or "stretch your hands further." In contrast, MI allowed for the quantitative measurement of strokes and pressure. Beyond only video demonstrations and audio descriptions, MI enabled instructors to offer more precise instructions that could not easily be conveyed with VI. I1 highlighted this by stating: *"Video instruction could only convey the basic elements of each massage stroke, such as hand placement and stroke technique. Traditionally, we would practice pressure control on learners' bodies, allowing them to feel the intensity directly, which is not possible with video instruction."*

5.1.2 Multimodal Instructions' Ease of Creation (For Instructors). Instructors reported that MI (Md = 6.5, IQR: 1) was slightly easier to use compared to VI (Md = 6, IQR: 1). I2 and I3 initially found that recording MI involved more steps than recording VI, such as wearing pressure sensors and setting up the Leap Motion camera.

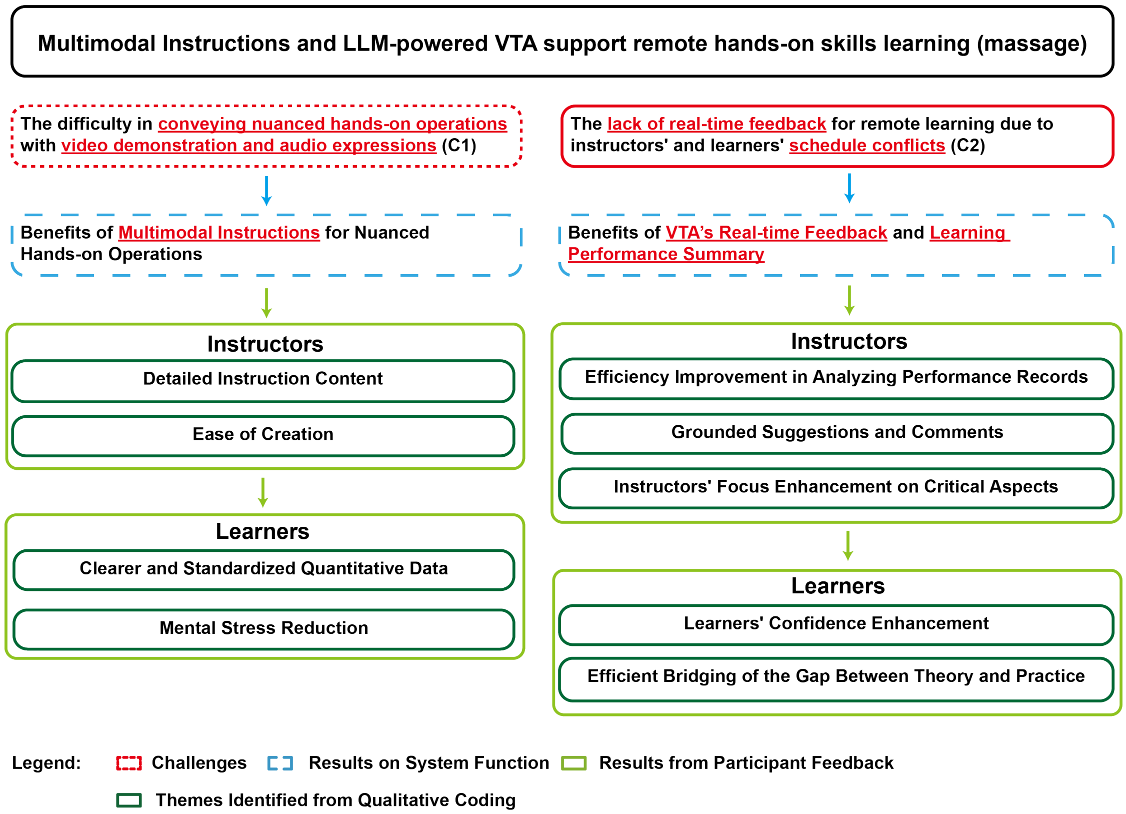

Figure 6: Summary of results showing how the multimodal instructions and VTA's real-time feedback and summary addressed the challenges and themes identified from qualitative coding.

They also noted a learning curve when first adapting to the system, especially because of their limited technical background. During this process, they spent time familiarizing themselves with the functions and operating procedures. However, once they mastered the technology, the operation became smooth and efficient. The simplicity of using a single button to start and stop the sensors made the operation less demanding than participants had anticipated.

I1 and I4 found that MI was easier to use than VI, mainly because they no longer needed to explain the vague details of stroke techniques and pressure control. I4 stated: *"When recording multimodal instruction, I can focus on the essence of each stroke and let the pressure and gesture data convey the basics."*

Additionally, the relative ease of creating MI was partly attributed to the instructors' educational backgrounds. Most instructors had not received a formal college education but had focused on mastering practical massage skills. This made it more difficult for them to teach complex theoretical concepts, but MI allowed them to focus on demonstrating hands-on techniques, which they found more intuitive and natural to teach.

5.1.3 Clearer and Standardized Quantitative Data using Multimodal Instructions (For Learners). The learners appreciated massage learning via quantitative data because it was clearer and standardized.

The learners reported that MI was clearer for learning massage strokes, with a median rating of 6 (IQR: 1), which was significantly higher than VI (Md = 5, IQR: 1.2), $t(11) = -5.42, p < 0.05$. They mentioned that the MI provided visualizations of necessary data, including the instructors' video demonstrations, and the stroke and pressure differences displayed on the interface. The multimodal data helped learners easily assess their performance and make precise corrections. By contrast, while VI offered clear verbal and visual demonstrations, they only provided a rough overview of each stroke, making them less effective for detailed understanding. L6 mentioned that: *"When I learned massage by myself, it was hard to perform precise strokes and I did not know where to apply pressure. This tool allows me to understand exactly where I should revise."*

In addition, they appreciated the standardized instructions provided by MIs. They mentioned their previous video-based remote learning experience that, since each instructor taught from a different perspective, learners had to watch multiple videos to understand the essence of the same stroke, leading to confusion and inconsistency in their learning. For example, L11 mentioned: *"I watched several videos of the same stroke, but each one had slightly different hand positions or stroke motions, which left me unsure of the correct technique."* This lack of standardization often resulted in

