

# Trinity: Synchronizing Verbal, Nonverbal, and Visual Channels to Support Academic Oral Presentation Delivery

Anonymous Author(s)

Submission Id: 23

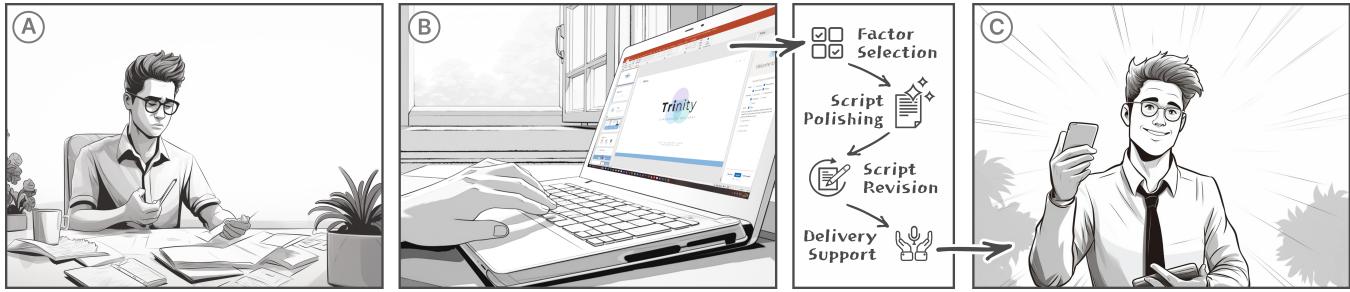


Figure 1: Example scenario of using *Trinity*. A) The user, lacking comprehensive knowledge and sufficient experiences, struggles to find an effective way to deliver his presentation. B) Seeking assistance, the user opens the *Trinity* PowerPoint add-in on their laptop. The user customizes the supportive features based on the specific needs, refines the presentation script, makes necessary revisions, and finally uploads the improved script to receive delivery support. C) The user confidently and smoothly delivers the presentation. *Trinity* App provides features like remote mobile visual control, multilevel speech pace modulation, and integrated delivery prompts, ensuring synchronized multichannel delivery and enhancing the overall effectiveness of the presentation.

## ABSTRACT

Academic Oral Presentation (AOP) allows English-as-Foreign-Language (EFL) students to express ideas, engage in academic discourse, and present research findings. However, while previous efforts focus on training efficiency or speech assistance, EFL students often face the challenge of seamlessly integrating verbal, nonverbal, and visual elements into their presentations to avoid coming across as monotonous and unappealing. Based on a need-finding survey, a design study, and an expert interview, we introduce *Trinity*, a hybrid mobile-centric delivery support system that provides guidance for multichannel delivery on-the-fly. On the desktop side, *Trinity* facilitates script refinement and offers customizable delivery support based on large language models (LLMs). Based on the desktop configuration, *Trinity* App enables a remote mobile visual control, multi-level speech pace modulation, and integrated delivery prompts for synchronized delivery. A controlled between-subject user study suggests that *Trinity* effectively supports AOP delivery and is perceived as significantly more helpful than baselines, without excessive cognitive load.

## CCS CONCEPTS

- Human-centered computing → Interactive systems and tools.

## KEYWORDS

Multichannel Communication, Academic Oral Presentation, Delivery Support

## ACM Reference Format:

Anonymous Author(s). 2018. *Trinity: Synchronizing Verbal, Nonverbal, and Visual Channels to Support Academic Oral Presentation Delivery*. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX')*. ACM, New York, NY, USA, 30 pages. <https://doi.org/XXXXXXXX.XXXXXXXX>

## 1 INTRODUCTION

Academic Oral Presentation (AOP) is frequently applied in universities to showcase students' understanding and mastery of a subject before an audience [43]. Particularly prevalent in subjects within institutions of higher education, AOP holds significant importance as it allows students to effectively communicate their ideas to the audience [87], assess and socialize them in genre-specific academic discourse [26, 58], and present research findings to peers or academic communities [100]. The omnipresence of AOPs in various courses and disciplines provides students with platforms to showcase their knowledge [51], foster self-assurance [47], and cultivate indispensable skills for academic and future professional pursuits [2, 58, 99].

As outlined in Table 1, previous literature has categorized the presentation delivery into three communication channels: **verbal** (e.g., vocal pitch), **nonverbal** (e.g., gesture), and **visual** (e.g., slides) [36, 66].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX', June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/XXXXXXXX.XXXXXXXX>

**Table 1: Communication channels and delivery factors in presentation.**

Channel	Factor	Utility
Verbal	Vocal Pitch	To place emphasis, convey emotions, demonstrate authority, or add vocal variety. [33, 37]
	Speech Rate	To capture attention, emphasize points, adjust to time, or match the content. [37, 38]
	Volume	To attract attention, emphasize points, adjust to environment, or match the content. [38]
Nonverbal	Eye Contact	To establish connection, convey message, project authority and confidence, facilitate engagement. [7, 76]
	Facial Expression	To enhance message, engage audience, show credibility, or communicate attitudes and intentions. [93]
	Composure	To overcome nervousness, handle unexpected situations, and project authority. [3]
Visual	Gesture	To enhance verbal message, engage audience, or show confidence and credibility. [6, 93]
	Posture	To show confidence, enhance voice and breathing, or engage audience and create rapport. [7, 56]
Slides		To organize the content, communicate message clearly and persuasively, or attract attention. [15, 25]

As indicated by Hunyadi et al. [39], “*if communication is to be effective, it is necessary to ensure that all the paraverbal dimensions are congruent.*” Similarly, Hentz [35] also argued that presenters should “*examine how their written, oral, visual, and electronic modes of communication inform one another to generate meaning.*” However, many presenters, especially EFL students, lack comprehensive knowledge and awareness of utilizing diverse communication channels in AOPs. Aside from watching presentation examples, current supportive tools only provide presentation assistance in certain channels (e.g., [89, 101]). They lack the tools to understand and effectively utilize diverse communication channels for delivering AOPs. As a result, EFL students still encounter various obstacles when dealing with diverse communication channels in AOPs to effectively communicate with the audience. **First, Absence of Certain Channels in Multichannel Delivery.** Effective presentations derive their potency from adequately utilizing various communication channels for delivery [18, 24]. However, due to inadequate preparation and unbalanced training, the act of EFL students delivering AOPs often regresses into a monotonous recitation or reading of content directly from slides or scripts, lacking engagement or dynamism [40, 81]. **Second, Inconsistencies and Discrepancies in Channel Integration.** Despite endeavors to exploit multiple communication channels, disparities in the execution of these channels by EFL students emerge [22], resulting in damages to the audience’s perceptions of their attitudes [4] and even undermining the credibility of their presentation content [32]. For instance, an EFL student may highlight a point on a slide without adjusting their tone of voice and maintain an inappropriate hand gesture with crossed hands. **Third, Linguistic Hurdles for EFL Students.** Language proficiency is another factor that adds to the difficulties faced by EFL students during oral presentations [91]. Specifically, having a limited vocabulary and struggling with sentence structure can hinder comprehension and make the presentation confusing, even when the topic is engaging and effectively delivered [81].

Our goal is to offer guidance for synchronizing multiple communication channels during presentations. However, offering such guidance encounters three significant research gaps: **G1. Empirical Gaps in Presentation Delivery Factors.** The existing literature has proposed and examined various factors that impact presentation delivery based on communication theories [12, 23, 31]. In the field of Human-Computer Interaction (HCI), several works aim to offer effective training or learning experience to improve presentation delivery [63, 85, 89]. However, there remains a notable lack of empirical investigation into the factors that EFL students need

the most support for and the factors that most affect the perception of AOPs. **G2. Limitations of Predefined Design.** Existing works often directly provide delivery guidance for a predetermined channel in a pre-defined design, leaving presenters with limited control over what prompts are presented. This lack of adaptability hinders the accommodation of presenters’ diverse needs regarding delivery support. Furthermore, audience’s perceptions and concerns towards these tools were not comprehensively taken into account from the start of system design in previous works. **G3. Unexplored Synchronization Approach.** Previous studies [5, 79, 80] have primarily focused on enhancing presentation delivery from the perspective of certain channels (e.g., verbal delivery). However, how to synchronize diverse communication channels to facilitate presentation delivery effectively and appropriately remains unexplored.

Therefore, our study aims to investigate the following five research questions (**RQ1-RQ5**). First, to investigate EFL students’ actual needs and preferences and the suitable approach of providing such support during AOPs, we have formulated two fundamental research questions: **RQ1: What factors do EFL students need the most support for and What factors most affect the perception of AOPs (G1)?** and **RQ2: What methods would be deemed acceptable by both EFL students and audience to provide on-the-fly delivery support and effectively synchronize diverse communication channels (G2, G3)?** To answer **RQ1**, we designed and launched a survey with 49 EFL students and 36 instructors from different majors in a local university. To address **RQ2**, we conducted a design study after identifying crucial AOP delivery factors and their ranking through a literature review and a survey. In the design study, we initially employed these factors to facilitate an exploratory design brainstorming session. Subsequently, we utilized storyboards to visually represent the design concepts generated from the brainstorming process. To further explore more detailed designs, we organized a design workshop to elaborate on the supportive approach. Additionally, to obtain professional perspectives on the output results and findings, we conducted interviews with five presentation experts. Moreover, during these interviews, we also performed a pilot test on potential design implementations. Considering LLMs have harvested rich knowledge from large-scale diverse corpus (e.g., website, books) and have demonstrated promising capabilities in understanding, generating, and expressing human languages [96], our experts were asked to evaluate the applicability of LLMs as an implementation choice. Incorporating feedback from instructors, EFL students, and experts, we identified six design goals and proposed *Trinity*, a hybrid mobile-centric system. On the PC end, *Trinity* refines scripts

and provides customizable delivery prompts using GPT-4. Based on the PC configuration, *Trinity* App enables remote visual control, speech pace modulation, and integrated delivery prompts for synchronized delivery.

With the proposed research prototype, we further investigate the following research questions: **RQ3: How are the usability and effectiveness of the supportive system?** Due to the lack of evaluation regarding usage patterns and influences on students, we further investigate **RQ4: How will EFL students interact with and be influenced by the system?** Since the system provides delivery support on-the-fly, we are interested in investigating **RQ5: How will students trust and collaborate with our system?** To answer **RQ3-RQ5**, we conducted a controlled between-subject user study with 33 EFL student presenters and 21 audience members, comparing our proposed *Trinity* with two baselines: 1) *IntelliPrompter*, a PC-based system with features like speech tracking and dynamic script display, and 2) *OfficeRemote*, a mobile office subsystem offering slide navigation, script reference, a timer, and a laser pointer. We used mixed-methods analyses, combining in-task and post-task questionnaires, and interviews. The results consistently showed that *Trinity* effectively supports multichannel presentation delivery and is perceived as significantly more useful than the baselines. The development and evaluation of *Trinity* contribute to our understanding of AOP factors. It offers valuable insights into system design considerations, usage patterns, and its overall impact. The successes and challenges associated with the design choices made in *Trinity* provide valuable guidance for the development of future presentation support tools within educational contexts and for enhancing public speaking tools beyond the classroom. In summary, this study makes four key contributions:

- (1) **Surveying EFL Students and Instructors:** We surveyed 49 EFL students to understand their specific AOP delivery needs and 36 instructors to identify factors significantly influencing the perception of AOP delivery.
- (2) **Design Study:** We held a design study with 9 EFL students and 4 instructors, aiming to explore the design space and refine our approach iteratively. The focus was on providing on-the-fly delivery support and synchronizing diverse communication channels.
- (3) **Trinity System:** We developed *Trinity*, a comprehensive delivery support system designed to enhance academic oral presentation delivery based on a hybrid infrastructure.
- (4) **User Study:** We conducted a controlled between-subject user study involving 33 EFL students and 21 audience to systematically investigate the usability, effectiveness, interaction, influence, trust, and collaboration of *Trinity*, as well as trade-offs in design choices.

## 2 RELATED WORK

### 2.1 Context and Challenges in AOPs for EFL Students

AOP, a well-established educational strategy, employs various platforms, from traditional mediums like chalkboards to modern tools like presentation software [43, 64]. In the implementation of AOPs, slideware tools such as Microsoft PowerPoint, Google Slides, and

Apple Keynote have become predominant. These tools allow students to create and display slides, offering real-time aids like markers, laser pointers, and timers [57]. However, challenges persist, especially regarding the suitability of slide-based AOPs for EFL students.

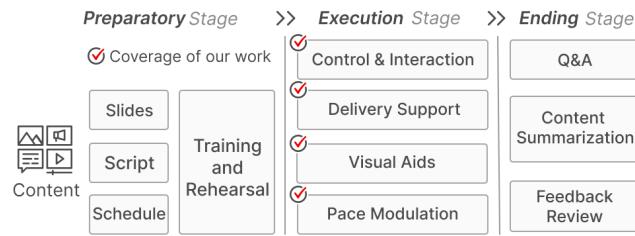
The primary concern of EFL students is the lack of sufficient preparation [81]. Despite their substantial effort, many EFL students are plagued by apprehensions about memory lapses and organizing ideas effectively, as thorough preparation is essential for effective oral presentations. Whai and Mei [91] found that inadequate preparation is the main challenge in oral presentations, which may be due to students prioritizing core subjects such as engineering and commerce. As a result, they tend to assign less importance to oral presentation tasks because they perceive them as irrelevant to their field of study [91]. Consequently, presentation preparation is compromised, negatively impacting delivery. Students often rely on notes or scripts for coherence [70]. However, this dependence leads EFL students to simply read from slides or scripts instead of engaging with the audience and maintaining eye contact [29, 40, 81]. These challenges arise from EFL students' apprehension and vulnerability when speaking in front of a large audience [81]. This study focuses on the issue of suboptimal presentation delivery among EFL students, where reading from slides or scripts occurs due to inadequate preparation and lack of effective learning approach. To address this, we propose an auxiliary method to enhance the AOP delivery of EFL students.

### 2.2 Requisites for Effective Presentation Delivery

Flawless delivery of a presentation is crucial for its success [20]. Both verbal and nonverbal components play a vital role in the effectiveness of a presentation [78]. Skillful speakers seamlessly blend spoken content with nonverbal cues, including body language, gestures, and facial expressions. For example, Dolan [24] emphasized the centrality of nonverbal communication within presentations and provided insights into effective presentation skills. In addition to verbal and nonverbal dimensions, visual aid constitute an additional pivotal avenue for delivery. Peeters et al. [66] delved into performance evaluation within an AOP course, culminating in the development of a standarized rubric that integrates visual aids into the evaluation schema for students' presentations. Furthermore, Bourne [10] developed a set of ten principles for proficient oral presentations, advocating for judicious yet impactful utilization of visual elements. Beyond the individual components, literature unveiled that inconsistencies among verbal, nonverbal, and visual communication channels can significantly affect audience perceptions concerning presenters' attitudes [4], emotions [30], and credibility [32]. These explorations highlight the crucial importance of aligning verbal, nonverbal, and visual elements throughout a presentation.

Nevertheless, most prior supportive systems have focused mainly on specific elements of presentation delivery, such as volume and speech tempo. This narrow focus has overlooked the overall cohesion and synchronicity among diverse communication channels. Therefore, in this study, we aim to coordinate verbal, nonverbal, and visual channels to enhance presentation delivery effectiveness.

### 349 2.3 Supportive Tools for Enhanced Presentation 350 Delivery



360 **Figure 2: Taxonomy of existing works in presentation do-**  
361 **main and coverage of our work.**

363 Numerous systems and tools have been developed to improve  
364 presentation performance across various stages, including prepara-  
365 tory, execution, and ending stages. We detail the taxonomy of exist-  
366 ing works and our scope in Figure 2. Careful planning, efficient time  
367 management, effective communication skills, and smart technology  
368 use, crucial for successful presentations [20], are primarily influ-  
369 enced by the preparation and execution stages. Hence, we discuss  
370 the related work for these stages.

371 During the preparation stage, one line of research has focused  
372 on supporting content formulation [13, 19, 27, 28, 42, 45, 68, 73, 88],  
373 such as slide generation and script creation. For instance, Khataei  
374 et al. [45] and Jokela et al. [42] introduced tools for sharing personal-  
375 ized narratives and creating mobile audio-visual presentations,  
376 respectively. Another line of work focuses on presentation skills  
377 training and rehearsal [49, 63, 85, 89, 97, 101]. However, EFL stu-  
378 dents often struggle with dedicating adequate time for thorough  
379 rehearsal, prompting our exploration of on-the-fly support during  
380 execution.

381 In the execution stage, existing efforts offer on-the-fly support  
382 in control and interaction, speech assistance, visual augmentation,  
383 and pace modulation. Liu et al. [52] proposed *IGScript*, an interac-  
384 tion grammar framework for scientific data presentations, while  
385 Cao et al. [14] and others explored gesture-based presentation con-  
386 trol [1, 69]. Some studies focused on enhancing visual aids, such  
387 as Kim et al.’s *HoloBox* [46], which spatially integrated presenters,  
388 and Thanyadit et al.’s mixed reality avatar presentation tools [82].  
389 Trinh et al. [84] introduced *DynamicDuo*, using virtual agents to  
390 assist inexperienced speakers. However, these works either require  
391 specialized, expensive media or use inflexible visual augmentations  
392 lacking adaptivity.