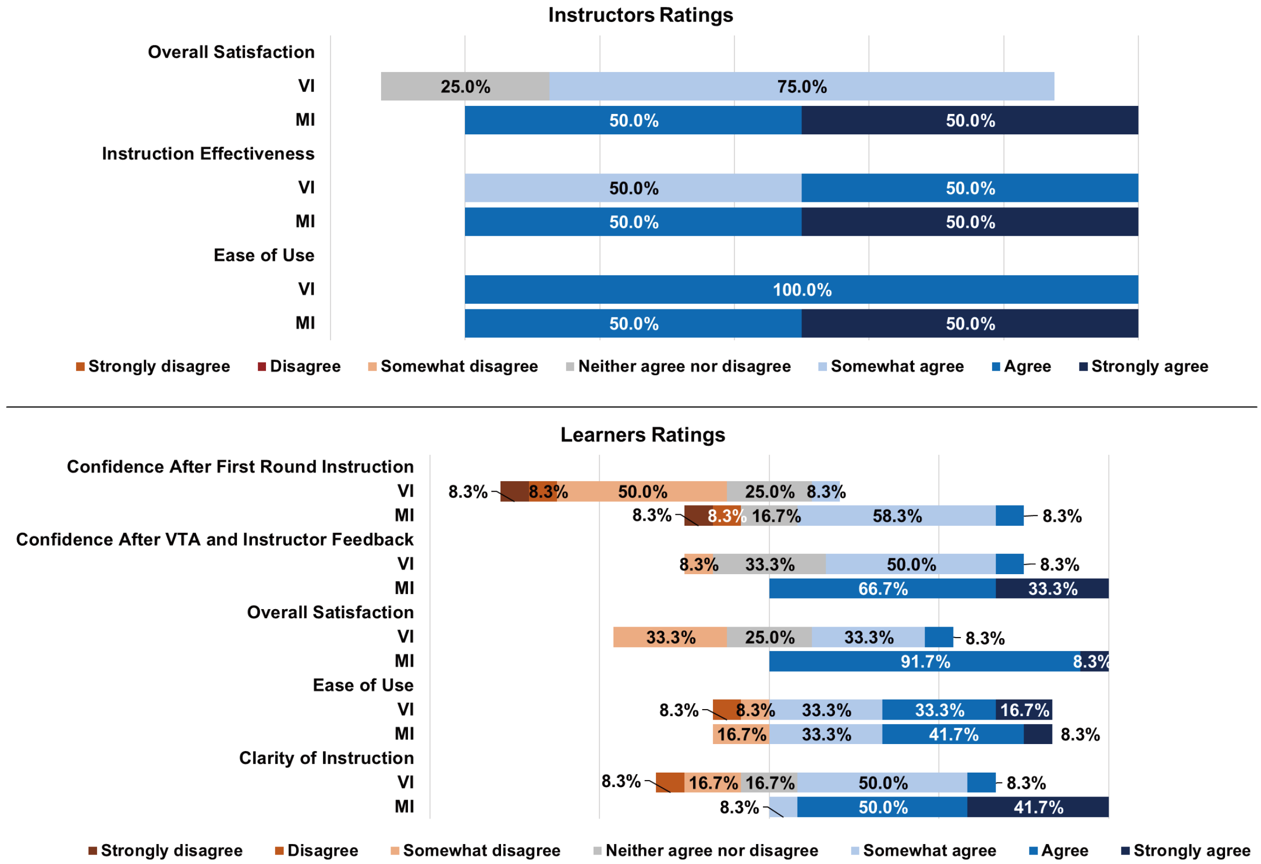

Figure 7: Diverging stacked bar chart showing the subjective ratings from Instructors and Learners: (top) Instructors ratings: overall satisfaction, instruction effectiveness, and ease of use; (bottom) Learners ratings: confidence after first round instruction, confidence after VTA and instructor feedback, overall satisfaction, ease of use, and clarity of instruction. For both instructors and learners, MI achieved higher median ratings.

gaps in their understanding, making their learning process more challenging and imprecise.

5.1.4 Mental Stress Reduction from Multimodal Instructions (For Learners). Most learners reported that learning through MI decreased their mental stress compared to traditional in-person methods. In traditional in-person learning, learners were often required to practice massage techniques on the bodies of instructors or partners, while being observed by others. This public environment created anxiety, as learners feared making mistakes in front of their peers, instructors, or partners, leading to higher stress levels and a less focused learning experience.

MI, however, offered a private learning experience where learners could practice alone. Without the pressure of being watched, learners felt more at ease and could concentrate better on their strokes and pressure control. One learner explained that practicing on a human model, rather than a real person, also reduced the stress of potentially causing discomfort or harm. The private and controlled environment allowed them to learn without fear of judgment or failure. L5 commented on how this environment improved their focus: *"Learning alone makes me more focused on the stroke and less likely to be disturbed by others. I preferred this learning method over the crowded classroom environment."* In this case study, VI achieved similar results, as the social pressure present in in-person instruction was also reduced with VI.

5.2 Benefits of VTA's Real-time Feedback and Learning Performance Summary

From our coding results, we identified three themes regarding how the VTA helped instructors and two themes for learners. The instructors three themes included: *efficiency improvement in analyzing performance records via VTA*, *grounded suggestions and comments from the VTA*, and *enhanced focus on critical aspects using VTA* . The learners two themes included: *enhanced confidence using VTA's real-time feedback and instructors' feedback*, and *bridging the gap between theory and practice through VTA's real-time feedback.*

5.2.1 Efficiency Improvements in Analyzing Performance Records via VTA (For Instructors). The VTA improved instructors' efficiency in reviewing learners' performance records. Initially, instructors individually analyzed each learner's performance records. Over time, they began to rely on the VTA to provide comments and recommend which strokes required further instruction or demonstration. Instructors based their decisions on the VTA's feedback and suggestions.

Each instructor was responsible for three learners, with each learner having eight performance records. In total, this meant each instructor needed to review 24 performance records, which typically took at least 30-45 minutes to analyze and summarize without the VTA. The VTA's feedback and recommendations helped reduce the time spent reviewing these records. Some instructors noted that the VTA saved more than 20 minutes in the analyzing and summarizing process (I1, I4). They also mentioned that the VTA would be even more beneficial with a larger number of learners. As I1 observed: *"Although I could review the performance records myself, it is much faster to see or hear the VTA's summary."*

5.2.2 Grounded Suggestions and Comments from the VTA (For Instructors). The VTA could provide grounded suggestions and comments based on learners' performance data. Most instructors, except I2, expressed confidence in the VTA's recommendations because they were grounded in detailed performance data for each learner.

For instance, I3 compared her own summary of stroke 8 with the VTA's summary and found the VTA's analysis more accurate. The VTA highlighted subtle differences that I3 had overlooked, such as the varying pressure applied to different parts of the hand. I3 had only focused on the pressure applied to the palm but missed that some learners exerted excessive pressure on the pinky finger. I4 also emphasized the VTA's effectiveness in summarizing quantitative data, stating: *"The VTA reviews all the data and summarizes learners' performance. I believe that AI can summarize quantitative data, such as gesture angles and pressure, better than me."*

However, I2 expressed concerns about over-relying on the VTA. She felt that while the AI's comments and suggestions were based on comprehensive data, it was still important for instructors to understand individual learners' differences in stroke techniques and pressure control. I2 felt it was more responsible to customize feedback instructions to each learner's drawbacks instead of recording common instructions for all learners based on the VTA's summary.

We also observed that the VTA occasionally hallucinated by generating inaccurate summaries of learners' performance. It tended to exhibit a lenient and overly favorable bias towards learners' performance. For instance, I3 pointed out that the VTA inaccurately described a learner's stroke as "perfect," despite her observation that some fingers were insufficiently bent. This discrepancy likely arose from the VTA's reliance on calculated data, which categorized degree differences smaller than 2 degrees as negligible. This threshold may have seemed reasonable when the required adjustment exceeded 20 degrees, as 2 degrees represented less than 10% of the total correction needed. However, I3, relying on photos and videos, could detect subtle variations that the VTA's numerical thresholds overlooked. To mitigate this issue, I3 asked follow-up questions by prompting the VTA to provide detailed degree differences. By offering precise numerical data, the VTA allowed instructors to identify and address subtle discrepancies.

5.2.3 Enhanced Focus on Critical Aspects using VTA (For Instructors). The VTA aided instructors by handling the correction of basic errors, allowing them to concentrate on more professional aspects of massage strokes. For instance, instructors could focus on enhancing the flexibility of learners' strokes and improving the continuity between different keyframes of a stroke to reduce recipients' nervousness and improve their overall experience.

By identifying and addressing obvious mistakes, such as incorrect hand positions and stroke deformations, during real-time reviews, the VTA streamlined the error correction process. This efficiency enabled instructors to focus on more nuanced, professional-level errors that the VTA might not detect. For example, they could study the flexibility of learners' strokes and the continuity of different strokes because it affected the recipients' experience. I4 expressed appreciation for this functionality, noting: *"The VTA is a valuable teaching assistant that reduces the repetitive task of correcting basic mistakes. It allows me to focus on more intricate errors that require my professional insight."* This approach enhanced the overall teaching process by ensuring that both basic and advanced errors are addressed effectively.

5.2.4 Enhanced Confidence using VTA's Real-time Feedback and Instructors' Feedback (For Learners). MI enhanced the learners' confidence in learning massage strokes after they received the VTA's real-time feedback and the instructors' feedback instructions.

Before receiving the VTA's review and the instructors' feedback (after first round instruction), learners reported significantly higher confidence when using MI (Md = 5, IQR: 1) compared to VI (Md = 3, IQR: 1), $t(11) = -5.63, p < 0.05$. After receiving the review and feedback (after VTA and instructor feedback), their learning confidence increased, and MI (Md = 6, IQR: 1) was significantly higher than VI (Md = 5, IQR: 1), $t(11) = -8.04, p < 0.05$.

The learners attributed their increased confidence to the VTA's real-time review, which helped them immediately correct basic errors, such as improper hand positions or stroke deformations. The VTA provided user-friendly feedback in natural language, much like a patient human instructor. Additionally, learners could ask the VTA about the essence of each stroke and other related questions, such as the stroke's benefits for the recipient or detailed information about the muscles involved. This extra information deepened their understanding of each stroke, giving them a stronger sense of truly learning something valuable.

Furthermore, after receiving detailed feedback instructions from instructors, learners could identify and correct more professional-level mistakes that the VTA had not detected, such as fine-tuning pressure applications in specific areas and stroke flexibility adjustments. L3 highlighted this by saying: *"I thought I had done well after the VTA's review, but my instructor pointed out more nuanced mistakes that I hadn't noticed. After receiving that feedback, I felt much more confident in my performance."* In contrast, with VI, they could only imitate the instructor's movements without fully understanding the appropriate strokes and related information that was not told by the instructors, making it harder for them to believe they had mastered the strokes.

5.2.5 Bridging the Gap Between Theory and Practice through VTA's Real-time Feedback (For Learners). VTA's real-time feedback could efficiently bridge the gap between theory and practice and thus could enhance the learners' massage learning efficiency.

In traditional remote learning, learners could only practice on themselves or friends and family members, but they lacked professional feedback. While they could sense basic aspects like hand placement or pressure, they did not receive expert guidance or corrections, limiting their ability to improve.

In traditional in-person learning, learners often struggled to immediately apply classroom lessons because they had to wait for instructors or partners to be available for practice. This process was inefficient and required multiple sessions with different people to fully connect theory with practice, which is time-consuming.

MI, however, addressed these issues by providing immediate, real-time feedback through the VTA, allowing learners to correct basic mistakes and refine their techniques independently. Once their practice reached a certain level, instructors could offer more in-depth, professional insights, focusing on subtler aspects of the massage strokes. This process accelerated the learning cycle, making it easier to integrate theoretical concepts with practical skills.

6 Discussion

We conducted a user study to explore how our multimodal instructions and LLM-powered VTA could help hands-on skills teaching and learning in the context of massage. The study revealed that multimodal instructions and LLM-powered VTA improved both the instructors' teaching efficiency and the learners' learning outcomes by effectively resolving the two challenges : (C1) the difficulty in conveying nuanced hands-on operations with video demonstration and audio expressions; (C2) the lack of real-time feedback for remote learning due to instructors' and learners' schedule conflicts. To address C1, we found that multimodal instructions could convey nuanced hands-on operations to improve the teaching and learning experience. To address C2, we found that VTA's real-time feedback and learning performance summaries could enhance the teaching and learning experience. We then discuss our work from three perspectives: *extending multimodal instructions to more hands-on skills teaching*, *cognitive process differences in various courses supported by LLM-powered TAs*, and *strategies to address incorrect responses from the VTA*.

6.1 Extending Mutimodal Instructions to More Hands-on Skills Teaching

Our results confirmed that multimodal instructions, which incorporate both gesture and hand pressure data, are effective in conveying nuanced operations for hands-on skills. These findings align with prior work suggesting that adding haptic feedback and accurate gesture tracking improves the teaching of hands-on skills [18]. However, the data required for effective instruction vary depending on the nature of the skill being taught. To illustrate this, we compare the instructional needs of three distinct hands-on skills: sign language, massage, and cooking.

Sign language instruction demands precise gesture capture and facial expression recognition. In American Sign Language (ASL), the same hand movement can have different meanings depending on the facial expression that accompanies it [43, 44]. Therefore, any multimodal instructions for sign language must include both hand and facial expression tracking to ensure accurate communication. Based on the themes identified in Section 5, all themes for instructors can be generalized to sign language, as it also requires specific movements, similar to massage. For learners, all themes except for clearer and standardized quantitative data can be generalized, since sign language does not rely on precise angle and pressure calculations—what matters is that the sign is correct and understandable [44].