393 Our study is more similar to works in the delivery support do-  
394 main, which focus on on-the-fly assistant for improved presenta-  
395 tion delivery. Tanveer et al. [79] introduced *Rhema* that provides real-  
396 time speech pace and volume cues. Microsoft’s *Office Remote*<sup>1</sup>  
397 facilitates mobile slide navigation and script referencing for Pow-  
398 erPoint. Asadi et al. [5] proposed *IntelliPrompter*, a system that  
399 tracks presenter’s slide content coverage and dynamically adapts  
400 the script display to highlight the next relevant topic, improving  
401 presentation coherence.

402 While existing efforts provide foundational contributions, they  
403 focus on a limited set of delivery factors within predefined channels,

404 <sup>1</sup><https://support.microsoft.com/en-us/topic/office-remote-for-pc-7e3d9342-61c7-4bc2-8bc0-fa47542d1bce>

407 such as loudness and speech rate [79], without exploring the inte-  
408 gration and synchronization of all three communication channels.  
409 Our work, shown in Figure 2, covers key aspects of the execu-  
410 tion stage and synchronizes three communication channels. We  
411 explored using LLMs to optimize linguistic expression, followed a  
412 user-centered design approach considering individual preferences  
413 and audience perceptions, and conducted a comprehensive user  
414 study to evaluate the system’s usability, effectiveness, usage pat-  
415 terns, and influence.

## 3 FORMATIVE STUDY

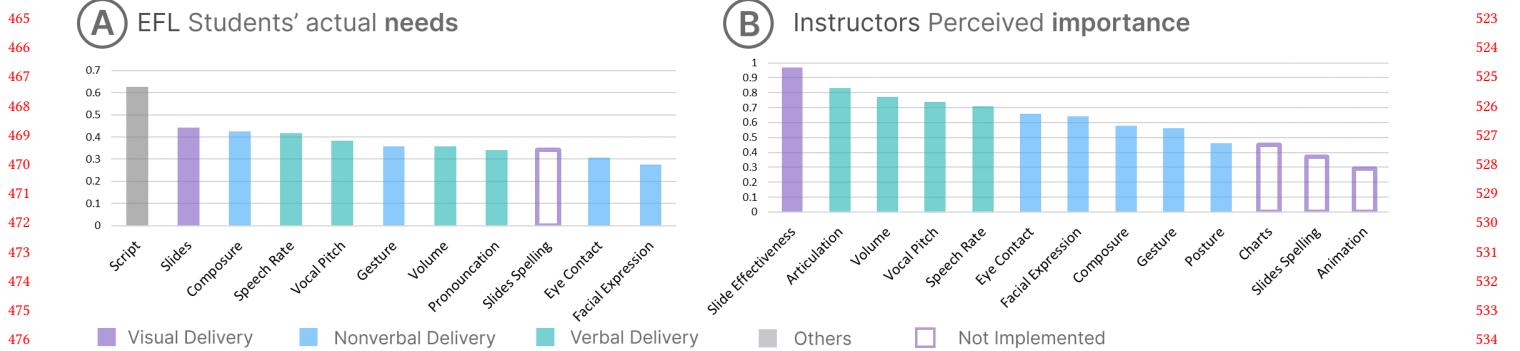
418 Previous research have shown the benefits of on-the-fly support  
419 for presentation delivery [5, 79]. However, existing designs face  
420 issues like limited audience engagement and fixed delivery factors.  
421 The formative study addresses: **RQ1**, exploring EFL student needs  
422 and audience perceptions through surveys, and **RQ2**, investigating  
423 student preferences and audience concerns regarding on-the-fly  
424 delivery support and channel synchronization via a design study.  
425 We included instructors as audience representatives considering  
426 their extensive experience in assessing presentations.

### 3.1 Exploratory Survey

428 **3.1.1 Method.** To explore **RQ1**, we first designed a survey to under-  
429 stand the AOP support needs of EFL students. From EFL students’  
430 perspective, we used a 3-point scale (*In demand*, *Somewhat need*,  
431 *No need*) to assess the level of support required for each AOP deliv-  
432 ery factor, considering the previously mentioned communication  
433 channels and delivery factors (Table 1). The survey also collected  
434 demographic details, familiarity with AOPs, perceived AOP delivery  
435 challenges, and usage of supportive tools. Additionally, participants  
436 were invited to offer comments or suggestions on a delivery support  
437 system design.

438 Likewise, from the audience’s perspective, we designed a survey  
439 for instructors to explore significant factors influencing presen-  
440 tation delivery perception, covering demographics, experience in  
441 overseeing AOPs, and perceptions of supportive systems during  
442 presentations. We assessed instructors’ perceived importance of  
443 each factor using a 3-point scale (*Very important*, *Somewhat im-  
444 portant*, and *Not important*) and solicited their suggestions for a  
445 real-time AOP delivery support system. After securing Institutional  
446 Review Board (IRB) approval, the surveys were distributed at a local  
447 university using snowball sampling via email and social media.

449 **3.1.2 Results.** Here we present ratings gathered from both stu-  
450 dents and instructors for each factor across the three channels,  
451 reflecting their perceived levels of need and importance. Full and  
452 detailed survey results can be found in Appendix A. Following [55],  
453 we employed normalized weighted importance scores to rank the  
454 supportive and perceptual factors. The resulting rankings are de-  
455 picted in Figure 3, showcasing the final score ranging from 0 to 1  
456 in descending order. It’s noteworthy that individual preferences  
457 vary due to differences in abilities and perceptions (**F1**). Even for  
458 the top-ranking factor, some participants might still consider them  
459 unnecessary or unimportant. As a result, this study acknowledged  
460 the existence of diverse individual perspectives. Notably, we imple-  
461 mented the majority of the highly ranked factors, as indicated by  
462 the colored columns in Figure 3.



**Figure 3: EFL student's needs and perceived importance from instructors based on a normalized weighted importance score (from 0 to 1, the closer to 1, the more important). The information without color filling represents the factor that is not implemented in the system. A) The ranking of students' actual needs in factors influencing AOP delivery. B) The ranking of instructors' perceived importance levels of delivery factors.**

### 3.2 Design Study

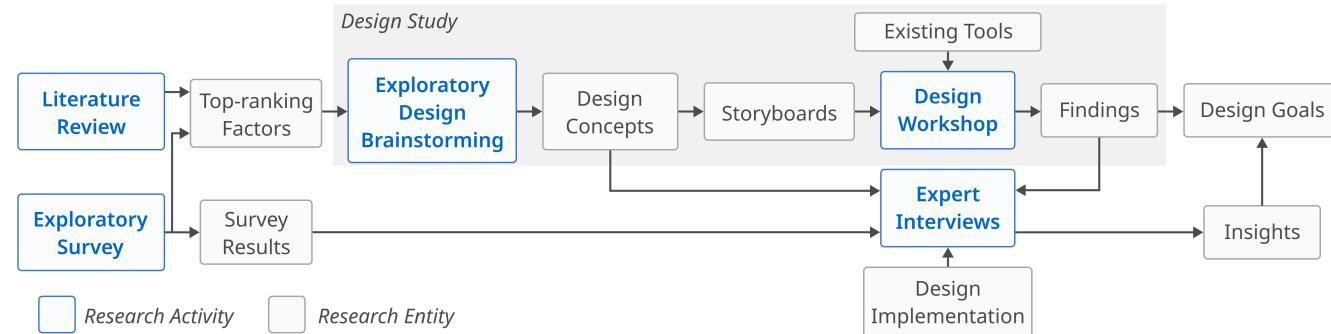
To address RQ2, we conducted a design study to explore methods for synchronizing communication channels and providing on-the-fly support. With this purpose in mind, we elaborate on the design development process in this subsection following our schema in Figure 4. According to the schema, we initially identified crucial factors and their ranking through a literature review and an exploratory survey. These factors served as the foundation for coordinating the exploratory design brainstorming. The brainstorming session produced design concepts as its output. Subsequently, we employed storyboards to visualize these design concepts. To explore more fine-grained designs, we organized a design workshop to elaborate on the supportive approach. In this workshop, both storyboards and designs of existing tools were utilized. Notably, both the design brainstorming and the design workshop were conducted through Zoom.

**3.2.1 Participants and Power Dynamics.** We recruited a group of 9 EFL students, all with a background in interaction design (S1-9; 6 male, 3 female, average age 22), by advertising within a local university's design course. Complementing this, we also recruited 4 instructors (I1-I4; 3 male, 1 female, average age 38.6) to serve as representatives of the audience. Notably, I1 and I2, among the instructors, were actively involved in teaching design courses at the local university's school of creativity and art. In our design study, our primary goal was to establish a balanced power dynamic between EFL students and instructors. We aimed to empower students to express their ideas and creativity while benefiting from the guidance of instructors with valuable experience and expertise. To achieve this equilibrium, we implemented several strategies. First, we encouraged both students and instructors to engage in open communication by asking questions, seeking clarification, and sharing their ideas. The initial step laid the groundwork for a collaborative environment. Building on this, we promoted peer feedback and assigned the role of facilitators to instructors, fostering a cooperative atmosphere throughout the design study. To address potential power imbalances, we considered Lukes's three dimensions of power [54] and implemented intentional interventions:

1) *Observable conflicts of interests*: Recognizing that conflicts might arise, particularly between students' focus on user experience and

functionality and instructors' concerns about audience perception, we employed a role reversal approach [41]. This involved asking students and instructors to switch roles (presenter and audience) until a consensus was reached. 2) *Hidden omission of possible alternatives*: This is about choices and agendas are narrowed in favour of more powerful parties' interests. In our scenario, acknowledging that certain design concepts might overshadow others, we diligently documented all proposed concepts. During the design study, we deliberately brought up any unmentioned concepts for discussion to ensure a comprehensive exploration of ideas. 3) *Deceptive and indiscernible characteristics of power*: This is about a more powerful person's interests may be internalised and sought after as one's own. In our scenario, recognizing the risk that students might overly align with instructors' comments and concepts, we closely monitored the design development process. If the design appeared to predominantly converge in a certain direction, we reminded both students and instructors to maintain a diverse range of perspectives. Given that two instructors possessed extensive design experience, we specifically tasked them with promptly correcting the direction if any deviations were identified.

**3.2.2 Exploratory Design Brainstorming.** The goal of our brainstorming session is to gather a substantial number of design concepts addressing the challenges associated with multichannel AOP delivery for EFL students. To achieve this, we adhered to three fundamental principles for group brainstorming [92]: 1) Aim for sheer quantity. Our focus was on generating a large number of concepts without any limitations. Participants were discouraged from taking personal notes during the brainstorming to ensure their full engagement in idea generation. Additionally, we requested that they refrain from checking emails or engaging in social media, minimizing distractions and fostering a more concentrated brainstorming environment. 2) Defer judgment about the quality of concepts. Participant were explicitly instructed not to criticize others' concepts either implicitly (through facial expressions or other nonverbal cues) or explicitly during the brainstorming session. 3) Encourage new and wild ideas. Participants were encouraged to propose concepts without considering feasibility, allowing for the inclusion of imaginative and unconventional ideas, even those reminiscent



**Figure 4: The design development process in the Formative Study, where the blue boxes are different research activities while the gray boxes are the research entities.**

of science fiction or movies. During the brainstorming, we initiated the discussion by exploring common presentation delivery challenges for EFL students, drawing insights from survey results. Subsequently, participants were prompted to suggest design concepts for the top-ranking factors identified in the survey, with an explicit directive to withhold criticism and avoid restricting the type or number of concepts. Active documentation of all emerging design concepts occurred throughout the session. Following the brainstorming, both authors and participants collaboratively reviewed all proposed concepts, refining them into a subset deemed potentially applicable to the identified problem.

**3.2.3 Design Concepts and Storyboards.** After obtaining the coarse-grained design concepts, we proceeded to enhance and refine these ideas through a design workshop. To facilitate this exploration, we employed the use of storyboards, providing a visual representation to enhance the clarity and depth of the design concepts. Storyboards, chosen for their ability to visualize design concepts quickly [86], were iteratively refined from broad design concepts to encapsulate different scale ideas. Our design workshop included 10 storyboards (exemplified in Figure 5B), each illustrating an AOP delivery problem, a potential solution, and hypothetical outcomes, and a summarized version is available in Figure 20 in Appendix.

**3.2.4 Design Workshop.** During the design workshop, we initiated the session by reviewing the design concepts derived from the earlier brainstorming session. Following [17], to provide context, we showcased previous designs of delivery support tools, as depicted in Figure 6. This approach aimed to offer participants a concrete understanding of how a delivery support system could be structured and function. The intention was to facilitate participants' comprehension of subsequent design concepts and, in turn, encourage their active contribution to design ideas during the ensuing sessions.

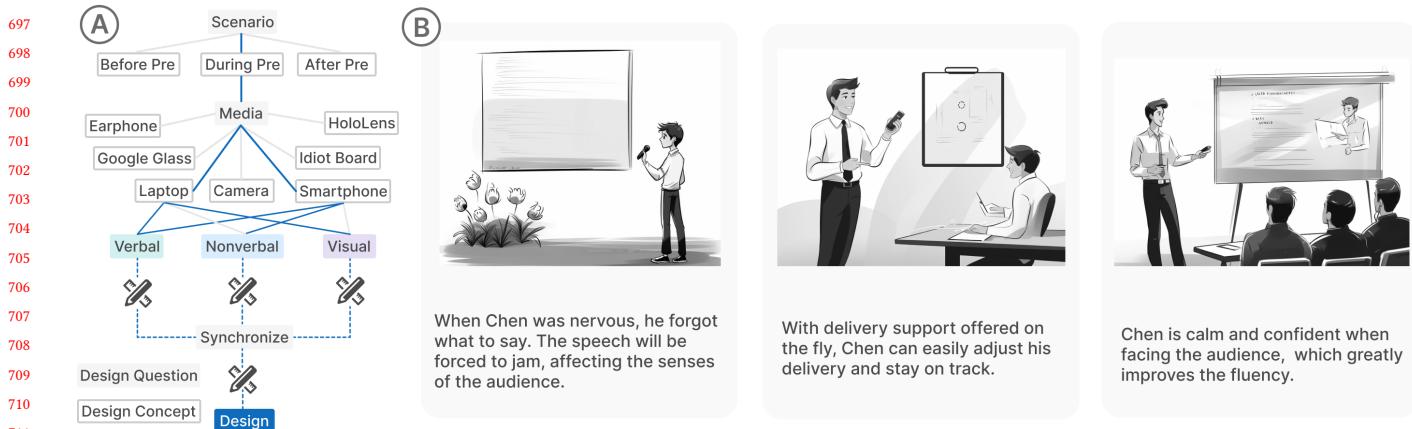
During the workshop, as illustrated in Figure 5A, one of the authors employed screen sharing to present the storyboards in a randomized sequence. Each storyboard was accompanied by an introduction from the author, providing details about both the scenario and the proposed solution. This presentation format allowed participants ample time to read and comprehend the information presented. Following the introduction of each storyboard, participants were encouraged to assess the respective advantages and disadvantages of the proposed solutions. Their insights were solicited, specifically seeking feedback on how each solution could be

effectively applied to the given scenario. Building on these insights, participants were then tasked with the challenge of sketching out a comprehensive design that addressed all relevant delivery factors. Emphasis was placed on encouraging participants to articulate how their designs intricately integrated multiple factors. Additionally, participants were prompted to iteratively refine their designs based on feedback received from their peers throughout the workshop. Throughout this collaborative process, the authors actively documented design insights that emerged from the participants' contributions, providing valuable, unintended insights into the design process.

The speed-dating interviews, averaging around 62.4 minutes, were conducted following IRB guidelines. Participants were compensated with \$20 USD. The interviews were transcribed and analyzed using an Affinity Diagram approach, with the authors collaboratively organizing quotes into themes, refining and structuring the data through iterative discussions.

**3.2.5 Findings.** Regarding the media aspect, participants favored using smartphones and laptops to improve their presentation delivery. For the approach aspect, they preferred inline prompts with scripts for delivery support. The study also highlighted participants' concerns about potential cognitive load challenges and managing different presentation genres.

**F2: Employing on-the-fly delivery support facilitates multichannel communication but requires careful management of cognitive load.** All participants agreed that on-the-fly delivery support would enhance their AOPs performance. S6 noted, “*Some presentation topics can completely throw me off track... having access to such delivery support would truly have been a game-changer.*” Participants also saw this support as a way to reduce performance anxiety, as S2 explained, “*At times, you might possess a deep understanding of the subject matter, but the pressure of the moment can lead to forgetting key details or losing track of progression. Having support to rely on would mitigate that pressure and elevate the overall presentation experience.*” This aligns with research indicating immediate feedback improves presenter awareness and behavior adjustment [13, 21, 44, 65, 67, 75, 79]. However, I1 and I2 cautioned, “*while delivery support provides valuable guidance, excessive cognitive burden may in turn affect the intended delivery. Therefore, on-the-fly patterns always require a careful management of cognitive load.*”



**Figure 5:** Flow diagram of design formation and Storyboard No.2. A) The design iteration started from storyboarding for scenario and media, then spanned to sketch out designs for verbal, nonverbal, and visual delivery support, ending up with a design for synchronizing diverse communication channels. B) Storyboard No.2: Offer delivery support on the fly. participants thought this would enhance their performance with reduced preparation effort.