Massage instruction, in contrast, focuses on the precise control of strokes and hand pressure. This requires multimodal instructions to collect and compare detailed data on both hand gestures and pressure. Video demonstrations supplement this data by showing how different strokes are connected to maintain the continuity of the massage, which is essential for the recipient's comfort [4, 18].

Cooking lessons, on the other hand, involve both simple (e.g., washing vegetables) and complicated gestures (e.g., chopping and slicing) that require various levels of precision according to the gesture selections. In addition, it is important to follow the correct sequence of actions. Therefore, multimodal instructions for cooking should focus on tracking the gestures, timing, and order of hand movements, ensuring learners understand the correct gestures and proper steps to complete a dish successfully [18, 26, 64]. For instructors, the theme of improved efficiency in analyzing performance records may be less relevant because, in cooking, the focus is more on the final result of the dish rather than on the precise movements during the process [26, 64]. Similarly, for learners, the theme of clearer and standardized quantitative data may not be as applicable in cooking, where the emphasis is more on timing and sequence rather than specific angles or pressure differences.

In summary, while the core concept of multimodal instructions remains consistent, the specific data and sensor requirements vary depending on the skill being taught. Instructors need to identify the key aspects of each skill—whether it be gesture precision, precise pressure control, or step sequencing—and adapt their multimodal instructions accordingly to provide the most effective learning experience.

6.2 Cognitive Process Differences in Various Courses Supported by LLM-powered TAs

We found that the VTA effectively supports the teaching and learning of hands-on skills, particularly massage instruction. Prior works have also explored LLM-powered TAs in various fields, such as sciences [1, 40, 47], maths [68, 73], and languages [17, 31]. Many of these works mentioned the importance of real-time feedback from the LLM-powered TA to the learning process.

Maiti et al. analyzed student questions posed to LLM-powered TAs across different disciplines, classifying them into six cognitive levels according to Bloom's taxonomy: Remember, Understand, Apply, Analyze, Evaluate, and Create. The types of questions reflect the learner's cognitive process, from basic to advanced levels [5, 52, 56]. Our study found that learners in massage instruction often asked questions at the "Analyze" and "Evaluate" levels, such as, *"How much more pressure should I add to the index finger"* or *"How much more should I bend my ring finger to match the instructor's* *stroke?"* In contrast, learners in sciences, maths, or language courses frequently ask questions that reach the highest "Create" level, such as how to design new solutions based on theoretical knowledge. This discrepancy may stem from two key factors: differences in course goals and potential negative effects.

Course Goals. Hands-on skills require understanding precise operations and step-by-step execution, so learners focus more on "how" rather than "why." In contrast, sciences, maths, and language learners are often expected to understand both "how" and "why" to eventually create something new. Therefore, it is appropriate that hands-on learners focus on "Analyze" level questions, especially in massage instruction.

Potential Negative Effects. In hands-on instructions like massage, machine assembly, cooking, and haircutting, incorrect execution may lead to harm to the operators and recipients, making the precision of operations crucial. Conversely, in sciences, maths, and language courses, the worst outcome is often limited to poor academic performance. As a result, hands-on skills learners prioritize nuanced operational questions, such as strokes and pressure control, over theoretical inquiries. In our study, learners frequently asked how to match the instructor's strokes, focusing on minimizing risks rather than exploring the theoretical basis for each stroke.

6.3 Strategies to Address Incorrect Responses from the VTA

Our findings revealed that the VTA provided performance summaries that contained inaccuracies, contradicting the quantitative data. These inaccuracies are consistent with the phenomenon of "factual contradiction" identified in previous studies, where LLM outputs are based on real-world information but introduce errors or inconsistencies [28]. Similar challenges have been reported in fields like massage diagnosis. For example, Chen et al. noted the risks of LLM misdiagnosis during the massage process, attributing these errors to prediction mechanisms and training data that may not generalize well to individual cases [63]. Future research could focus on making the VTA's summaries more accurate by incorporating quantitative data and prompting it to be more precise.

Additionally, participants asked follow-up questions when they suspected a hallucination, but if they missed the error, it could lead to inaccurate outcomes. This suggests that future VTAs could benefit from incorporating cognitive forcing functions—strategies that prompt critical thinking during decision-making and have been shown to help users recognize AI errors and reduce overreliance on AI [8]. For instance, the VTA could prompt instructors to ask follow-up questions or reference instructional materials, which would help mitigate overreliance on the VTA's summaries.

6.4 System Improvement

Currently, our system primarily supports strokes performed with the fingertips and palms. However, certain types of massage, such as Thai massage, involve using additional body areas, such as the knees or feet, to assist in stretching or applying pressure. To effectively teach these techniques, it is important to capture pressure data from various body areas and whole-body postures. Since wearing traditional stiff pressure sensors across the entire body may be impractical, future research could investigate alternative methods for supporting a wider range of body areas and postures, while maintaining the freedom of movement essential for massage instruction. Therefore, there is a need for flexible sensors with high deformation capacity and biocompatibility [62, 67]. Refereded from prior werable devices, these sensors can be worn on various body areas with excellent conformability, enabling the collection of pressure data for a wider range of massage strokes, [13, 34, 35, 62]. For instance, Sharma et al. developed a wearable pressure-sensing system using MXene composite nanofibrous scaffolds, capable of detecting pressure and strain on areas such as the wrist, throat, and face [62].

Additionally, we collected and analyzed the pressure data from five fingers and the palms of both hands. However, to capture more nuanced variations in pressure distribution across the smaller areas in palms or fingers, we need to develop a pressure sensing system with higher spatial resolution. To achieve this goal, improved fabrication processes and materials need to be explored [12, 42]. For example, Chen et al. developed a high-resolution sensing array based on piezoresistive strain transducers, achieving a sensitivity of 0.13 kPa^{-1} [12]. This advancement may enable a more refined understanding of pressure dynamics during massage, facilitating improved feedback and instruction methods.

Moreover, a challenge of our system is its reliance on a head-mounted camera, a Leap Motion hand-tracking camera, and piezoresistive pressure sensors attached to gloves. This setup is resource-intensive and could be costly for practical applications. In the future, we aim to reduce the overall system cost to make it scalable and more accessible for learners and instructors. For instance, we plan to explore the use of cost-effective cameras for gesture recognition and develop lower-cost sensor fabrication techniques.

6.5 Limitations and Future Work

In this work, we used massage as a case study to explore the effectiveness of an LLM-based synthesis tool that integrates and provides real-time feedback for teaching and learning hands-on skills. Consequently, we addressed several limitations specific to massage instruction in our current study and proposed directions for future development.

This case study focused on two conditions: MI (which included multimodal information and an LLM-supported VTA) and VI (which included video and audio information), with VI serving as a baseline representing current real-world practices. Future research could explore additional conditions to isolate specific components of MI, such as pressure differences or gesture differences. This would allow for a more detailed evaluation of which component is most beneficial for remote massage instruction.

In addition, each learner in our study received instruction from only one instructor. Future research could examine whether learning from multiple instructors, who might perform the same massage stroke with slight variations, causes confusion, and explore ways to address the challenges this may introduce. For instance, pressure and gesture data might need to accommodate multiple reference points rather than relying on a simple binary comparison between one instructor and the learner.

Furthermore, this study focused on massage instruction with a limited sample of four instructors and twelve learners. In the future, we aim to expand this work through both large-scale and longitudinal studies. For a large-scale study, we plan to involve a greater number of instructors and learners while examining additional factors, such as gender, age, and variations across different types of massage, including Thai massage and Chinese traditional massage. For a longitudinal study, we intend to explore how our system supports learners and instructors over an extended learning period. This will include comparing the effectiveness of our system with traditional face-to-face and remote instruction methods to better understand its long-term benefits and limitations.

Additionally, we plan to employ both theoretical and practical tests to evaluate the effectiveness of the instruction [18]. For the theoretical tests, we aim to assess learners' understanding of massage-related theoretical knowledge, including muscle and skeletal anatomy, as well as the indications and contraindications of massage. For the practical tests, we intend to evaluate learners' performance on real massage recipients and gather feedback from both the recipients and experienced massage instructors. Various parameters will be assessed to ensure a comprehensive evaluation. For the learners, we will examine factors such as the continuity and diversity of their strokes and whether they apply the appropriate strokes to the correct positions. For the massage recipients, we will measure physiological changes, including muscle tension, blood pressure, and heart rate, both before and after the massage.

7 Conclusion

In conclusion, we conducted a within-subjects case study to understand how a synthesis tool an LLM-based VTA to integrate multimodal data for remote massage instruction and learning might help instructors and learners by comparing it to traditional video instruction. The results showed that our system improved the instructors' and the learners' teaching and learning experience by effectively resolving the two challenges: (C1) the difficulty in conveying nuanced hands-on operations with video demonstration and audio expressions; (C2) the lack of real-time feedback for remote learning due to instructors' and learners' schedule conflicts. For C1, we found that multimodal instructions could convey nuanced hands-on operations to improve the teaching and learning experience. For C2, we found that VTA's real-time feedback and learning performance summaries further enhanced the teaching and learning process. Moreover, we discussed extending multimodal instructions to more hands-on skills teaching and cognitive process differences in various courses supported by LLM-powered TA. Our work serves as a valuable reference for the future development of multimodal remote instructions and LLM-powered virtual teaching assistants to facilitate the teaching and learning of other hands-on skills.

Acknowledgments

This work is partially supported by the Guangzhou-HKUST(GZ) Joint Funding Project (No. 2024A03J0617), Guangzhou Higher Education Teaching Quality and Teaching Reform Project (No. 2024YBJG070), Education Bureau of Guangzhou Municipality Funding Project (No. 2024312152), Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No. 2023B1212010007), the Project of DEGP (No.2023KCXTD042), the Guangzhou Science and Technology Program City-University Joint Funding Project (No. 2023A03J0001), and Guangdong Scientific Research Platform and Projects for the Higher-educational Institution & Education Science Planning Scheme (No. 2023KTSCX170).

References

[1] Bashaer Alsafari, Eric Atwell, Aisha Walker, and Martin Callaghan. 2024. Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants. *Natural Language Processing Journal* 8 (2024), 100101. doi:10.1016/j.nlp.2024.100101

[2] David Baidoo-anu and Leticia Owusu Ansah. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI* 7, 1 (2023), 52–62. doi:10.61969/jai.1337500

[3] Amanda Baskwill. 2023. Navigating Generative AI: Opportunities, Limitations, and Ethical Considerations in Massage Therapy and Beyond. *International Journal of Therapeutic Massage & Bodywork* 16, 4 (2023), 1. doi:10.3822/ijtmb.v16i4.949

[4] Rachel Beider. 2022. *Press Here! Massage for Beginners.* Fair Winds Press.

[5] Benjamin Samuel Bloom. 2010. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

[6] Thorsten Brants, Ashok Popat, Peng Xu, Franz Josef Och, and Jeffrey Dean. 2007. Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).* 858–867. https://aclanthology.org/D07-1090.pdf [Last accessed: 27/01/2025].

[7] Larike H Bronkhorst, Paulien C Meijer, Bob Koster, and Jan D Vermunt. 2014. Deliberate practice in teacher education. *European Journal of Teacher Education* 37, 1 (2014), 18–34. https://doi.org/10.1080/02619768.2013.825242

[8] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 188:1–188:21. doi:10.1145/3449287

[9] William Cai, Josh Grossman, Zhiyuan Jerry Lin, Hao Sheng, Johnny Tian-Zheng Wei, Joseph Jay Williams, and Sharad Goel. 2021. Bandit algorithms to personalize educational chatbots. *Machine Learning* 110, 9 (2021), 2389–2418.

[10] Yuanzhi Cao, Anna Fuste, and Valentin Heun. 2022. MobileTutAR: a Lightweight Augmented Reality Tutorial System using Spatially Situated Human Segmentation Videos. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. New York, NY, USA, Article 396, 8 pages. doi:10.1145/3491101.3519639

[11] Bei Chen, Fengji Zhang, Anh Nguyen, Daoguang Zan, Zeqi Lin, Jian-Guang Lou, and Weizhu Chen. 2022. CodeT: Code Generation with Generated Tests. arXiv:2207.10397 [cs.CL]

[12] Minrui Chen, Weifeng Luo, Zhongqi Xu, Xueping Zhang, Bo Xie, Guanghou Wang, and Min Han. 2019. An ultrahigh resolution pressure sensor based on percolative metal nanoparticle arrays. *Nature communications* 10, 1 (2019), 4024. https://www.nature.com/articles/s41467-019-12030-x.pdf [Last accessed: 27/01/2025].