**Figure 6:** Designs of previous delivery support tools. A) *Office Remote*, an Office subsystem that turns phone into a smart remote that interacts with Microsoft Office on PC. Picture adopted from [74]. B) *Rhema*, an intelligent user interface for Google Glass to help people with public speaking. Picture adopted from [79]. C) *IntelliPromter*, a speech-based note display system that automatically tracks a presenter's coverage of each slide's content. Picture adopted from [5].

**F3: Smartphones and laptops emerge as the more feasible and fitting choices for delivery media.** S2, S4, and I2 found AR/VR devices like HoloLens unsuitable for delivery support due to weight and audience perceptions. Tripod-mounted cameras and idiot boards faced disapproval due to site constraints, but some saw potential in adapting their features. Smartphones and laptops were favored for AOP delivery for their accessibility and capabilities. S3 called it a "*low learning cost*" method, and I1 found it "*visually unintrusive*." S5 envisioned integrating teleprompter features into smartphones, comparing it to "*holding cue cards like hosts do*."

**F4: In-line delivery prompts may enhance my perception of support.** In the workshop, students favored designs with in-line delivery prompts over separate sections, citing prompt recognition as the main reason. S1 noted that "*integrated prompts help me scan the script seamlessly, preventing prompt omission*" (S1). Interest was also expressed in how in-line prompts could aid script refinement. I2 suggested that "*in-line prompts and script refinement suggestions could encourage presenters to improve their content, thereby enhancing presentation quality*".

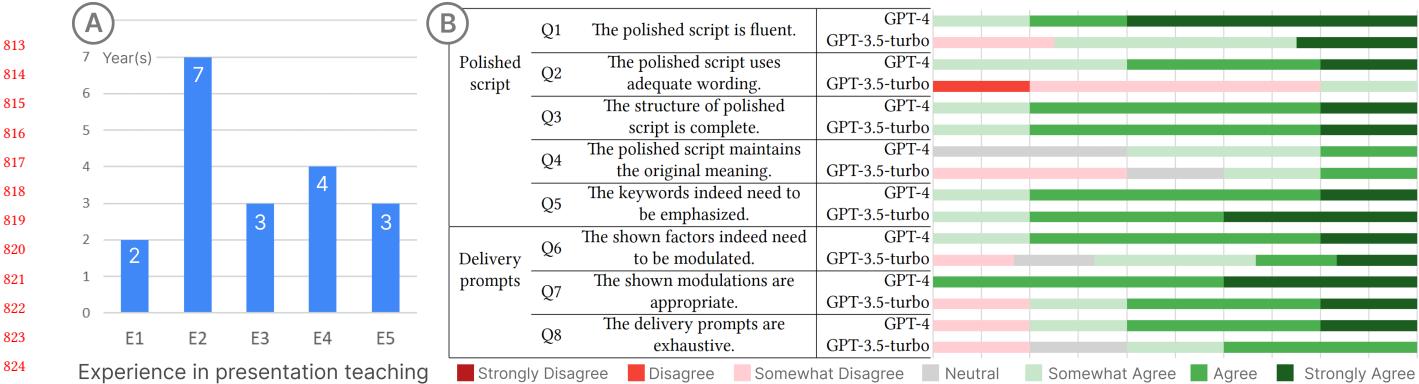
**F5: Polished scripts may have a double-edged sword effect.** Instructors (I1, I3, I4) highlighted concerns about students' AOPs, citing excessive use of slang and improper language. They acknowledged the benefits of script refinement for presentations, such as improved fluency and syntax. However, I1 pointed out a potential issue: "*We have certain speech styles of our own; however, a refined script may 'standardize' our speech style, making us feel awkward*".

*to use them.*" Additionally, students like S9 (male, age: 22) worried about the challenges of understanding and pronouncing unfamiliar words in polished scripts, saying, "*While the idea of having a polished script is appealing, what if it includes challenging or unfamiliar words?*"

**F6: Enhanced speech progress monitoring and hints could assist me in maintaining better pacing.** Participants (S2, S4, S5, S7, S8, I3, and I4) frequently recognized time management as a common issue in AOPs, especially for inexperienced presenters. The idea of features to track delivery progress and provide real-time timing hints was widely supported for its practicality. S7 highlighted the potential benefits, stating, "*Such hints would encourage me to focus on the overall pace of delivery, which is crucial when adhering to strict time limits.*"

**F7: Intuitive prompts may enhance my comprehension speed.** All the participants unanimously agreed on the importance of quick response for fluent AOP and emphasized reducing cognitive recognition demands in design. S2, S8, and I2 preferred *Rhema*'s design approach, using large font sizes and distinct color blocks for intuitiveness. S8 shared, "*Initially, my design for verbal factors resembled a fountain-like graph. However, feedback from others highlighted its lack of user-friendliness. Subsequently, I developed a sketch incorporating icons to signify delivery prompts, which proved to be a much more effective approach.*"

**F8: Enhanced control over visuals on smartphone may facilitate my delivery.** S3, S4, and I2 were concerned about the



**Figure 7: A) Experience of experts in presentation teaching. B) Questions (Q1 - Q8) and experts' feedback on the quality of output script and delivery prompts for GPT-4 and GPT-3.5-turbo.**

lack of dynamism in AOP delivery and the challenge of maintaining engagement. They recognized the value of nonverbal elements like gestures and movement but faced difficulties synchronizing them with the presentation due to the need to stay near the laptop. S4 proposed a solution where "the tool can automatically add pictures, charts, and animations into the slides while adjusting them based on your speech," which was criticized for the potential loss of control. S5 countered with the idea of using smartphones for visual control, acting as "a handheld teleprompter providing both verbal and non-verbal delivery prompts... allowing free movement, gesture, and audience interaction. A simple click on the phone would seamlessly control the visuals without concerns regarding smartphone weight or size."

### 3.3 Expert Interview

We conducted semi-structured interviews with five experts in presentation instruction (E1-5), and their teaching experiences are detailed in Figure 7A. A comprehensive version of the expert interview is available in Appendix B. Our expert interviews had two primary objectives. First, we aimed to gather their professional perspectives on the outcomes of our initial exploratory surveys and design study, thereby enhancing our understanding of the potential applications and implications of the technology. Second, we sought to conduct a pilot test on potential design implementations. since LLMs, like ChatGPT, store diverse and vast knowledge about different areas and are capable of understanding, generating, and expressing human languages flexibly, we were particularly interested in exploring the feasibility of using ChatGPT for refining presentation scripts and providing delivery prompts. We evaluated two LLMs, GPT-3.5-turbo and GPT-4 (both March 14 versions), using 10 EFL student presentation scripts collected from social networks. We adopted a 7-point Likert scale to assess the quality of the LLMs' output scripts and delivery prompts, and open-ended questions to capture insights from the experts' experiences and perceptions about our survey and design outcomes (Figure 7B). We present primary insights obtained from the interviews.

**Insight1: GPT-4 outperformed GPT-3.5-turbo on all metrics, suggesting potential for integration into practice.** GPT-3.5-turbo's complex language and sentence structure were criticized. Both LLMs were favorably rated for fluency and meaning preservation, with GPT-4 slightly leading. The LLMs were praised for

enhancing script clarity, coherence, and organization. The in-line delivery prompts by the GPT-4 were also positively rated. Experts generally preferred GPT-4 for factor judgment and modulation during presentations.

**Insight2: Incorporating visual cues like arrows or highlights to guide students' focus.** Experts also proposed design suggestions for an LLM-based tool to aid EFL students in real-time presentations, such as integrating multimodal feedback. E1 suggested using auditory cues like beeps or pitches to adjust speech volume or speed, and somatosensory cues like vibrations for emphasizing points. E3 proposed incorporating visual cues like arrows or highlights to guide students' focus on relevant slides or audience members.

**Insight3: Pronunciation of unfamiliar or polysyllabic words can pose challenges for EFL students.** E4 mentioned that EFL students often struggle more with verbal communication due to limited opportunities to enhance these skills in academic or professional contexts. E5 added that pronunciation of unfamiliar or polysyllabic words can pose significant challenges for EFL students during presentations.

### 3.4 Design Goals

Based on the findings and insights derived, we summarized the following design goals for a novel delivery support system. In this system, the smartphone takes on the role of an augmented prompter, while the laptop functions as the responder. The aim is to synchronize diverse communication channels:

**[D1] Customizable Delivery Factors (F1):** Users should have the ability to personalize delivery factors based on their individual preferences and needs.

**[D2] Moderate Script Polishing (F5):** The system should enhance the fluency, wording, and sentence structure of users' presentation script while maintaining a balanced level of embellishment.

**[D3] Remote Mobile Visual Control (F8):** Users need to be empowered with remote slide control capabilities on their smartphones.

**[D4] Flexible Speech Pace Modulation (F6):** The system should allow for multilevel modulation of speech pace to accommodate diverse presentation scenarios.

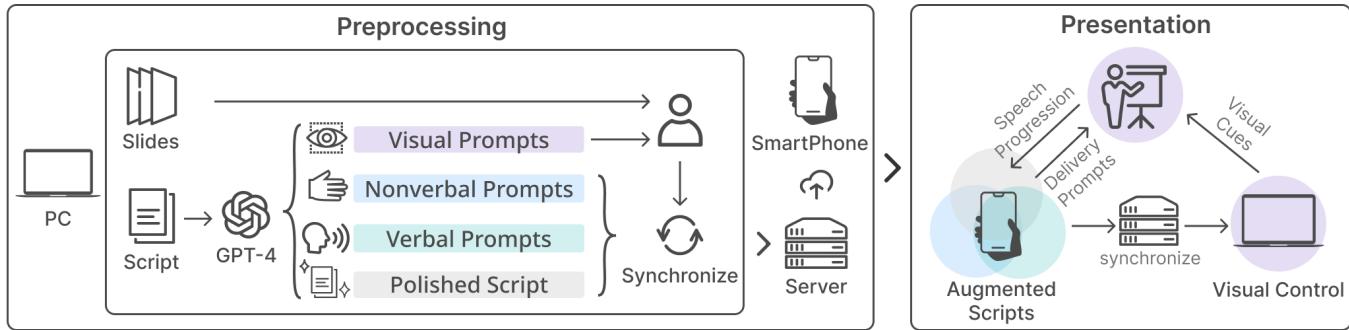


Figure 8: Overview workflow of *Trinity*, including a Preprocessing phase and a Presentation phase.

**[D5] Integrated Delivery Prompts (F4):** Essential prompts for both verbal and nonverbal channels, along with corresponding scripts, should be presented to users to minimize cognitive load.

**[D6] Rapid Pronunciation Acquisition (Insight3):** The system should aid users in quickly mastering the pronunciation of unfamiliar and polysyllabic words.

## 4 TRINITY

### 4.1 Approach Overview

Based on the design goals derived from our formative study, we propose *Trinity*, an interactive system that synchronizes verbal, nonverbal, and visual channels to improve users' AOP delivery on the fly. The workflow of using *Trinity* consists of two distinct phases, namely the **Preprocessing** phase and the **Presentation** phase. This system comprises three integral components: a smartphone-based prompter, a central server, and a PC-end interface (**F3**) (Figure 8).

**1) In the Preprocessing phase,** the PC end uses GPT-4 to process the script, which can be provided separately or written in the remark column. The enhanced script with in-line textual prompts for all three communication channels is generated and sent to the server. Users then add or adjust visual cues on PowerPoint according to the provided prompts. The final package of slides, refined script, and their mappings is uploaded to the smartphone via the server.

**2) In the Presentation phase,** the smartphone-based prompter tracks the speaker's progress and presents the script with **verbal** and **nonverbal** prompts for real-time guidance. Users can remotely control slides via the smartphone, with synchronization data transmitted to the PC end through the server, enabling manipulation of slides and **visual cues**.

### 4.2 Interface Design

**4.2.1 Presentation Script Augmentation.** The PC interface of *Trinity*, an add-in for PowerPoint, caters to delivery support prerequisites and integrates seamlessly as a sidebar (Figure 9). This add-in, comprising **factor checkboxes**, **script input field**, and **operational buttons**, provides comprehensive support.

**[D1] Customizable Delivery Factors.** Recognizing diverse user preferences for delivery support, *Trinity* enables users to customize their selection of delivery factors with a simple click on the corresponding checkboxes (Figure 9-1). *Trinity* also includes a recommended preset, derived from our extensive survey, for balanced

and effective delivery support. It is worth noting that *Trinity* will pop up a reminder when the user selects too many factors (e.g., over 5), signaling the potential risk of overload (**F2**). Moreover, the selected factors are not obligatory for prompt but rather intended for consideration solely in cases where modulation is deemed necessary.

**[D2] Moderate Script Polishing.** *Trinity* uses GPT-4 (**Insight1**) to refine presentation scripts based on expert interview insights. Users can either input scripts paragraph by paragraph or select a local file by clicking the **Select from...** (Figure 9-2) button. If no input is given, *Trinity* uses the script from the Remark column. Users can set a presentation time limit using the slider **—●—** (Figure 9-3) and refine the script by clicking the **Augment** button (Figure 9-4). The server returns a refined script with in-line prompts, with GPT-4 ensuring no overuse of embellishments or excessive expansion. Being aware of potential issues arising from GPT (e.g., odd words and unfamiliar expressions), we grant users full autonomy to adjust the refined scripts in accordance with their personal preferences and habits. Users can toggle between the original and refined scripts for comparison (Figure 9-5), make revisions (Figure 9-6), optimize slide visuals based on prompts (Figure 9-7), label unfamiliar words (Figure 9-8), and upload the revised script by clicking the **Confirm & upload** button (Figure 9-9) for transfer to their smartphones.

**4.2.2 On-the-fly Delivery Support.** The smartphone interface of *Trinity* integrates key delivery support features into an App, including **slide thumbnails**, **progress bars**, and the **augmented script**. Notably, the interface can be adjusted for users' handedness, as is illustrated by the left-handed orientation in Figure 10.

**[D3] Remote Mobile Visual Control.** The smartphone interface of *Trinity* allows filmstrip-style interactions for [visual]visual control. The central static gray box (Figure 11A) serves as a viewfinder to capture slides, and a simple tap (Figure 11B) triggers a click operation at the PC end. Users can switch slides by swiping up and down on the thumbnails (Figure 11C). To reduce cognitive load, only the most relevant thumbnails are shown: the previous, current, and next slide. The thumbnails of the non-current slides are slightly transparent to maintain user focus.

**[D4] Flexible Speech Pace Modulation.** *Trinity* employs a dual-layered approach to regulate speech pace, addressing both broad and fine-grained aspects. Central progress bars (Figure 10A)

929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

1084

1085

1086

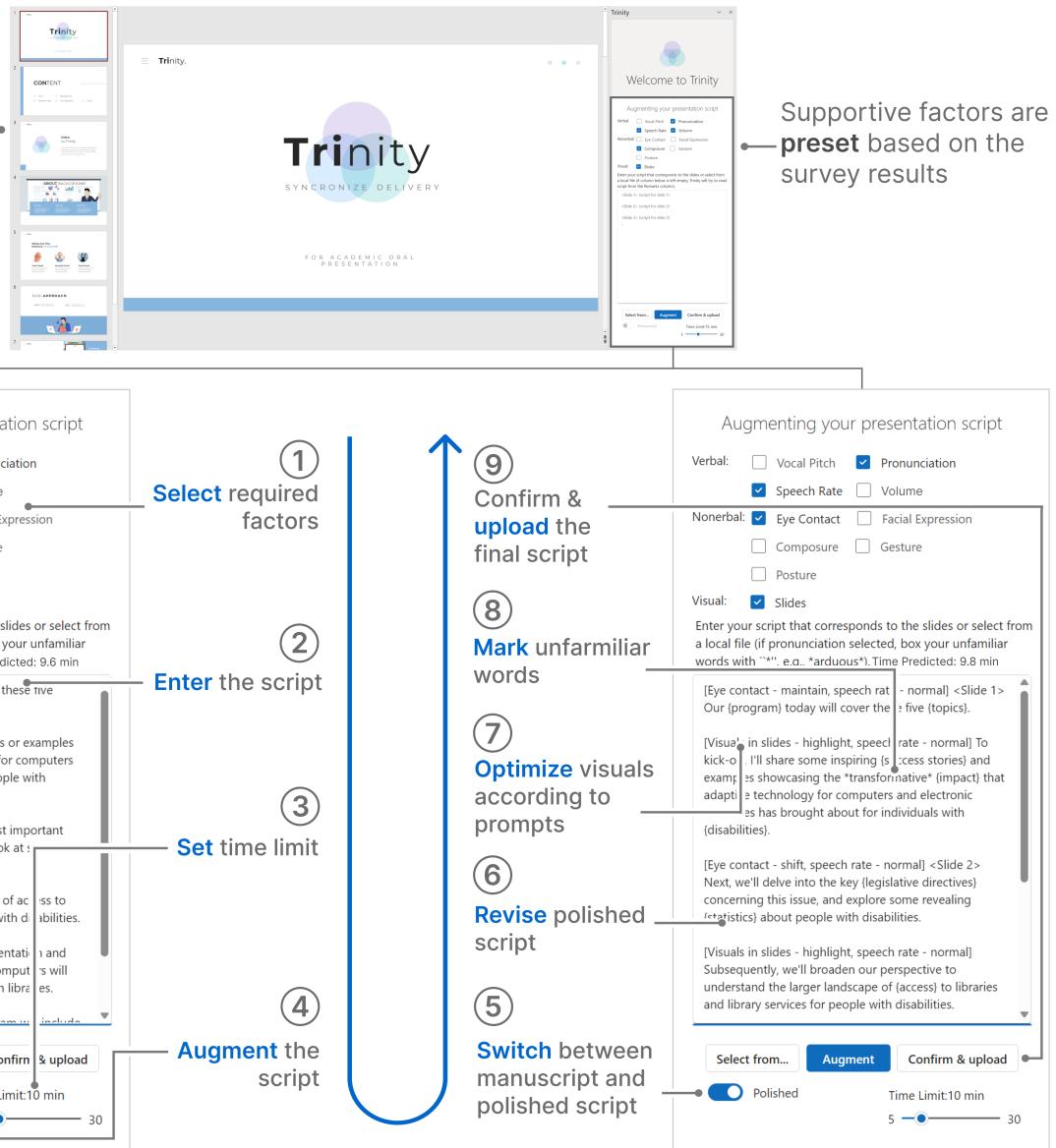
1087

1088

1089

illustrate the overall pace, while the augmented script's underpainting provides detailed pace control. The left (red) and right (blue) progress bars show current and ideal pace respectively, allowing users to quickly gauge their pace. The underpainting highlights the current sentence for reading at the optimal pace. Fast and slow paces are visualized with upward and downward triangle groupings (**Insight2**) in the underpainting (Figure 10B▲▲▲, D▼▼▼). Moreover, *Trinity* synchronizes the speech with the script displayed at the screen's upper part. If the pace is slightly slow, the underpainting gradually fades (Figure 10C).

## Overview of the PC end of Trinity



Supportive factors are preset based on the survey results

**Figure 9: PC end interface of *Trinity*.** A complete workflow on PC end encompasses 9 steps, users 1) select required delivery factors on the checkboxes, 2) enter the script in the input field, 3) set time limit, 4) augment the script, 5) switch between the manuscript and polished script, 6) perform necessary revisions, 7) optimize visuals in the slides according to prompts, 8) mark unfamiliar words, and 9) confirm and upload the final script.

[D5] **Integrated Delivery Prompts.** Drawing from previous work [89, 101], we present **verbal** and **nonverbal** delivery prompts using an integrated approach. To avoid overwhelming users with on-the-fly delivery support, we generate prompts at the start of each sentence (e.g., [volume - loud]), using emojis for encoding due to their rich meanings and familiarity [53, 77, 83]. We enlarge emojis for saliency and resolve any potential ambiguity [8, 71] through author discussions, ensuring the selection of suitable emojis for various modulation scenarios (Table 2). Users have a lookup table within the smartphone interface (Figure 21 in Appendix) to aid

1100

1101

1102

1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159

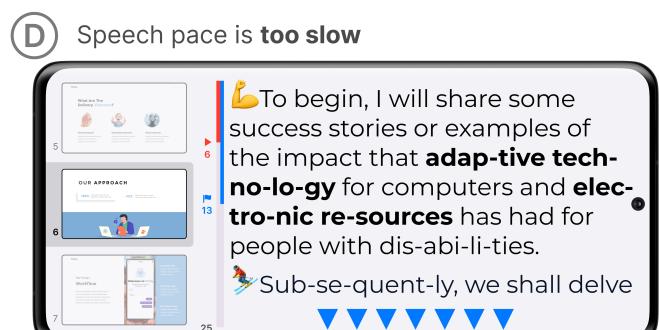
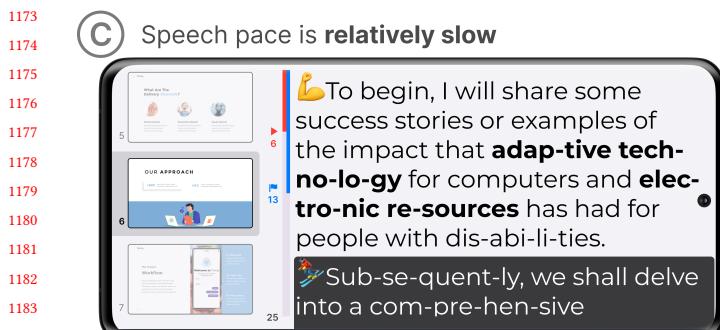
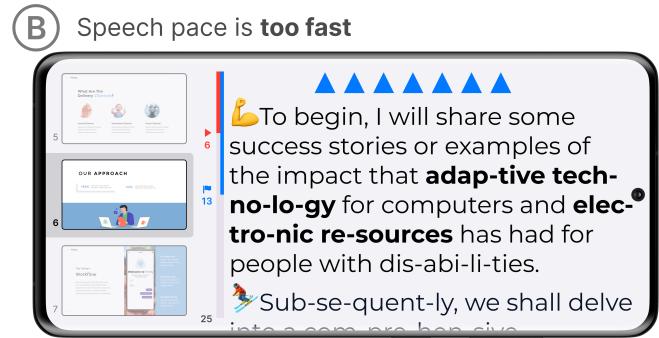
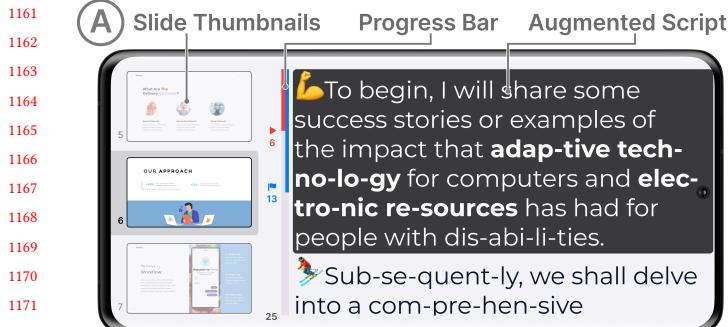


Figure 10: Interface design of *Trinity* App. A) Interface of *Trinity* App consists primarily of three components: *slide thumbnails*, *progress bars*, and the *augmented script*. B) Upward triangle group indicates that speech pace is *too fast*. C) Underpainting fades out as speech pace is *relatively slow*. D) Downward triangle group indicates that speech pace is *too slow*.

Table 2: Emoji mapping for delivery prompts.

Factor	Modulation	Prompt	Factor	Modulation	Prompt	Factor	Modulation	Prompt	Factor	Modulation	Prompt
Vocal pitch	High	↗	Facial expression	Happy	😊	Volume	Loud	🔊	Gesture	Wave	👋
	Normal	—		Sad	😢		Normal	🔉		Unfold	👉
	Low	↘		Angry	😡		Soft	🔈		Point	👉
Speech rate	Fast	➡	Composure	Surprised	😲		Calm	🧘	Posture	Stand	🚶
	Normal	➡️		Embarrassed	😳		Relaxed	😌		Walk	🚶
	Slow	➡➡		Serious	😐		Confident	⭐		Eye contact	👀

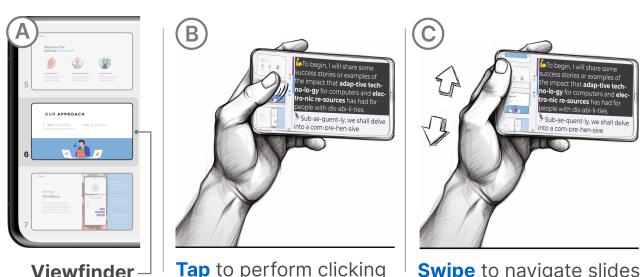


Figure 11: Interactions for *Trinity* App. A) The viewfinder appears as a static gray box to capture the slides intended for display. B) Tapping in the viewfinder to perform clicking. C) Swiping the slide thumbnails to promptly switch among slides.

emoji interpretation and the option to deactivate emoji transcriptions. Moreover, we bold the keywords within the script to further inform users. Notably, the use of GPT-4 for delivering prompts is optional; users can manually input prompts or deactivate this feature by deselecting associated checkboxes. Once the final script with

1248 textual delivery prompts is uploaded, the server end transcribes the 1249 delivery factors and their modulations generated by GPT-4 into 1250 corresponding emojis (e.g., [gesture - point] →👉), and then transfers 1251 the script with emojis to the mobile end.

**[D6] Rapid Pronunciation Acquisition.** To improve presentation 1252 scripts and mitigate disruptions from unfamiliar or polysyllabic 1253 words, we initially considered using phonetic symbols. However, 1254 user feedback indicated that while helpful for pronunciation, this 1255 could hinder fluent delivery. Instead, we used phonics to segment 1256 unfamiliar words marked (Figure 9–8) and polysyllabic words into 1257 syllables (e.g., "Subsequently" → "Sub-se-quently"). This approach 1258 leverages users' existing phonological knowledge, facilitating swift 1259 pronunciation acquisition without imposing a significant learning 1260 curve.

### 4.3 Implementation Details

The PC end of *Trinity* is an office project created with Yeoman generator<sup>2</sup>, designed as a PowerPoint add-in using Office.js for

<sup>2</sup><https://learn.microsoft.com/en-us/office/dev/add-ins/develop/yeoman-generator-overview>

1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210  
2211  
2212  
2213  
2214  
2215  
2216  
2217  
2218  
2219  
2220  
2221  
2222  
2223  
2224  
2225  
2226  
2227  
2228  
2229  
2230  
2231  
2232  
2233  
2234  
2235  
2236  
2237  
2238  
2239  
2240  
2241  
2242  
2243  
2244  
2245  
2246  
2247  
2248  
2249  
2250  
2251  
2252  
2253  
2254  
2255  
2256  
2257  
2258  
2259  
2260  
2261  
2262  
2263  
2264  
2265  
2266  
2267  
2268  
2269  
2270  
2271  
2272  
2273  
2274  
2275  
2276  
2277  
2278  
2279  
2280  
2281  
2282  
2283  
2284  
2285  
2286  
2287  
2288  
2289  
2290  
2291  
2292  
2293  
2294  
2295  
2296  
2297  
2298  
2299  
2300  
2301  
2302  
2303  
2304  
2305  
2306  
2307  
2308  
2309  
2310  
2311  
2312  
2313  
2314  
2315  
2316  
2317  
2318  
2319  
2320  
2321  
2322  
2323  
2324  
2325  
2326  
2327  
2328  
2329  
2330  
2331  
2332  
2333  
2334  
2335  
2336  
2337  
2338  
2339  
2340  
2341  
2342  
2343  
2344  
2345  
2346  
2347  
2348  
2349  
2350  
2351  
2352  
2353  
2354  
2355  
2356  
2357  
2358  
2359  
2360  
2361  
2362  
2363  
2364  
2365  
2366  
2367  
2368  
2369  
2370  
2371  
2372  
2373  
2374  
2375  
2376  
2377  
2378  
2379  
2380  
2381  
2382  
2383  
2384  
2385  
2386  
2387  
2388  
2389  
2390  
2391  
2392  
2393  
2394  
2395  
2396  
2397  
2398  
2399  
2400  
2401  
2402  
2403  
2404  
2405  
2406  
2407  
2408  
2409  
2410  
2411  
2412  
2413  
2414  
2415  
2416  
2417  
2418  
2419  
2420  
2421  
2422  
2423  
2424  
2425  
2426  
2427  
2428  
2429  
2430  
2431  
2432  
2433  
2434  
2435  
2436  
2437  
2438  
2439  
2440  
2441  
2442  
2443  
2444  
2445  
2446  
2447  
2448  
2449  
2450  
2451  
2452  
2453  
2454  
2455  
2456  
2457  
2458  
2459  
2460  
2461  
2462  
2463  
2464  
2465  
2466  
2467  
2468  
2469  
2470  
2471  
2472  
2473  
2474  
2475  
2476  
2477  
2478  
2479  
2480  
2481  
2482  
2483  
2484  
2485  
2486  
2487  
2488  
2489  
2490  
2491  
2492  
2493  
2494  
2495  
2496  
2497  
2498  
2499  
2500  
2501  
2502  
2503  
2504  
2505  
2506  
2507  
2508  
2509  
2510  
2511  
2512  
2513  
2514  
2515  
2516  
2517  
2518  
2519  
2520  
2521  
2522  
2523  
2524  
2525  
2526  
2527  
2528  
2529  
2530  
2531  
2532  
2533  
2534  
2535  
2536  
2537  
2538  
2539  
2540  
2541  
2542  
2543  
2544  
2545  
2546  
2547  
2548  
2549  
2550  
2551  
2552  
2553  
2554  
2555  
2556  
2557  
2558  
2559  
2560  
2561  
2562  
2563  
2564  
2565  
2566  
2567  
2568  
2569  
2570  
2571  
2572  
2573  
2574  
2575  
2576  
2577  
2578  
2579  
2580  
2581  
2582  
2583  
2584  
2585  
2586  
2587  
2588  
2589  
2590  
2591  
2592  
2593  
2594  
2595  
2596  
2597  
2598  
2599  
2600  
2601  
2602  
2603  
2604  
2605  
2606  
2607  
2608  
2609  
2610  
2611  
2612  
2613  
2614  
2615  
2616  
2617  
2618  
2619  
2620  
2621  
2622  
2623  
2624  
2625  
2626  
2627  
2628  
2629  
2630  
2631  
2632  
2633  
2634  
2635  
2636  
2637  
2638  
2639  
2640  
2641  
2642  
2643  
2644  
2645  
2646  
2647  
2648  
2649<br

1277 seamless use. It comprises around 1,000 lines of TypeScript code  
 1278 developed within ReactJS. The server side, implemented in about  
 1279 500 lines of Python code, uses Flask and the OpenAI API<sup>3</sup> to re-  
 1280 fine scripts, transcribe delivery prompts, and connect the PC and  
 1281 smartphone ends.

1282 The Smartphone end of *Trinity* is based on Android due to its  
 1283 market dominance and popularity. It captures voice input, converts  
 1284 it to text using Android's SpeechRecognizer<sup>4</sup>, and processes the  
 1285 text with the BM25 string matching ranking algorithm for accurate  
 1286 speech progress monitoring and synchronized script scrolling [72].  
 1287 The Android App comprises around 1,000 lines of Kotlin<sup>5</sup> code  
 1288 within the Compose framework<sup>6</sup>.

## 1289 5 USER STUDY

1290 To provide a thorough assessment of the proposed *Trinity* in relation  
 1291 to its **usability and effectiveness (RQ3)**, **usage patterns and**  
 1292 **system impact (RQ4)**, as well as **trust and collaboration (RQ5)**,  
 1293 we executed a controlled between-subject user study. In this study,  
 1294 EFL students were tasked with delivering AOPs using *Trinity*, while  
 1295 two other baseline systems were employed as control conditions.  
 1296

### 1297 5.1 Conditions

1298 *Trinity* is built upon the most common practice in presentation,  
 1299 PowerPoint, and integrates both mobile and PC devices for sup-  
 1300 port, with the focus of providing on-the-fly guidance on the mobile  
 1301 end. To evaluate the realization of our six design goals, we com-  
 1302 pared *Trinity* with two baseline systems which are also built upon  
 1303 PowerPoint, *IntelliPrompter* and *OfficeRemote*. 1) *IntelliPrompter*  
 1304 (Figure 6C) enhances PowerPoint's presenter view, automatically  
 1305 tracking the presenter's slide coverage and adjusting the note dis-  
 1306 play on a PC. This baseline was chosen due to its shared features  
 1307 with *Trinity*, such as speech tracking and dynamic script display,  
 1308 facilitating comparison of effects introduced by mobile support. 2)  
 1309 *OfficeRemote* (Figure 6A) is a mobile app that connects with Pow-  
 1310 erPoint on a desktop, designed for slide navigation, script reference,  
 1311 timer control, and laser pointer feature, all accessible through a  
 1312 mobile phone. We selected this baseline because it operates on mo-  
 1313 bile like *Trinity* App and offers basic on-the-fly delivery supports  
 1314 as well, facilitating the comparison of designed delivery support  
 1315 provided in *Trinity*. Initially, EFL students were randomly divided  
 1316 into three groups, each using a different system. These groups were  
 1317 then randomly and evenly allocated to three sessions to deliver two  
 1318 5-minute AOPs without Q&A sessions.  
 1319

1320 **Table 3: Comparison of baseline systems.**

<i>IntelliPrompter</i>	<i>Office Remote</i>	<i>Trinity</i>
Dynamic Script Dispaly	Static	
Mobile Silde Navigation	.	
Speech Pace Modulation	.	
Script Polishing	.	
Delivery Prompts	.	
Built upon Powerpoint		
Media	PC	Mobile & PC
		Mobile & PC

1330 <sup>3</sup><https://platform.openai.com/docs/api-reference/chat>

1331 <sup>4</sup><https://developer.android.com/reference/kotlin/android/speech/SpeechRecognizer>

1332 <sup>5</sup><https://kotlinlang.org/>

1333 <sup>6</sup><https://developer.android.com/jetpack/compose>

## 1334 5.2 Participants

1335 Following IRB approval, we recruited 65 EFL student presenters  
 1336 (29 females, 36 males, S1-S65, average age 21.5, *SD* = 2.13) and 25  
 1337 audience members (11 females, 14 males, A1-A25, average age 25.4,  
 1338 *SD* = 4.83) from a local university through email and social media.  
 1339 The presenters came from various majors, and all had AOP expe-  
 1340 rience. The audience comprised undergraduates, graduates, and  
 1341 researchers from diverse fields. Five instructors (2 female, 3 males,  
 1342 I1-I5, average age 41.6, *SD* = 6.12) also joined as audience. The  
 1343 presenters were initially divided randomly into three groups of 22  
 1344 for the *Trinity* and *IntelliPrompter* conditions, and 21 for the *OfficeRe-*  
 1345 *mote* condition. After three dropouts, *Trinity* and *IntelliPrompter*  
 1346 conditions had 21 presenters, and *OfficeRemote* conditions had 20  
 1347 presenters.  
 1348

## 1349 5.3 Procedure

1350 In this study, each EFL student delivered two 5-minute AOPs in  
 1351 English, separated by a brief break. Students prepared presentation  
 1352 materials, including slides and scripts, with specific criteria: 1) They  
 1353 needed a good understanding of the content without extensive re-  
 1354hearsal, maintaining moderate preparation. 2) The AOPs covered  
 1355 diverse genres: one as a paper sharing session (discussing others'  
 1356 work) and the other as a project report (covering their own work).  
 1357 Materials were submitted at least one day in advance for compli-  
 1358 ance checks with three criteria: 1) Each presenter must report an  
 1359 approximate preparation time to ensure moderate preparation. 2)  
 1360 The presentation content must be complete (i.e. from introduction  
 1361 to discussion). 3) Materials must be in English and cover both pa-  
 1362 per sharing and project report. For the *Trinity* group, participants  
 1363 installed the app on their laptops and smartphones, completing pre-  
 1364 processing with screen recording based on provided instructions.  
 1365 Two authors facilitated the experiment and addressed technical  
 1366 issues as needed.