[13] Yanjun Chen, Xuewei Liang, Si Chen, Yuwen Chen, Hongnan Lin, Hechuan Zhang, Chutian Jiang, Feng Tian, Yu Zhang, Shanshan Yao, and Teng Han. 2022. HapTag: A Compact Actuator for Rendering Push-Button Tactility on Soft Surfaces. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 70, 11 pages. doi:10.1145/3526113.3545644

[14] Neil Chulpongsatorn, Thien-Kim Nguyen, Nicolai Marquardt, and Ryo Suzuki. 2024. HoloDevice: Holographic Cross-Device Interactions for Remote Collaboration. arXiv:2405.19377 [cs.HC]

[15] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1 (1990), 3–21. https://doi.org/10.1007/BF00988593

[16] Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jianghao Yin, Jiaju Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, Aimin Zhou, Ze Zhou, Qin Chen, Jie Zhou, Liang He, and Xipeng Qiu. 2023. EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education. arXiv:2308.02773 [cs.CL]

[17] Yaxin Fan, Feng Jiang, Peifeng Li, and Haizhou Li. 2023. GrammarGPT: Exploring Open-Source LLMs for Native Chinese Grammatical Error Correction with Supervised Fine-Tuning. arXiv:2307.13923

[18] Mehrad Faridan, Bheesha Kumari, and Ryo Suzuki. 2023. ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. New York, NY, USA, Article 203, 13 pages. doi:10.1145/3544548.3581381

[19] Mais Farkhadov and Sofiya Khayitova. 2019. How a Multilingual Remote Teaching System Can Take into Account the Specifics of National Education. In *2nd International Conference on Education Science and Social Development (ESSD 2019)*. Atlantis Press, 230–235. https://doi.org/10.2991/essd-19.2019.51

[20] Salome Flegr, Jochen Kuhn, and Katharina Scheiter. 2023. How to foster STEM learning during Covid-19 remote schooling: Combining virtual and video experiments. *Learning and Instruction* 86 (2023), 101778. doi:10.1016/j.learninstruc.2023.101778

[21] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Article 662, 14 pages. doi:10.1145/3411764.3445576

[22] Mor Geva, Ankit Gupta, and Jonathan Berant. 2020. Injecting Numerical Reasoning Skills into Language Models. arXiv:2004.04487

[23] Richard Gold. 2006. *Thai massage: a traditional medical technique.* Elsevier Health Sciences.

[24] Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. 2019. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. *Sensors* 19, 5 (2019). doi:10.3390/s19051072

[25] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi coaches: an augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference (AH '17)*. New York, NY, USA, Article 25, 4 pages. doi:10.1145/3041164.3041194

[26] Meng He, Chen Qiu, Zhanghua Liao, Zhongquan Sui, and Harold Corke. 2018. Impact of cooking conditions on the properties of rice: Combined temperature and cooking time. *International Journal of Biological Macromolecules* 117 (2018), 87–94. doi:10.1016/j.ijbiomac.2018.05.139

[27] Gaoping Huang, Xun Qian, Tianyi Wang, Fagun Patel, Maitreya Sreeram, Yuanzhi Cao, Karthik Ramani, and Alexander J. Quinn. 2021. AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Article 417, 15 pages. doi:10.1145/3411764.3445283

[28] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2024. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Trans. Inf. Syst.* (Nov. 2024). doi:10.1145/3703155

[29] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and Ryo Suzuki. 2023. HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. New York, NY, USA, Article 119, 12 pages. doi:10.1145/3586183.3606727

[30] Keiichi Ihara, Kyzyl Monteiro, Mehrad Faridan, Rubaiat Habib Kazi, and Ryo Suzuki. 2024. Video2MR: Automatically Generating Mixed Reality 3D Instructions by Augmenting Extracted Motion from 2D Videos. arXiv:2405.18565 [cs.HC]

[31] Jaekwon Park, Jiyoung Bae, Unggi Lee, Taekyung Ahn, Sookbun Lee, Dohee Kim, Aram Choi, Yeil Jeong, Jewoong Moon, and Hyeoncheol Kim. 2024. How to Align Large Language Models for Teaching English? Designing and Developing LLM based-Chatbot for Teaching English Conversation in EFL, Findings and Limitations. (2024). doi:10.13140/RG.2.2.13490.82883

[32] Rahul Jain, Jingyu Shi, Andrew Benton, Moiz Rasheed, Hyungjun Doh, Subramanian Chidambaram, and Karthik Ramani. 2024. Visualizing Causality in Mixed Reality for Manual Task Learning: An Exploratory Study. arXiv:2310.13167 [cs.HC]

[33] Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–18. https://doi.org/10.1145/3586183.3606793

[34] Chutian Jiang, Yanjun Chen, Mingming Fan, Liuping Wang, Luyao Shen, Nianlong Li, Wei Sun, Yu Zhang, Feng Tian, and Teng Han. 2021. Douleur: Creating Pain Sensation with Chemical Stimulant to Enhance User Experience in Virtual Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 66 (June 2021), 26 pages. doi:10.1145/3463527

[35] Chutian Jiang, Yinan Fan, Junan Xie, Emily Kuang, Kaihao Zhang, and Mingming Fan. 2024. Designing Unobtrusive Modulated Electrotactile Feedback on Fingertip Edge to Assist Blind and Low Vision (BLV) People in Comprehending Charts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. New York, NY, USA, Article 425, 20 pages. doi:10.1145/3613904.3642546

[36] Chayapol Kamyod and Parichart Hongsing. 2023. The Proposed Wind-gate Monitoring System for Teaching and Learning Thai Traditional Massage. In *2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT NCON)*. 353–356. doi:10.1109/ECTIDAMTNCON57770.2023.10139348

[37] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274. doi:10.1016/j.lindif.2023.102274

[38] Majeed Kazemitabaar, Xinying Hou, Austin Henley, Barbara Jane Ericson, David Weintrop, and Tovi Grossman. 2023. How novices use LLM-based code generators to solve CS1 coding tasks in a self-paced learning environment. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*. 1–12. https://doi.org/10.1145/3631802.3631806

[39] Emily Kuang, Minghao Li, Mingming Fan, and Kristen Shinohara. 2024. Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive Dialogue and Timing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16. https://dl.acm.org/doi/full/10.1145/3613904.3642168

[40] Ehsan Latif, Ramviyas Parasuraman, and Xiaoming Zhai. 2024. PhysicsAssistant: An LLM-Powered Interactive Learning Robot for Physics Lab Investigations. arXiv:2403.18721 [cs.RO]

[41] Hang Li, Tianlong Xu, Chaoli Zhang, Eason Chen, Jing Liang, Xing Fan, Haoyang Li, Jiliang Tang, and Qingsong Wen. 2024. Bringing Generative AI to Adaptive Learning in Education. arXiv:2402.14601 [cs.CY]

[42] Yihao Li, Junyu Long, Yun Chen, Yan Huang, and Ni Zhao. 2022. Crosstalk-Free, High-Resolution Pressure Sensor Arrays Enabled by High-Throughput Laser Manufacturing. *Advanced Materials* 34, 21 (2022), 2200517. doi:10.1002/adma.202200517

[43] Scott K Liddell. 2021. *American sign language syntax.* Vol. 52. Walter de Gruyter GmbH & Co KG.

[44] Scott K Liddell and Robert E Johnson. 1989. American sign language: The phonological base. *Sign language studies* 64, 1 (1989), 195–277.

[45] Valentin Liévin, Christoffer Egeberg Hother, Andreas Geert Motzfeldt, and Ole Winther. 2024. Can large language models reason about medical questions? *Patterns* 5, 3 (2024). https://doi.org/10.1016/j.patter.2024.100943

[46] Weikang Lin, Biao Wang, Guoxiang Peng, Yao Shan, Hong Hu, and Zhengbao Yang. 2021. Skin-inspired piezoelectric tactile sensor array with crosstalk-free row+ column electrodes for spatiotemporally distinguishing diverse stimuli. *Advanced Science* 8, 3 (2021), 2002817. https://doi.org/10.1002/advs.202002817

[47] Mengqi Liu and Faten M'Hiri. 2024. Beyond Traditional Teaching: Large Language Models as Simulated Teaching Assistants in Computer Science. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024)*. 743–749. https://doi.org/10.1145/3626252.3630789

[48] Tiedong Liu and Bryan Kian Hsiang Low. 2023. Goat: Fine-tuned LLaMA Outperforms GPT-4 on Arithmetic Tasks. arXiv:2305.14201 [cs.LG]

[49] Haipeng Luo, Qingfeng Sun, Can Xu, Pu Zhao, Jianguang Lou, Chongyang Tao, Xiubo Geng, Qingwei Lin, Shifeng Chen, and Dongmei Zhang. 2023. WizardMath: Empowering Mathematical Reasoning for Large Language Models via Reinforced Evol-Instruct. arXiv:2308.09583 [cs.CL]

[50] Ziyang Luo, Can Xu, Pu Zhao, Qingfeng Sun, Xiubo Geng, Wenxiang Hu, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. 2023. WizardCoder: Empowering Code Large Language Models with Evol-Instruct. arXiv:2306.08568 [cs.CL]

[51] Tian Ma and Ming Guo. 2019. Research on Kinect-based Gesture Recognition. In *2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. 1–5. doi:10.1109/ICSPCC46631.2019.8960806

[52] Pratyusha Maiti and Ashok K. Goel. 2024. How Do Students Interact with an LLM-powered Virtual Teaching Assistant in Different Educational Settings? arXiv:2407.17429 [cs.CY]

[53] Atsushi Mizumoto and Masaki Eguchi. 2023. Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics* 2, 2 (2023), 100050. doi:10.1016/j.rmal.2023.100050

[54] Sritoomma Netchanok, Moyle Wendy, Cooke Marie, and O'Dwyer Siobhan. 2012. The effectiveness of Swedish massage and traditional Thai massage in treating chronic low back pain: A review of the literature. *Complementary Therapies in Clinical Practice* 18, 4 (2012), 227–234. doi:10.1016/j.ctcp.2012.07.001

[55] OpenAI. 2023. Models - OpenAI API. https://platform.openai.com/docs/models. [Last accessed: 27/01/2025].

[56] Margaret M Plack and MaryAnne Driscoll. 2011. *Teaching and learning in physical therapy: From classroom to clinic.* Slack Incorporated.

[57] Cristina Ramírez-Fernández, Victoria Meza-Kubo, Eloísa García-Canseco, Alberto L Morán, Oliver Pabloff, David Bonilla, and Nirvana Green. 2017. Massage Therapy of the Back Using a Real-Time Haptic-Enhanced Telerehabilitation System. *Mobile Information Systems* 2017, 1 (2017), 5253613. https://doi.org/10.1155/2017/5253613

[58] Baptiste Rozière, Jonas Gehring, Fabian Gloeckle, Sten Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi Adi, Jingyu Liu, Romain Sauvestre, Tal Remez, Jérémy Rapin, Artyom Kozhevnikov, Ivan Evtimov, Joanna Bitton, Manish Bhatt, Cristian Canton Ferrer, Aaron Grattafiori, Wenhan Xiong, Alexandre Défossez, Jade Copet, Faisal Azhar, Hugo Touvron, Louis Martin, Nicolas Usunier, Thomas Scialom, and Gabriel Synnaeve. 2024. Code Llama: Open Foundation Models for Code. arXiv:2308.12950 [cs.CL]

[59] Mose Sakashita, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew David Wilson. 2024. SharedNeRF: Leveraging Photorealistic and View-dependent Rendering for Real-time and Remote Collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3613904.3642945

[60] Vincenzo Scotti and Mark James Carman. 2024. LLM Support for Real-Time Technical Assistance. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 388–393. https://doi.org/10.1007/978-3-031-70371-3_26

[61] Nigam H Shah, David Entwistle, and Michael A Pfeffer. 2023. Creation and adoption of large language models in medicine. *Jama* 330, 9 (2023), 866–869. https://doi.org/10.1001/jama.2023.14217

[62] Sudeep Sharma, Ashok Chhetry, Md Sharifuzzaman, Hyosang Yoon, and Jae Yeong Park. 2020. Wearable capacitive pressure sensor based on MXene composite nanofibrous scaffolds for reliable human physiological signal acquisition. *ACS applied materials & interfaces* 12, 19 (2020), 22212–22224. https://pubs.acs.org/doi/10.1021/acsami.0c05819 [Last accessed: 27/01/2025].

[63] Yichun Shen, Shuyi Wang, Yuhan Shen, and Hua Xing. 2024. Integrating Traditional Chinese Medicine massage therapy with machine learning: A new trend in future healthcare. (2024). https://doi.org/10.61189/721472czacxf

[64] Aldona Sobota and Piotr Zarzycki. 2013. Effect of Pasta Cooking Time on the Content and Fractional Composition of Dietary Fiber. *Journal of Food Quality* 36, 2 (2013), 127–132. doi:10.1111/jfq.12023

[65] Aarohi Srivastava and David Chiang. 2024. We're Calling an Intervention: Exploring the Fundamental Hurdles in Adapting Language Models to Nonstandard Text. arXiv:2404.07304 [cs.CL]

[66] Supanita Sudsaward and Chalabhorn Suwansumrit. 2018. Interactive Multimedia Electronic Learning (IMEL) Packages on Health Thai Massage Course. In *Proceedings of the 2nd International Conference on Education and Multimedia Technology*. 64–68. https://doi.org/10.1145/3206129.3239424

[67] Xia Sun, Shaoshuai He, Zhihui Qin, Junjie Li, and Fanglian Yao. 2021. Fast self-healing zwitterion nanocomposite hydrogel for underwater sensing. *Composites Communications* 26 (2021), 100784. doi:10.1016/j.coco.2021.100784

[68] Edi Supriyadi and KS Kuncoro. 2023. Exploring the future of mathematics teaching: Insight with ChatGPT. *Union: Jurnal Ilmiah Pendidikan Matematika* 11, 2 (2023), 305–316. https://doi.org/10.30738/union.v11i2.14898