1367 The experiment, illustrated in Figure 12, had three phases: prepara-  
 1368 tory, execution, and closing. In the preparatory phase, participants  
 1369 filled out a consent form and a pre-task survey. All presenters  
 1370 received a 5-minute system tutorial. In the execution phase, pre-  
 1371 senters in the *Trinity*, *IntelliPrompter*, and *OfficeRemote* conditions  
 1372 had 10 minutes for system setup and familiarization. Presenters  
 1373 were then randomly and evenly assigned to three AOP sessions.  
 1374 A 5-10 minute introduction was given before the AOPs, during  
 1375 which instructors and audience assessed presenter performance  
 1376 using evaluation forms (Figure 22). Post-AOP, presenters completed  
 1377 an in-task questionnaire. In the closing phase, a post-task question-  
 1378 naire was filled by all participants, followed by semi-structured  
 1379 interviews with 12 audience members and all the presenters in  
 1380 *Trinity* condition separately. To maximize the blindness of the au-  
 1381 dience, we didn't disclose experimental conditions from the outset  
 1382 of recruitment to the end of the interviews.

1383 The whole process of this study took around 11 hours in total.  
 1384 The AOPs are scheduled with a 15-minute break inserted every  
 1385 40 minutes to mitigate the fatigue effect. Each day, there is an  
 1386 approximate duration of 2.2 hours dedicated to AOPs. We ensured  
 1387 consistency by using a 70-inch monitor for slide projection and a  
 1388 tripod-mounted camera to record presentations, with participants'  
 1389 consent. Presenters were asked to bring their own laptops and  
 1390

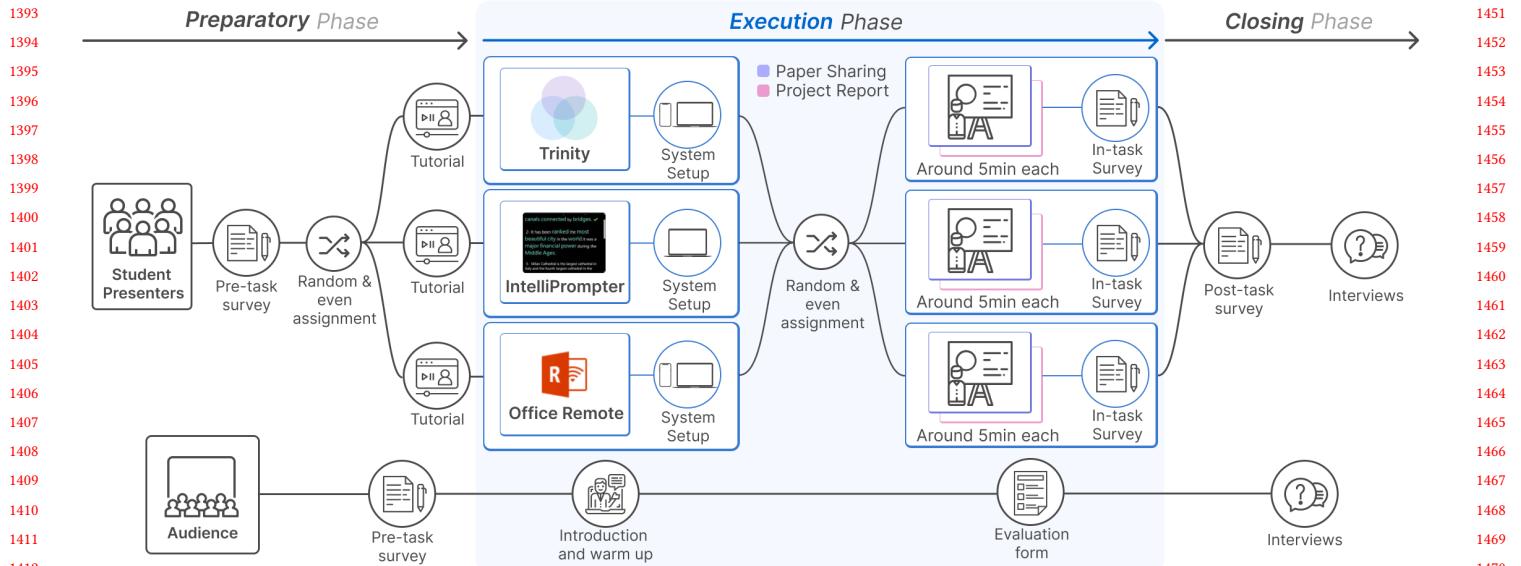


Figure 12: Design and procedure of user study. 1) In the preparatory phase, student presenters were given a pre-task survey and assigned to one of three conditions randomly and evenly with tutorials. 2) In the execution phase, presenters across three conditions were randomly and evenly assigned to three sessions to give a 5-minute paper sharing and a 5-minute project report. An in-task survey was conducted alongside these presentations. Following the warm-up session, audience members evaluated the performance of presenters using evaluation forms. 3) In the closing phase, both presenters and audience members participated separately in post-task surveys and interviews.

smartphones for use, with standby devices available for iOS users and emergencies. Participants were compensated at a rate of \$15 USD/hr.

#### 5.4 Measurement

**Student Presenters' Self-report.** As detailed in Table 4, in our study, we employed a 7-point Likert in-task questionnaire to gather feedback from student presenters regarding their experiences using different support systems during AOPs [55, 60]. This questionnaire covered three key aspects: **RQ3. Usability and Effectiveness**. We designed questions on the usability and effectiveness of the systems based on prior research [9, 55, 60, 60, 61, 65]; **RQ4. Usage Pattern and System Impact**. We collected interaction logs to track usage patterns and assessed students' emotions and workload using questions inspired by existing studies [5, 55, 60, 79] and the NASA-TLX survey [34]; **RQ5. Trust and Collaboration**. We administered a post-task questionnaire to gauge presenters' trust in our system [55]. Additionally, semi-structured interviews were conducted with student presenters, addressing topics related to **RQ3-RQ5**.

**Presenter Evaluation by Audience.** We employed a standardized rubric [66] for the audience to evaluate students' presentation performance. This rubric included assessments of 1) verbal delivery, 2) nonverbal delivery, 3) visual aids, 4) delivery consistency, and 5) content&organization. The evaluation form (as detailed in Figure 22 in Appendix) utilized a 4-point scale with detailed descriptions for each option. Additionally, we conducted interviews with the audience, inviting them to provide further comments through open-ended questions.

#### 5.5 User Study Limitations

**Variable control.** Our system is developed upon the new design goals derived in the Formative Study, making it challenging to strictly control the variables for evaluating newly designed features. In the early iteration of our user study, we considered applying an ablation study manner to strictly configure the conditions for newly designed features. However, the total number of conditions (8 conditions) makes it unaffordable for researchers to apply in a user study and the fatigue effect becomes inevitable for participants. Therefore, we follow the practices in HCI research (e.g., [55, 95, 98]) and use two systems with shared features (Table 3) for baseline comparison. In the following, we report the directional results, while more in-depth and fine-grained evaluation of each feature's impact may leave for the future works.

**Audience blindness.** Presentation is a process through which presenters communicate information to an audience on the spot, making it challenging to strictly ensure the blindness of audience. In our user study, EFL students are expected to employ various communication channels while the audience evaluates their actual performance. In the early iteration of our user study, we considered using certain apparatus (e.g., a board erected in front of the presenter) to mask the experimental conditions; however, this configuration proved impractical and significantly hindered both presentation delivery and evaluation. Therefore, although we did not disclose the experimental conditions throughout recruitment and interviews, the audience may still infer some condition-related information from presenters' usage patterns (i.e., how they interact with devices), potentially influencing their ratings. Future works may focus on addressing the challenge of audience blindness in

1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507

**Table 4: Metrics of student presenter' self-report to investigate RQ3-RQ5, organized by research question, aspect, and approach.**

Research Question	Aspect	Approach	Metric
<b>RQ3.</b> How are the usability and effectiveness of the supportive system?	Usability	In-task questionnaire	1) ease-of-use, 2) helpfulness in supporting AOP delivery, 3) distraction caused by the system, 4) system satisfaction, 5) likelihood for future use.
	Effectiveness		1) making the speech more fluid, coherent and modulated, 2) facilitating good eye contacts with audience, 3) facilitating appropriate facial expressions, 4) facilitating natural gestures, 5) enhancing easiness of visual control, 6) enhancing effectiveness of slides, 7) enhancing control of the presentation progress, 8) making the presentation more engaging
<b>RQ4.</b> How will EFL students interact with and be influenced by the system?	Interaction	System Impact	Collect students' interaction logs on the interface and counted the frequency of students' glancing at the interface.
			Emotions: 1) noticed, 2) excited, 3) exhausted, 4) frustrated, 5) happy, 6) hopeful, 7) overwhelmed, 8) safe, 9) nervous, 10) anxious, 11) confident.
<b>RQ5.</b> How will students trust and collaborate with our system?	Trust	Post-task questionnaire	Load: 1) cognitive load, 2) attentional load, 3) workload, 4) self-perceived performance
	Collaboration	Semi-structured interviews	1) perceptions of the usefulness of the provided system features, 2) perceived accuracy of the system and overall trust in the system, 3) trust in different features embodied in the system.
			Student presenters' feedback

**Table 5: Summary of qualitative insights on design choices of *Trinity*, organized by design goals, to enlighten future presentation support tools.**

Design Goal	Trinity Choice	Success	Challenges
[D1] Customizable Delivery Factors	selectable factors; recommended preset	support participants' diverse individual preferences; incorporates recommended preset	require to understand self-needs; excessive number of factors can overwhelm presenters
[D2] Moderate Script Polishing	script polishing with GPT-4; time limit control; script comparison and revision	improve the overall quality and effectiveness; save preparation time; boost the success in conveying messages or ideas.	LLM may entail script-related potential misunderstandings; LLM may mismatch with participants' speech style;
[D3] Remote Mobile Visual Control	thumb tapping and swiping to navigate slides	effortlessly control visuals via smartphone; minimize cognitive load	lacks the display of each animation step; limits smartphone holding posture
[D4] Flexible Speech Pace Modulation	progress bars; background underpainting; triangle grouping	manage and adjust the speaking pace; cater to both macroscopic and microscopic aspects	central progress bars are prone to be overlooked; speech recognition jitters may misdirects to incorrect scripts
[D5] Integrated Delivery Prompts	emojis; keywords highlighting	vividly present emotions and nonverbal cues; ensure a personalized experience	nonverbal cues prone to be more irresponsive than verbal cues; response effect is somewhat subpar
[D6] Rapid Pronunciation Acquisition	phonics-based segmentation	allows participants to acquire correct pronunciation quickly	somewhat intensify the workload; presenters exhibit a noticeable spelling process

presentation scenarios to enhance the rigor of subsequent investigations.

## 6 ANALYSES AND RESULTS

This section analyzes RQ3-RQ5 using a mixed-methods approach. Quantitatively, Kruskal-Wallis with post-hoc Dunn's tests and Bonferroni correction were used for in-task questionnaires, and descriptive statistics for post-task responses. One-way ANOVA with post-hoc Tukey tests studied the impact of delivery support tools

on perceived AOP performance. Qualitatively, we reviewed video recordings of AOP sessions with *Trinity*, identified interaction timestamps, and discussed with presenters to understand motivations and reactions. Feedback was audio-recorded, transcribed, and thematically analyzed by two authors until consensus. We present results addressing tensions in tool usage, noting how design choices both facilitated and introduced challenges for *Trinity*. Achievements and obstacles are categorized based on *Trinity*'s design goals, providing

1625 insights for future research. Primary findings are presented here,  
 1626 with minor findings in Appendix C.

## 1628 6.1 RQ3. How are the usability and effectiveness 1629 of the supportive system?

1630 We present findings from our mixed-methods analysis, evaluating  
 1631 the usability of *Trinity* for student presenters compared to baseline  
 1632 systems. We also analyse the effectiveness of *Trinity* features as  
 1633 reported by students and assess AOP performance from the audi-  
 1634 ence's perspective.

1635 **i) Usability.** The results of Kruskal-Wallis tests indicate signif-  
 1636 icant differences among the systems in terms of their perceived  
 1637 helpfulness in supporting AOP delivery ( $H = 9.471, p = 0.007$ )  
 1638 and the likelihood of future use ( $H = 10.382, p = 0.005$ ). However,  
 1639 there is no significant difference in terms of ease-to-use ( $H = 4.523,$   
 1640  $p > 0.05$ ):

1641 “*Trinity is not a plug-and-play for me. It took me some time to get  
 1642 familiar with the utilities and figure out how to it works before I  
 1643 put it into practical use.*” (S10, male, age: 21)

1644 Post-hoc analysis revealed that *Trinity* was significantly more help-  
 1645 ful for supporting AOP delivery and received higher ratings for  
 1646 intended future use compared to both *IntelliPrompter* ( $p = 0.032$  and  
 1647  $p = 0.010$ ) and *OfficeRemote* ( $p = 0.024$  and  $p = 0.041$ ). However,  
 1648 there were no significant differences across the three systems in  
 1649 terms of distraction caused by the system and system satisfaction,  
 1650 as indicated by Kruskal-Wallis tests.

1651 **ii) Effectiveness.** Figure 13 presents participant-perceived use-  
 1652 fulness of *Trinity*'s features in descending order. Most features were  
 1653 deemed useful but posed some challenges. Notably, participants  
 1654 highly rated flexible remote visual control, keyword highlighting,  
 1655 and the fine-grained speech pace indicator (underpainting), while  
 1656 also finding integrated delivery prompts helpful.

1657 Subsequently, we conducted an in-depth analysis of how the  
 1658 *Trinity* system supported EFL student presenters in delivering AOP  
 1659 from various perspectives. Presenter ratings on relevant questions  
 1660 are presented in Figure 16. Kruskal-Wallis tests with post-hoc tests  
 1661 showed that *Trinity* outperformed both *IntelliPrompter* and *OfficeRe-  
 1662 mote* in all aspects (Figure 16). This findings align with participants'  
 1663 qualitative feedback from interviews. Participants found *Trinity* su-  
 1664 perior due to its “*comprehensive and customizable delivery support*”  
 1665 (S2, S28, S37, S45), which “*facilitates adaptive delivery*” (S14, S23,  
 1666 S28), “*enriches presentation content*” (S27, S47), and “*enhances control  
 1667 over presentations*” (S3, S31, S39). They also noted the interface as  
 1668 “*easy to understand and interact with*” (S14, S55).

1669 regarding making the speech fluid, coherent and modulated  
 1670 ( $H = 8.726, p = 0.012$ , pairwise:  $p_{TI} = 0.060, p_{TO} = 0.025$ <sup>7</sup>), facil-  
 1671 itating good eye contacts with audience ( $H = 9.198, p = 0.010$ , pair-  
 1672 wise:  $p_{TI} = 0.18, p_{TO} = 0.060$ ), facilitating appropriate facial expres-  
 1673 sions ( $H = 11.075, p = 0.004$ , pairwise:  $p_{TI} = 0.071, p_{TO} = 0.005$ ),  
 1674 facilitating natural gestures ( $H = 10.104, p = 0.006$ , pairwise:  
 1675  $p_{TI} = 0.009, p_{TO} = 0.065$ ), enhancing easiness of visual control  
 1676 ( $H = 8.870, p = 0.012$ , pairwise:  $p_{TI} = 0.009, p_{TO} = 0.316$ <sup>8</sup>), en-  
 1677 hancing effectiveness of slides ( $H = 10.824, p = 0.004$ , pairwise:  
 1678

1679 <sup>7</sup> $p_{TI}$ :  $p$  value between *Trinity* and *IntelliPrompter*;  $p_{TO}$ :  $p$  value between *Trinity* and  
 1680 *OfficeRemote*.

1681 <sup>8</sup>*OfficeRemote* received higher scores here due to its simplicity.