[69] Takashige Suzuki, Alarith Uhde, Takuto Nakamura, Takuji Narumi, Tomohiro Amemiya, and Hideaki Kuzuoka. 2024. Be sensei, my friend: Aikido training with a remotely controlled proxy trainer. *Frontiers in Virtual Reality* 5 (2024). doi:10.3389/frvir.2024.1392635

[70] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. *Nature medicine* 29, 8 (2023), 1930–1940. https://doi.org/10.1038/s41591-023-02448-8

[71] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 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* (New Orleans, LA, USA) *(UIST '19)*. New York, NY, USA, 161–174. doi:10.1145/3332165.3347872

[72] Rubén Gil Vera, José Luis Lázaro Galilea, Alfredo Gardel Vicente, Álvaro De-La-Llana-Calvo, and Ignacio Bravo Muñoz. 2024. A Flexible Framework for the Deployment of STEM Real Remote Laboratories in Digital Electronics and Control Systems. *IEEE Access* 12 (2024), 14563–14579. doi:10.1109/ACCESS.2024.3357991

[73] Yousef Wardat, Mohammad A Tashtoush, Rommel AlAli, and Adeeb M Jarrah. 2023. ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education* 19, 7 (2023), em2286. https://doi.org/10.29333/ejmste/13272

[74] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf [Last accessed: 27/01/2025].

[75] Maryellen Weimer. 2013. *Learner-centered teaching: Five key changes to practice.* John Wiley & Sons.

[76] Karsten Weiß and Heinz Worn. 2005. The working principle of resistive tactile sensor cells. In *IEEE International Conference Mechatronics and Automation, 2005*, Vol. 1. IEEE, 471–476. https://doi.org/10.1109/ICMA.2005.1626593

[77] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144 [cs.CL]

[78] Xiaona Xia and Wanxue Qi. 2024. Driving STEM learning effectiveness: dropout prediction and intervention in MOOCs based on one novel behavioral data analysis approach. *Humanities and Social Sciences Communications* 11, 1 (2024), 1–19. https://doi.org/10.1057/s41599-024-02882-0

[79] Changrong Xiao, Wenxing Ma, Qingping Song, Sean Xin Xu, Kunpeng Zhang, Yufang Wang, and Qi Fu. 2024. Human-AI Collaborative Essay Scoring: A Dual-Process Framework with LLMs. arXiv:2401.06431 [cs.CL]

[80] Yuanyuan Yang and Wei Bao. 2023. Application of Human-Computer Interaction Technology in Remote Language Learning Platform. *International Journal of Human–Computer Interaction* 0, 0 (2023), 1–11. https://doi.org/10.1080/10447318.2023.2194709

[81] Longhui Yu, Weisen Jiang, Han Shi, Jincheng Yu, Zhengying Liu, Yu Zhang, James T. Kwok, Zhenguo Li, Adrian Weller, and Weiyang Liu. 2024. MetaMath: Bootstrap Your Own Mathematical Questions for Large Language Models. arXiv:2309.12284 [cs.CL]

[82] Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. In *International Conference on Machine Learning*. PMLR, 41092–41110. https://proceedings.mlr.press/v202/zhang23m/zhang23m.pdf [Last accessed: 27/01/2025].

[83] Hong Zhang and Eric So. 2002. Hybrid resistive tactile sensing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 32, 1 (2002), 57–65. https://doi.org/10.1109/3477.979960

8 Appendix

8.1 Prompts for VTA

8.1.1 Prompts for the Role of the VTA. You are a helpful massage teaching assistant. You need to help with analyzing students' massage training data. The following is a stroke description used by teachers and students: certainStroke. For each stroke, we have the stroke data representing the accuracy of the action posture and pressure data representing the accuracy of the action force.

For stroke data, we provide data on the finger's bending angle.

For pressure data, you categorize the human hand into the right and left hands, each further subdivided into several specific areas. Each hand has areas designated for the tips and rests of the fingers: the thumb, index, middle, ring, and pinky fingers have their respective tip and rest regions identified. Additionally, the palm is divided into upper, middle, and lower palm areas, which cover different vertical segments of the palm itself. Two specific areas, the hyperthenar and thenar, are noted for their position on the edge of the palm near the pinky and thumb, respectively.

8.1.2 Learner Mode Review Prompts. The *learner mode review prompts* include the **stroke data calculation prompts**, **stroke data response prompts**, **pressure data calculation prompts**, and **Pressure data response prompts**.

Stroke data input prompts The following is a JSON file of the stroke difference between the student and the instructor.

This dictionary describes the angle difference in the degree of curvature of each finger on both hands between student and instructor. The dictionary is divided into two parts: "right hand" and "left hand." Each section contains five keys representing five fingers: thumb, index finger, middle finger, ring finger, and little finger.

The positive values mean that the finger of the student is bent more than the instructor's standard finger, the negative values mean that the finger of the student is not curved enough compared to the instructor's standard finger, and a value of 0 indicates that the bending angle of the fingers is the same.

The content of the JSON file is: strokeDiffInfo

Stroke data response prompts

Now please tell me how far I am from the instructor's standard movements based on the above information, i.e. what do I need to do to improve. Please note: your response should be on the colloquial side, not directly stating a numerical value, but describing the magnitude in more general terms. Your response should match the tone of an instructor teaching students.

Keep your answers as concise as possible.

Pressure data input prompts The following is a JSON file of differences between teachers and students. By comparing hand pressure data from both groups, the data differences are calculated.

A negative value means that the student's force was less than the teacher's standard force and should apply more pressure next time. A positive value means that the student's force was more than the teacher's standard force and should apply less pressure next time. If the dictionary is empty, it means that the student used the same force as the teacher. For example, if "leftHand": , it means that all fingers of the student's left hand perform as well as the teacher's.

The content of the JSON file is: pressureDiffInfo

Pressure data response prompts Now, based on this difference JSON file, tell me how to adjust the pressure of my hand for the next time. Please note: your response should be on the colloquial side, not directly stating a numerical value, but describing an extent. Don't mention anything that doesn't need to be changed.

Note that your description should use intensity adverbs appropriately and describe according to the hand sections I provided. Your response should match the tone of an instructor teaching students. Keep your answers as concise as possible.

8.1.3 Instructor Mode Summary Prompts. You are a teaching assistant specializing in massage, and your task is to report back to the teacher on the student's learning based on the pressure data generated by each student using the same gesture, and to select two to four of the eight gesture data for the teacher to re-demonstrate based on the student's performance, giving reasons for it.

The descriptions of the differences in each student's stress data are stored in the JSON files.