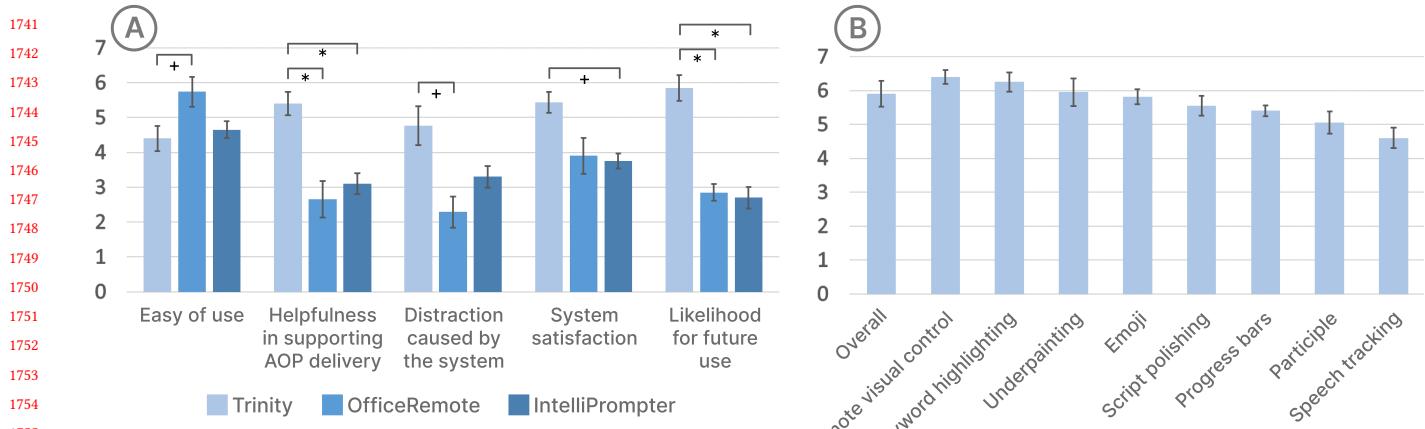
1682  $p_{TI} = 0.010, p_{TO} = 0.028$ ), enhancing control of the presentation  
 1683 progress ( $H = 10.853, p = 0.004$ , pairwise:  $p_{TI} = 0.005, p_{TO} =$   
 1684  $0.066$ ), making the presentation more engaging ( $H = 7.207, p =$   
 1685  $0.027$ , pairwise:  $p_{TI} = 0.094, p_{TO} = 0.005$ ). These findings align  
 1686 with participants' qualitative feedback obtained through interviews.  
 1687 Participants perceived *Trinity* to be more beneficial than the two  
 1688 baselines due to its provision of “*comprehensive and customizable  
 1689 delivery support*” which effectively “*facilitates adaptive delivery*”,  
 1690 “*enriches presentation content*”, and “*enhances control over presenta-  
 1691 tions*”. Additionally, participants found the interface design to be  
 1692 “*easy to understand and interact*”.

1693 Finally, we assess presenters' AOP performance from the **audi-  
 1694 ence's perspective**. As depicted in Figure 14, regarding **nonverbal  
 1695 delivery**, one-way ANOVA with post-hoc Tukey tests revealed that  
 1696 presenters in the *Trinity* condition received significantly higher rat-  
 1697 ings for eye contact ( $F(2, 60) = 16.673, p = 0.001$ , pairwise:  $p_{TI} =$   
 1698  $0.014, p_{TO} = 0.001$ <sup>9</sup>), facial expression ( $F(2, 60) = 23.416, p <$   
 1699  $0.001$ , pairwise:  $p_{TI} = 0.001, p_{TO} = 0.001$ ), composure ( $F(2, 60) =$   
 1700  $14.728, p = 0.003$ , pairwise:  $p_{TI} = 0.015, p_{TO} = 0.018$ ), and gesture  
 1701 ( $F(2, 60) = 13.532, p = 0.003$ , pairwise:  $p_{TI} = 0.016, p_{TO} = 0.003$ ).  
 1702 Specifically, seven audience members observed presenters using  
 1703 “*consistent eye contact*” (A3, A5) and a “*sweeping gaze*” (A13) while  
 1704 holding a phone. They showed adaptability with facial expressions,  
 1705 “*frowning when discussing challenges*” (A15, A17) and “*smiling when  
 1706 presenting findings*” (A5). These presenters appeared “*well-prepared  
 1707 and imperturbable*” (A21), demonstrating “*control over their pres-  
 1708 entations*” (A24). They used active gestures to “*emphasize key points  
 1709 through pointing*” (A15) and “*enhance explanations by opening their  
 1710 arms*” (A5). Typically, they kept “*one hand in a stationary position*”  
 1711 (A18) to secure the phone and used “*the other hand for gestures  
 1712 and movements*” (A18, A24). Despite differences in eye contact, the  
 1713 audience found no variations in notes usage, as all presenters relied  
 1714 on scripts, “*whether they read them from computers or phones*” (A2,  
 1715 A5, A9, A20). Three audience members also noted little difference  
 1716 in posture, possibly due to minimal posture prompts.

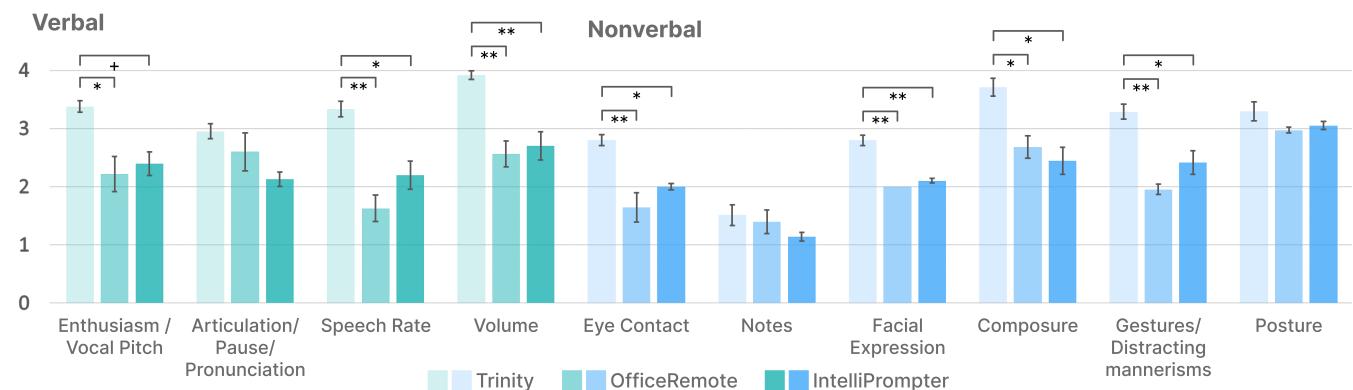
1717 In **verbal delivery** (Figure 14), our tests indicate that presenters  
 1718 in *Trinity* condition received significantly higher ratings for vocal  
 1719 pitch ( $F(2, 60) = 10.434, p = 0.010$ , pairwise:  $p_{TI} = 0.053, p_{TO} =$   
 1720  $0.012$ ), speech rate ( $F(2, 60) = 13.215, p = 0.002$ , pairwise:  $p_{TI} =$   
 1721  $0.045, p_{TO} = 0.002$ ), and volume ( $F(2, 60) = 15.362, p = 0.001$ ,  
 1722 pairwise:  $p_{TI} = 0.002, p_{TO} = 0.002$ ). Specifically, nine audience  
 1723 members noted that presenters who “*held the phone sideways*” (A2,  
 1724 A3) showed more vocal variety, using a “*louder voice to emphasize  
 1725 important points*” (A9, A13), speaking “*faster when discussing previ-  
 1726 ous works*” (A15, A17, A18), and “*raising pitch while posing questions*”  
 1727 (A3, A24, A25), unlike those who “*read their notes monotonously*”  
 1728 (A2, A5, A18, A25).

1729 In **visual delivery** (Figure 15), our tests indicate that *Trinity*  
 1730 user received significantly higher ratings for slides effectiveness  
 1731 ( $F(2, 60) = 7.352, p = 0.024$ , pairwise:  $p_{TI} = 0.033, p_{TO} = 0.010$ ),  
 1732 with no significant difference in slide spelling and references. Specif-  
 1733 ically, seven audience members noted that some presenters over-  
 1734 whelmed them with “*excessive text or bullet points*” (A5, A24), leading  
 1735 to questioning the value of attending (A9). However, exceptional  
 1736

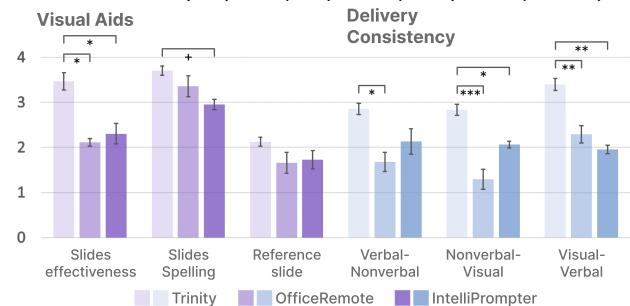
1737 <sup>9</sup> $p_{TI}$ :  $p$  value between *Trinity* and *IntelliPrompter*;  $p_{TO}$ :  $p$  value between *Trinity* and  
 1738 *OfficeRemote*.



**Figure 13: The usability and usefulness of systems. A) The usability of different systems perceived by student presenters. B) The perceived usefulness of different features provided by *Trinity*. The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).**

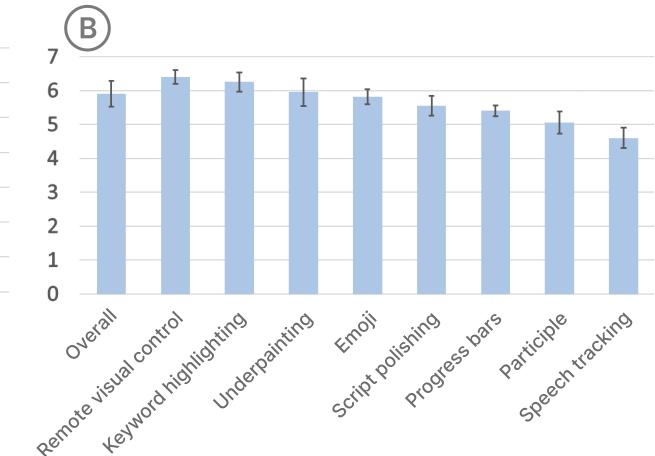


**Figure 14: Audience's ratings on verbal and nonverbal delivery of presenters using different systems. The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).**



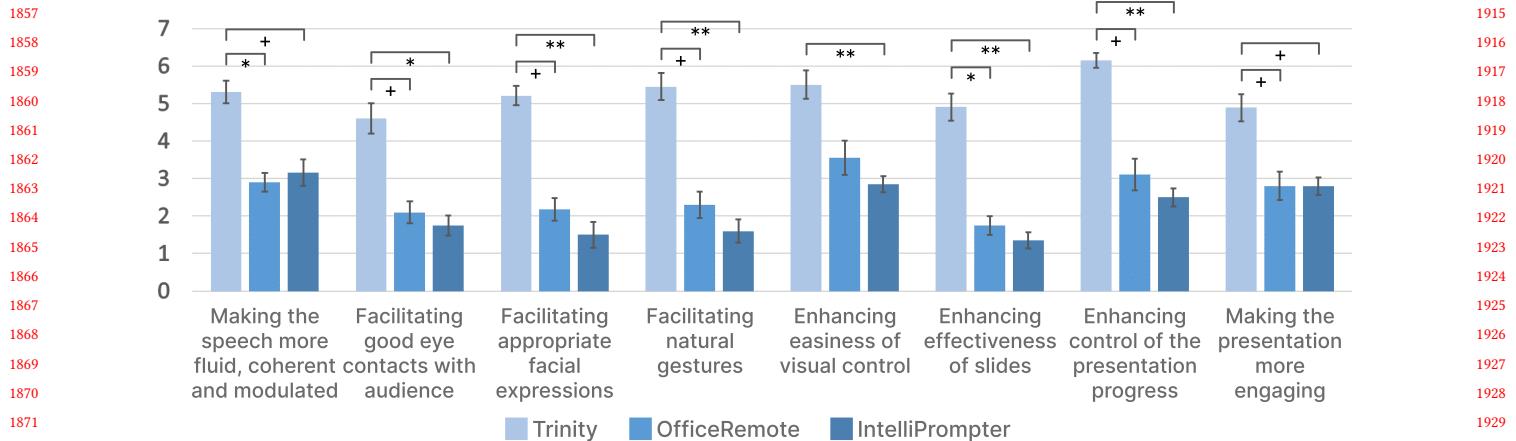
**Figure 15: Audience's ratings on visual delivery and delivery consistency of presenters using different systems. The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).**

presenters use "highlighting key terms" (A2, A5, A9) and "incorporating diagrams" (A2, A15) to make content "more accessible" (A24) and facilitate "better understanding" (A18). They create an environment where the audience can "easily grasp complex concepts" (A5) without feeling overwhelmed.



For **delivery consistency**, our tests indicate that *Trinity* user received significantly higher ratings for speech-behavior consistency ( $F(2, 60) = 9.532, p = 0.018$ , pairwise:  $p_{TI} = 0.143, p_{TO} = 0.023$ ), behavior-visual consistency ( $F(2, 60) = 18.351, p < 0.001$ , pairwise:  $p_{TI} = 0.012, p_{TO} < 0.001$ ), and speech-visual consistency ( $F(2, 60) = 17.341, p < 0.001$ , pairwise:  $p_{TI} = 0.001, p_{TO} = 0.003$ ). Many presenters "read from their scripts or slides" (A3, A20), but some also "not only present but also coordinate" (A5), aligning "what they say, what they do, and what they show" (A13). Examples include pointing to bold words with higher volume (A18), giving opening statements with a smile and open arms (A9), and pointing and zooming in on mentioned locations (A24).

For **content&organization** (Figure 24 in appendix), our tests indicate that *Trinity* users received significantly higher ratings for the opening statement ( $F(2, 60) = 12.357, p = 0.001$ , pairwise:  $p_{TI} = 0.003, p_{TO} = 0.004$ ), transitions ( $F(2, 60) = 21.628, p < 0.001$ , pairwise:  $p_{TI} < 0.001, p_{TO} = 0.002$ ), and organization ( $F(2, 60) = 10.379, p = 0.004$ , pairwise:  $p_{TI} = 0.001, p_{TO} = 0.021$ ), with no significant difference in other factors. Six audience members noted that some presenters, unlike others who "gave a quick hello" (A9) and "skimmed through the title" (A2, A9), used a phone to "start



**Figure 16: Student presenters' ratings on how different systems support their delivery. The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).**

with the background" (A2, A5, A9) and "gradually lead into the main content" (A5). This approach helped those not "well-versed in a particular field or specialty" (A21) to "grasp the main ideas" (A2, A18) and "follow along more easily" (A9). Additionally, these presenters "provided an outline of content" (A24) after introduction and "added smooth transitions" (A21), avoiding "jumping next directly" (A9), making their content "clear and coherent" (A9, A18).

In summary, audience and student presenter ratings and feedback illustrate *Trinity*'s usability and its effectiveness in supporting verbal, nonverbal, and visual delivery while improving multichannel consistency.

## 6.2 RQ4. How will EFL students interact with and be influenced by the system?

Building on the mixed methods mentioned earlier, we begin by exploring how students' interactions with our system affect their AOP delivery, followed by an analysis of how the system impacts students' behaviors, emotions, and cognitive workloads.

**i) Usage pattern.** We share insights into how student presenters use the PC interface to customize delivery factors and improve presentation scripts, their interactions with the *Trinity* App during AOPs, and the motivations driving their actions.

**F7: Presenters generally customize delivery factors to suit their individual needs.** Out of 21 presenters, 17 tailored delivery factors to fit their preferences, reflecting diverse needs. Ten incorporated verbal, nonverbal, and visual factors, emphasizing the importance of content presentation and engagement. Four focused on verbal and nonverbal cues, aiming for a cohesive style where "body language plays a crucial role" (S14) and gestures "aligned with their spoken words" (S34). Three chose verbal and visual elements, believing visuals "aid in understanding" (S27) and make presentations "more engaging" (S44, S52). Four presenters used presets, possibly due to time or personal preference.

**F8: Presenters typically adapt their script length to match their allocated time and expected duration.** On the PC end, presenters adjust scripts to fit their time slots, and our PC tool predicts script durations to help. Of 21 presenters, 18 used this feature, with nine replacing text with diagrams, six cutting demos,

and three adding discussions. This ensured their presentations "met the required time limit" (S3, S27, S34) without "running out of time or having nothing left to say" (S14, S62).

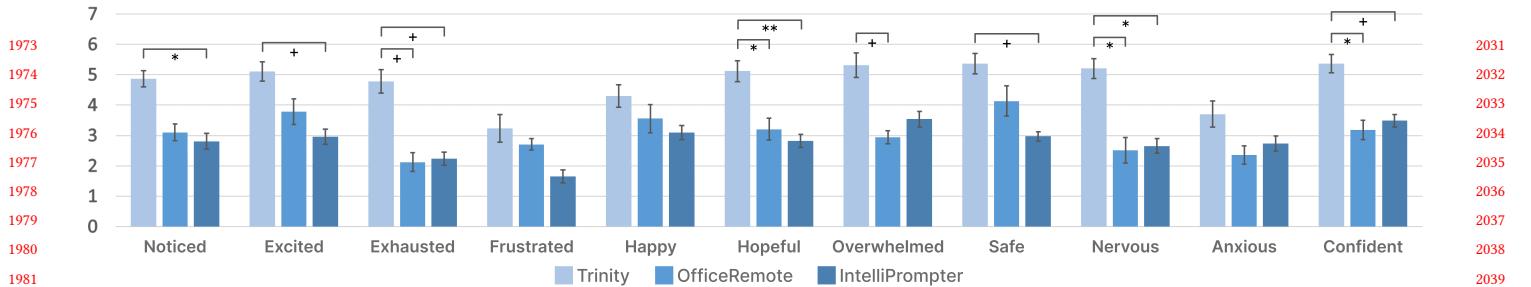
**F9: Presenters commonly annotate terminology but substitute embellished words in the polished script.** Presenters often simplify terminology in polished scripts to "connect better with their audience" (S23) and "ensure smooth communication" (S14). Six emphasized clearer pronunciation for easier concept "grasp" (S37, S39). Five noted personal preferences, using transitions for "more comfort and confidence" (S21, S37, S47), aiding a "natural flow" (S14). Revisions were driven by content confirmation (nine) and personal language alignment (seven).

**F10: Presenters manually scroll through the script when they experience speech recognition jitters.** Despite speech recognition's help in delivering speeches "without having to constantly refer back to physical scripts" (S17, S31), glitches can occur. Presenters then intervene by scrolling or "swiping the screen with the other hand" (S23, S39, S45). One reported a script that "suddenly jumped down" (S39), necessitating a scroll back to resume smoothly.

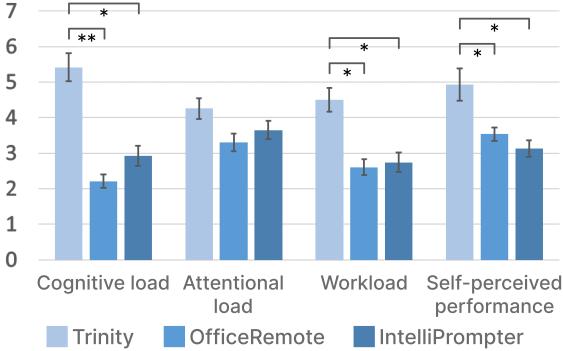
**ii) Effects on behavior and emotions.** We investigated how *Trinity* influenced both students' presentation behaviors and emotions, resulting in the following insights.

**F11: Emojis prompted presenters to adapt their delivery.** Emojis in presentations enhanced delivery and audience engagement in the *Trinity* condition. They helped presenters "effectively convey emotions" (S42) and "emphasize key points" (S21, S64). Emojis acted as visual cues for "how to deliver the content at different times" (S14), prompting appropriate facial expressions and gestures that improved "communication with the audience" (S28). This nonverbal layer made presentations "dynamic and engaging" (S37), aiding audience understanding and maintaining attention as they "followed along with both verbal and nonverbal cues" (S31).

**F13: Presenters had a more positive outlook on their presentation experience with *Trinity*.** Figure 17 shows that different presentation support systems significantly affected presenters' feelings of being noticed ( $H = 10.236, p = 0.022$ ), hopeful ( $H = 8.431, p = 0.034$ ), nervous ( $H = 11.237, p = 0.013$ ), and confident ( $H = 9.842, p = 0.019$ ). Post-hoc tests reveal that *Trinity*



**Figure 17: The effects on presenters' emotions and related feelings.** The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).



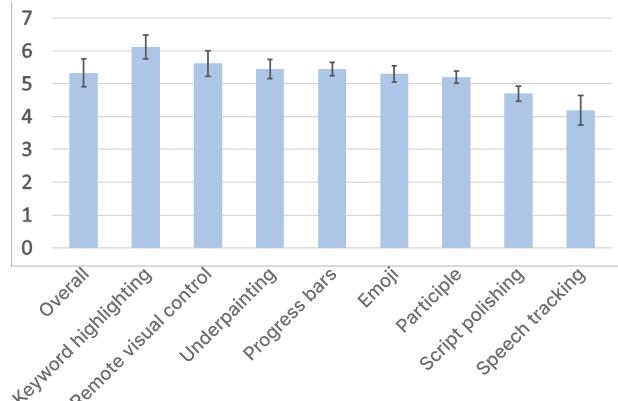
**Figure 18: The effects on presenters' cognitive load, attentional load, workload, and self-perceived performance.** The error bars indicate standard errors. (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ ).

made presenters feel more noticed than *IntelliPrompter* (all  $p < 0.1$ ) and potentially increased engagement compared to *OfficeRemote*. Presenters felt more hopeful using *Trinity* than *IntelliPrompter* ( $p = 0.003$ ) and *OfficeRemote* ( $p = 0.017$ ). Also, *Trinity* made presenters feel less nervous and more confident than *IntelliPrompter* (nervous:  $p = 0.015$ , confident:  $p = 0.540$ ) and *OfficeRemote* (nervous:  $p = 0.028$ , confident:  $p = 0.022$ ). No significant differences were found across the systems regarding feelings of excitement, exhaustion, frustration, happiness, being overwhelmed, or anxiety.

**F14: Presenters in-situ emotions during presentation can be influenced by the system's performance.** In retrospective reviews of the *Trinity* condition, most presenters (18 out of 21) saw on-the-fly delivery support as "a reliable backup for presentation" (S17) and "pillar to learn on" (S2). During smooth system operation, they felt confident and more engaged with their audience, "feeling more relaxed and perform better" (S34). However, technical glitches like an "unresponsive App" (S7, S42) led to "feeling flustered and frustrated" (S7).

**iii) Effects on workload, attentional load, and cognitive load.** Figure 18 displays average ratings of student presenters' workload, attentional load, and cognitive load. Kruskal-Wallis tests reveal significant differences in workload ( $H = 9.395, p = 0.010$ ), cognitive load ( $H = 12.365, p = 0.004$ ), and self-perceived performance ( $H = 11.545, p = 0.010$ ) across different support systems. Post-hoc tests indicate *Trinity* incurs significantly higher workload ( $M = 4.54, SD = 0.83, p = 0.025$  with *IntelliPrompter*,  $p = 0.034$  with *OfficeRemote*), cognitive load ( $M = 5.68, SD = 1.24, p = 0.036$  with *IntelliPrompter*,  $p = 0.007$  with *OfficeRemote*), and improves

self-perceived performance ( $M = 5.89, SD = 0.94, p = 0.021$  with *IntelliPrompter*,  $p = 0.014$  with *OfficeRemote*). No significant difference in attentional load implies similar attention allocation across conditions. Evidently, the baselines impose less workload and cognitive load as they only involve script reading without additional operations.



**Figure 19: Student presenters' trust levels (in 7-point Likert) on different system features.**

**F15: Presenters primarily concentrate on their scripts and briefly glance at emojis during presentations.** Presenters prioritized scripts for fluency and confidence, believing it helped them deliver messages effectively without "missing any important points" (S21). Many admitted to being "not fully prepared" (S33) and felt "more comfortable relying on a structured plan" (S55). In the *Trinity* condition, presenters still focused on scripts but "briefly glanced at the emojis" (S23, S37, S39, S64), which did not increase attentional load and seemed to enhance comprehension without "distracting from following along with the script" (S14).

**F17: Cognitive load increases when malfunctions or recognition jitters occur.** Cognitive load is the mental effort needed for information processing, which can rise during presentation issues. Problems like the "*Trinity App got stuck*" (S7, S42) require presenters to troubleshoot, "interrupting the train of thought" (S7, S42) and adding cognitive burden as they "understand what went wrong" (S7) and seek solutions, such as "check connectivity" (S7) or "seek technical assistance" (S42). Sudden topic shifts also increase cognitive load, as presenters need to "backtrack and find where they left off" (S7), requiring effort to "recall previous content" (S42) for a smooth transition.

**F18: Cognitive load diminishes as presenters become more acquainted with the system.** In the *Trinity* condition, presenters initially felt a cognitive burden, but this lessened with practice. They were instructed to preprocess at least half a day before, and four found that practicing with different content helped them "*better understand its functionalities*" (S21, S45, S52) and adapt their presentations. This familiarity "*reduced the cognitive load*" (S45) and "*improved the overall performance*" (S21).

### 6.3 RQ5. How will students trust and collaborate with our system?

Finally, as *Trinity* is a hybrid mobile-centric system offering on-the-fly delivery support, we aim to explore how students trust and engage with our system, as well as the factors influencing their trust and collaboration.

**i) Trust.** Figure 19 shows presenters' average ratings on a 7-point Likert scale regarding their trust in the system in the post-task questionnaire. We combine these ratings with qualitative insights from the interviews to present the following findings.

**F20: Students' inherent perceptions of LLMs influence their trust.** Most student presenters (16 out of 21) trust the system more when they know it uses a high-quality LLM. Initially, students may doubt the effectiveness of an unknown LLM, questioning if the output is "*reliable*" (S28) or if there "*could be potential issues*" (S39). However, once informed of an advanced LLM, they gain a "*sense of confidence and assurance*" (S47), trusting the system to be "*accurate and dependable*" (S28). This leads to increased trust and a stronger bond with the system. The use of an advanced LLM also signifies "*continuous improvements*" (S39), which students see as a "*testament to the reliability and credibility*" (S62) of the system.

**F21: System malfunctions could impair students' trust.** Five presenters worried about system malfunctions affecting their trust, noting it could lead to a "*significant decrease in trust*" (S7, S23, S39) and an "*unmanageable sense of crisis*" (S23, S42). Four shared experiences of voice tracking jitters, which made them "*lose confidence in the voice tracking*" (S23, S45) feature. To avoid issues, they manually "*scrolled through the script*" (S23, S39). This shows minor glitches can erode trust, as seen in Figure 19 where "*Speech Tracking*" had lower trust ratings.

**ii) Collaboration.** Beyond trust assessment, our study uncovered valuable insights into the collaborative dynamics between students and our system, gleaned from their open-ended feedback during the interviews.

**F22: The system could serve as a coach for students, offering guidance on their presentation content.** Of 21 student presenters, 16 found the system's guidance on presentation strategies valuable. One student (S3) noted, "*Trinity illuminated a clear path for me by refining the script and generating prompts for delivery, enabling me to enhance my ability to express myself effectively*." This indicates the system's role in improving language proficiency and communication skills. Additionally, students like (S34, S47) would "*study the polished script*" and (S17) "*identify expressions and sentence structures worth learning*." The script offered "*concrete examples*" (S28) of enhanced wording and expression within their "*own context*" (S40), suggesting the system's potential to aid in overall language acquisition.

**F23: Some students rely less on the system when presenting their own work.** Analysis from video replays showed that presenters reporting on their own work tended to refer to scripts less and improvise more: "*I'm way more familiar with the stuff I've worked on, so I don't stick to the script and just throw in my own improvisations*" (S21). This suggests that as they gain confidence in their material, they "*feel empowered*" (S31) to deviate from scripts, leading to a more natural and engaging presentation style (S27). Such improvisation "*tests the ability*" (S14) to think critically and adapt on the spot, fostering creativity and innovation beyond memorized material (S64).

## 7 DISCUSSION AND LIMITATION

### 7.1 Implications for Offering On-the-fly Support in Presentations

**Implication II: Emphasizing intuitive and attention-grabbing designs for on-the-fly support tools.** To ease presenters' cognitive load, *Trinity* offers user-friendly features like thumbnail navigation, larger fonts, emojis for clarity, and a darker underpainting to show progress. We bolded keywords for emphasis. These underscores the importance of helping presenters handle prompts without cluttering their presentations, enhancing delivery quality and emphasizing the need for intuitive, eye-catching real-time support tool designs.

**Implication III: Empowering presenters with autonomy over both content and presentation flow.** During our research's early stages, we considered creating a fully automated system for script and slide optimization. However, feedback from participants led us to reconsider autonomy in content and control. Our user study showed that many presenters want to review and refine their scripts. Thus, *Trinity* was designed for smooth transitions from manuscript to polished script, allowing for content revisions. Additionally, *Trinity* provides manual control via smartphone interactions like tapping and swiping, ensuring control autonomy. We believe future presentation tools should focus on empowering users with autonomy in content and control to build trust and improve the user experience.

**Implication VI: Strengthen system robustness to uphold user trust.** On-the-fly support systems like *Trinity* face challenges in maintaining user trust due to their fragile nature. Our study showed that presenters who experienced system malfunctions during presentations had a significant drop in trust. This underscores the importance of system robustness to prevent trust erosion. We advise designers to focus on thorough testing to fix potential issues, ensuring a reliable and long-lasting user-trust relationship.

### 7.2 Usability of LLMs from the Perception of Linguistic Expression

LLMs like ChatGPT have expanded the potential for real-time support during presentations, demonstrating language comprehension and generation skills, including unfamiliar word association, sentence denoising, and syntactic reuse [96]. Expert feedback and user studies have validated these capabilities, highlighting improved fluency and syntax in edited scripts. However, for formal and objective language, the polished scripts align with Muñoz et al.'s findings [59],

which is beneficial for academic presentations but requires further investigation for other genres like business talks.

When it comes to linguistic expression, there's a clear preference for verbal delivery prompts over nonverbal and visual ones, suggesting LLMs excel at verbal communication. The perception of linguistic expression also varies by the specific LLM, with GPT-3.5-turbo showing less proficiency in nonverbal and visual cues. For tasks involving linguistic expression, we recommend using LLMs with multi-modal training. Improving the consistency between LLM and human perceptions is expected to be a growing area of interest in both machine learning and HCI.

### 7.3 Presentation Delivery Learning Potential for EFL Students

Our focus is on the methodological aspect of presentation delivery learning for EFL students, diverging from traditional research that focuses on communication channel training before presentations. *Trinity* aims to empower EFL students to use various communication channels effectively in their presentations, potentially fostering the acquisition of multichannel presentation skills through experiential learning [48]. However, user study participants suggested that while delivery prompts are effective, their trust and understanding could be improved with explanations of the rationale behind them. This insight points towards a more effective approach to experiential delivery learning for EFL students, one that integrates abstract concepts with concrete experiences enhances experiential learning for EFL students. By reviewing video playbacks and adjusting delivery prompts after presentations, they can better understand the logic behind their performance. This approach helps them build a personalized yet generalizable model of presentation logic.

### 7.4 Towards a Comprehensive Ecosystem for Presentation Support

Students experience multiple stages in presentations, and *Trinity* supports them from preparation to reflection. While it enhances delivery, concerns exist about potential overreliance on its real-time support without structured training. To mitigate this, we suggest *Trinity* be integrated into a comprehensive presentation support system. It could complement tools like *Slide4N* [88] and *HyperSlides* [27] by optimizing slides and scripts, reducing the need for extensive tutorials. This allows students to incorporate delivery patterns into their scripts and compare with *Trinity*'s prompts for personalized learning. Additionally, feedback mechanisms such as *SlideSpecs* [90] and *AutoManner* [80] could provide audience feedback for performance reflection, completing the experiential learning cycle. By becoming part of a larger ecosystem, *Trinity* encourages consideration of how tools fit into broader systems and promotes open designs for seamless integration with existing practices.

### 7.5 Limitations and Future Work

**Towards Better Hardware Support.** In our design study, storyboards evaluated design concepts, with the mobile and PC combination standing out. Chosen for their common use and minimal

visual distraction, this pairing is not the ultimate solution for on-the-fly delivery support. It must meet three criteria: providing clear support without disrupting presenters' communication, remaining unnoticed by the audience, and being portable. Current media may not fully satisfy these needs. Our user study indicates that mobile devices could still limit presenters' nonverbal cues. Thus, we see the mobile and PC combination as progress towards better hardware support. Future research should seek media that can meet all three requirements.

**Adapt LLMs to Personal Contexts.** *Trinity* uses LLMs to enhance script quality and create delivery cues. Most users in our study appreciated the improvements in wording and syntax, but some faced challenges with unfamiliar language and mismatched preferences. To remedy this, *Trinity* allows users to edit the script freely. Yet, this might not be ideal for all EFL students, who may lack the willingness or patience to review thoroughly. A possible solution is LLM few-shot prompting, where a small set of relevant examples is provided to tailor the model's output. This could involve uploading past scripts as references, similar to custom GPT features. However, it's unclear if this approach can fully adapt the LLM to users' vocabulary and style. Another option is fine-tuning the LLM, which, despite being resource-intensive and potentially causing knowledge loss, could be more effective. Beyond script development, LLMs could also help users refine their presentation delivery styles, drawing inspiration from figures like Obama. Future research could aim to find more efficient ways to adapt LLMs to individual needs, enhancing personalization.

**Technical Stability and Extendability.** The *Trinity* App, which uses real-time speech recognition to track presentations, faced issues with jitters and delays, affecting smooth speech tracking. Future studies could look into content-based navigation as a solution [16, 62]. Additionally, the app's slide thumbnail feature struggled with complex animations not fully visible due to PowerPoint interface constraints. Collaboration with Microsoft to integrate better animation support is suggested. Lastly, *Trinity* enhances scripts based on user manuscripts but found that the manuscript's semantic richness and integrity greatly affect the presentation's perceived "content & organization." Improving this could involve cross-referencing slides with scripts to optimize content and enhance underdeveloped manuscripts.

## 8 CONCLUSION

In this study, we introduced *Trinity*, a hybrid system designed to synchronize verbal, nonverbal, and visual elements in academic oral presentations (AOPs). Developed through a formative study with a survey, design research, and expert interviews, *Trinity* features a PC PowerPoint add-in for script enhancement and delivery with GPT-4, alongside an Android App for remote visual control, speech pacing, and delivery cues. User studies showed that *Trinity* surpassed two baseline systems. The research deepened our understanding of AOP factors, provided design insights, and highlighted the system's impact on presentation practices. The experiences and challenges in *Trinity*'s design offer key lessons for future presentation support tool development. We see *Trinity* as an innovative step towards future presentation practices, potentially leading to more advanced human-AI collaboration in educational settings.

## REFERENCES

- [1] Darko Martinovikj Nevena Ackovska. 2013. Gesture recognition solution for presentation control. In *Proc. of 10th Conference for Informatics and Information Technology*.
- [2] Olwyn Alexander, Sue Argent, and Jenifer Spencer. 2008. *EAP essentials: A teacher's guide to principles and practice*. Garnet Education.
- [3] Nur Lina Amalia and Nadiah Ma'mun. 2020. The anxiety of EFL students in presentation. *ELITE JOURNAL* 2, 1 (2020), 65–84.
- [4] Michael Argyle, Florisse Alkema, and Robin Gilmour. 1971. The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European Journal of Social Psychology* 1, 3 (1971), 385–402.
- [5] Reza Asadi, Ha Trinh, Harriet J Fell, and Timothy W Bickmore. 2017. IntelliPrompter: speech-based dynamic note display interface for oral presentations. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 172–180.
- [6] Elizabeth E Austin and Naomi Sweller. 2014. Presentation and production: The role of gesture in spatial communication. *Journal of experimental child psychology* 122 (2014), 92–103.
- [7] Steven A Beebe. 1976. Effects of Eye Contact, Posture and Vocal Inflection upon Credibility and Comprehension. (1976).
- [8] Laurence Bich-Carriére. 2019. Say it with [a smiling face with smiling eyes]: judicial use and legal challenges with emoji interpretation in Canada. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique* 32, 2 (2019), 283–319.
- [9] Timothy Bickmore, Everlyne Kimani, Ameneh Shamekh, Prasanth Murali, Dhaval Parmar, and Ha Trinh. 2021. Virtual agents as supporting media for scientific presentations. *Journal on Multimodal User Interfaces* 15 (2021), 131–146.
- [10] Philip E Bourne. 2007. Ten simple rules for making good oral presentations. , e77 pages.
- [11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [12] Berndt Brehmer. 1976. Social judgment theory and the analysis of interpersonal conflict. *Psychological bulletin* 83, 6 (1976), 985.
- [13] Mark Bubel, Ruiwen Jiang, Christine H Lee, Wen Shi, and Audrey Tse. 2016. AwareMe: addressing fear of public speech through awareness. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, 68–73.
- [14] Xiang Cao, Eyal Ofek, and David Vronay. 2005. Evaluation of alternative presentation control techniques. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 1248–1251.
- [15] Matt Carter. 2012. *Designing science presentations: A visual guide to figures, papers, slides, posters, and more*. Academic Press.
- [16] Minsoo Chang, Mina Huh, and Juho Kim. 2021. Rubyslippers: Supporting content-based voice navigation for how-to videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14.
- [17] Xinyue Chen, Shuo Li, Shipeng Liu, Robin Fowler, and Xu Wang. 2023. MeetScript: Designing Transcript-based Interactions to Support Active Participation in Group Video Meetings. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–32.
- [18] Yin Ling Cheung. 2008. Teaching effective presentation skills to ESL/EFL students. *The Internet TESL Journal* 14, 6 (2008), 1–2.
- [19] Pei-Yu Chi, Bongshin Lee, and Steven M Drucker. 2014. DemoWiz: re-performing software demonstrations for a live presentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1581–1590.
- [20] Barbara Chivers and Michael Shoolbred. 2007. *A student's guide to presentations: Making your presentation count*. Sage.
- [21] Ionut Damian, Chiew Seng Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*, 565–574.
- [22] Luc De Grez, Martin Valcke, and Irene Rooden. 2014. The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. *Higher Education Research & Development* 33, 2 (2014), 256–271.
- [23] Alan R Dennis and Susan T Kinney. 1998. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research* 9, 3 (1998), 256–274.
- [24] Robert Dolan. 2017. Effective presentation skills. *FEMS microbiology letters* 364, 24 (2017), fmx235.
- [25] Nancy Duarte. 2008. *Slide: ology: The art and science of creating great presentations*. Vol. 1. O'Reilly Media Sebastopol.
- [26] Patricia A Duff. 2007. Second language socialization as sociocultural theory: Insights and issues. *Language teaching* 40, 4 (2007), 309–319.
- [27] Darren Edge, Joan Savage, and Koji Yatani. 2013. HyperSlides: dynamic presentation prototyping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 671–680.
- [28] Darren Edge, Xi Yang, Yasmine Kotturi, Shuoping Wang, Dan Feng, Bongshin Lee, and Steven Drucker. 2016. Slidespace: Heuristic design of a hybrid presentation medium. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 3 (2016), 1–30.
- [29] Ayman Hassan Abu El Enein. 2011. Difficulties encountering English majors in giving academic oral presentations during class at Al-Aqsa University. *Unpublished Master's thesis, Islamic University of Gaza* (2011).
- [30] Adrian Furnham, Robert Trevelyan, and George Gaskell. 1981. The relative contribution of verbal, vocal, and visual channels to person perception: Experiment and critique. (1981).
- [31] Howard Giles, Tania Ogay, et al. 2007. Communication accommodation theory. (2007).
- [32] Randall L Gillis and Elizabeth S Nilsen. 2017. Consistency between verbal and non-verbal affective cues: A clue to speaker credibility. *Cognition and Emotion* 31, 4 (2017), 645–656.
- [33] George M Glasgow. 1952. A semantic index of vocal pitch. *Communications Monographs* 19, 1 (1952), 64–68.
- [34] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, Vol. 52. Elsevier, 139–183.
- [35] Brian S Hentz. 2006. Enhancing presentation narratives through written and visual integration. *Business Communication Quarterly* 69, 4 (2006), 425–429.
- [36] Brigitte Hertz. 2015. *Spotlight on the presenter: a study into presentations of conference papers*. Ph.D. Dissertation, Wageningen University and Research.
- [37] Rebecca Hincks. 2005. Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System* 33, 4 (2005), 575–591.
- [38] Bill Hoogterp. 2014. *Your Perfect Presentation: Speak in Front of Any Audience Anytime Anywhere and Never Be Nervous Again*. McGraw Hill Professional.
- [39] Laszlo Hunyadi and István Szekrényes. 2020. *The Temporal Structure of Multimodal Communication*. Springer.
- [40] Ikhfi Imaniah. 2018. The students' difficulties in presenting the academic speaking presentation. *Globish (An English-Indonesian Journal for English, Education and Culture)* (2018).
- [41] David W Johnson. 1971. Role reversal: A summary and review of the research. *International Journal of Group Tensions* (1971).
- [42] Tero Jokela, Jaakko T Lehtinen, and Hannu Korhonen. 2008. Mobile multimedia presentation editor: enabling creation of audio-visual stories on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 63–72.
- [43] Kuldip Kaur and Afida Mohamad Ali. 2018. Exploring the genre of academic oral presentations: A critical review. *International Journal of Applied Linguistics and English Literature* 7, 1 (2018), 152–162.
- [44] Ken Kelch. 1985. Modified input as an aid to comprehension. *Studies in second language acquisition* 7, 1 (1985), 81–90.
- [45] Amirsam Khataei and Ali Arya. 2014. Personalized presentation builder. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 2293–2298.
- [46] Minju Kim and Kwangyun Wohm. 2018. HoloBox: Augmented visualization and presentation with spatially integrated presenter. *Interacting with Computers* 30, 3 (2018), 224–242.
- [47] Jane King. 2002. Preparing EFL Learners for Oral Presentations. *Dong Hwa Journal of Humanistic Studies* 4 (2002), 401–418.
- [48] David A Kolb. 2014. *Experiential learning: Experience as the source of learning and development*. FT press.
- [49] Kazutaka Kurihara, Masataka Goto, Jun Ogata, Yosuke Matsusaka, and Takeo Igarashi. 2007. Presentation sensei: a presentation training system using speech and image processing. In *Proceedings of the 9th international conference on Multimodal interfaces*, 358–365.
- [50] Timothy R Levine. 2014. Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology* 33, 4 (2014), 378–392.
- [51] Peter Levrai and Averil Bolster. 2015. Developing a closer understanding of academic oral presentations. *Folio* 16, 2 (2015), 65–72.
- [52] Richen Liu, Min Gao, Shunlong Ye, and Jiang Zhang. 2021. IGScript: An interaction grammar for scientific data presentation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13.
- [53] Shao-Kang Lo. 2008. The nonverbal communication functions of emoticons in computer-mediated communication. *Cyberpsychology & behavior* 11, 5 (2008), 595–597.
- [54] Steven Lukes. 2021. *Power: A radical view*. Bloomsbury Publishing.
- [55] Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An adaptable system for instructors to grasp student learning status in synchronous online classes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–25.
- [56] Albert Mehrabian. 1968. Relationship of attitude to seated posture, orientation, and distance. *Journal of personality and social psychology* 10, 1 (1968), 26.

- [57] Microsoft. [n. d.]. Start the presentation and see your notes in Presenter view. <https://support.microsoft.com/en-us/office/start-the-presentation-and-see-your-notes-in-presenter-view-4de90e28-487e-435c-9401-eb49a3801257> (2023, Aug 12).
- [58] Naoko Morita. 2000. Discourse socialization through oral classroom activities in a TESL graduate program. *Tesol Quarterly* 34, 2 (2000), 279–310.
- [59] Alberto Muñoz-Ortiz, Carlos Gómez-Rodríguez, and David Vilares. 2023. Contrasting Linguistic Patterns in Human and LLM-Generated Text. *arXiv preprint arXiv:2308.09067* (2023).
- [60] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [61] Prasanth Murali, Lazlo Ring, Ha Trinh, Reza Asadi, and Timothy Bickmore. 2018. Speaker hand-offs in collaborative human-agent oral presentations. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 153–158.
- [62] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1–7.
- [63] Jeungmin Oh, Darren Edge, and Uichin Lee. 2020. ScriptFree: Designing Speech Preparation Systems with Adaptive Visual Reliance Control on Script. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [64] Brian Paltridge and Sue Starfield. 2013. *The handbook of English for specific purposes*. Vol. 592. Wiley Online Library.
- [65] Dhaval Parmar and Timothy Bickmore. 2020. Making it personal: Addressing individual audience members in oral presentations using augmented reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–22.
- [66] Michael J Peeters, Eric G Sahloff, and Gregory E Stone. 2010. A standardized rubric to evaluate student presentations. *American journal of pharmaceutical education* 74, 9 (2010).
- [67] Yi-Hao Peng, JiWoon Jang, Jeffrey P Bigham, and Amy Pavel. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [68] Larissa Pschetz, Koji Yatani, and Darren Edge. 2014. TurningPoint: narrative-driven presentation planning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1591–1594.
- [69] J Ram Rajesh, R Sudharshan, D Nagarjunan, and R Aarthi. 2012. Remotely controlled PowerPoint presentation navigation using hand gestures. In *Proc. of International Conference on Advances in Computer, Electronics and Electrical Engineering*.
- [70] Nurul Amilin Razawi, Luqmanul Hakim Zulkornain, and Razifa Mohd Razlan. 2019. Anxiety in oral presentations among ESL students. *Journal of Academia* 7, 1 (2019), 31–36.
- [71] Monica A Riordan. 2017. The communicative role of non-face emojis: Affect and disambiguation. *Computers in Human Behavior* 76 (2017), 75–86.
- [72] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Found. Trends Inf. Retr.* 3, 4 (apr 2009), 333–389. <https://doi.org/10.1561/1500000019>
- [73] Reinout Roels and Beat Signer. 2014. MindXpres: An extensible content-driven cross-media presentation platform. In *Web Information Systems Engineering—WISE 2014: 15th International Conference, Thessaloniki, Greece, October 12–14, 2014, Proceedings, Part II* 15. Springer, 215–230.
- [74] Sam Sabri. [n. d.]. Microsoft releases Office Remote to allow you to control your PowerPoint presentation and more from your Windows Phone. <https://www.windowscentral.com/office-remote-windows-phone> (2023, Aug 30).
- [75] Jan Schneider, Dirk Börner, Peter Van Rosmalen, and Marcus Specht. 2015. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*. 539–546.
- [76] Atsushi Senju and Mark H Johnson. 2009. The eye contact effect: mechanisms and development. *Trends in cognitive sciences* 13, 3 (2009), 127–134.
- [77] Diane J Skiba. 2016. Face with tears of joy is word of the year: are emoji a sign of things to come in health care? *Nursing education perspectives* 37, 1 (2016), 56–57.
- [78] Eva Strangert and Joakim Gustafson. 2008. What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations. In *Ninth Annual Conference of the International Speech Communication Association*.
- [79] M Iftekhar Tanveer, Emi Lin, and Mohammed Hoque. 2015. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th international conference on intelligent user interfaces*. 286–295.
- [80] M Iftekhar Tanveer, Ru Zhao, Kezhen Chen, Zoe Tiet, and Mohammed Ehsan Hoque. 2016. Automanner: An automated interface for making public speakers aware of their mannerisms. In *Proceedings of the 21st international conference on intelligent user interfaces*. 385–396.
- [81] Hashmatullah Tareen. 2022. Investigating EFL learners' perceptions towards the difficulties in oral presentation at Kandahar university. *ESI Preprints* 11 (2022), 535–535.
- [82] Santawat Thanyadit, Matthias Heintz, and Effie LC Law. 2023. Tutor In-sight: Guiding and Visualizing Students' Attention with Mixed Reality Avatar Presentation Tools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.
- [83] Dominic Thompson, Ian G Mackenzie, Hartmut Leuthold, and Ruth Filik. 2016. Emotional responses to irony and emoticons in written language: Evidence from EDA and facial EMG. *Psychophysiology* 53, 7 (2016), 1054–1062.
- [84] Ha Trinh, Lazlo Ring, and Timothy Bickmore. 2015. Dynamicduo: co-presenting with virtual agents. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1739–1748.
- [85] Ha Trinh, Koji Yatani, and Darren Edge. 2014. PitchPerfect: integrated rehearsal environment for structured presentation preparation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1571–1580.
- [86] Khai N Truong, Gillian R Hayes, and Gregory D Abowd. 2006. Storyboarding: an empirical determination of best practices and effective guidelines. In *Proceedings of the 6th conference on Designing Interactive systems*. 12–21.
- [87] Wan Nurul Fazliza Wan Zakaria and Siti Shazlin Razak. 2016. English as a Second Language (ESL) Learner's Perceptions of the Difficulties in Oral Commentary Assessment. *Journal of Contemporary Social Science Research* 1, 1 (2016), 1–15.
- [88] Fengjie Wang, Xuye Liu, Oujing Liu, Ali Neshati, Tengfei Ma, Min Zhu, and Jian Zhao. 2023. Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [89] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. 2020. Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [90] Jeremy Warner, Amy Pavel, Tonya Nguyen, Maneesh Agrawala, and Björn Hartmann. 2023. SlideSpecs: Automatic and Interactive Presentation Feedback Collation. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 695–709.
- [91] Marcus Kho Gee Whai and Leong Lai Mei. 2015. Causes of academic oral presentation difficulties faced by students at a polytechnic in Sarawak. *The English Teacher* 44, 3 (2015).
- [92] Chauncey Wilson. 2013. *Brainstorming and beyond: a user-centered design method*. Newnes.
- [93] Torsten Wörtwein, Mathieu Chollet, Boris Schauerte, Louis-Philippe Morency, Rainier Stiefelhagen, and Stefan Scherer. 2015. Multimodal public speaking performance assessment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 43–50.
- [94] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–22.
- [95] Meng Xia, Qian Zhu, Xingbo Wang, Fei Nie, Huamin Qu, and Xiaojuan Ma. 2022. Persua: A visual interactive system to enhance the persuasiveness of arguments in online discussion. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–30.
- [96] Qihui Xu, Yingying Peng, Minghua Wu, Feng Xiao, Martin Chodorow, and Ping Li. 2023. Does Conceptual Representation Require Embodiment? Insights From Large Language Models. *arXiv preprint arXiv:2305.19103* (2023).
- [97] Shengzhou Yi, Hiroshi Yumoto, Xuetong Wang, and Toshihiko Yamasaki. 2020. Presentationtrainer: Oral presentation support system for impression-related feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 13644–13645.
- [98] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
- [99] Sandra Carolina Zappa-Hollman. 2007. Academic presentations across post-secondary contexts: The discourse socialization of non-native English speakers. *Canadian Modern Language Review* 63, 4 (2007), 455–485.
- [100] Alla Zareva. 2011. 'And so that was it': Linking adverbials in student academic presentations. *RELC Journal* 42, 1 (2011), 5–15.
- [101] Haipeng Zeng, Xingbo Wang, Yong Wang, Aoyu Wu, Ting-Chuen Pong, and Huamin Qu. 2022. GestureLens: Visual analysis of gestures in presentation videos. *IEEE Transactions on Visualization and Computer Graphics* (2022).

## Appendix

## A DETAILS OF EXPLORATORY SURVEY RESULTS

We analyzed the quantitative data with descriptive statistics, and two authors coded the responses to open-ended questions using thematic analysis methods [11]. Subsequently, the two authors engaged in collaborative discussions to synthesize their findings.

In the **survey conducted among EFL students**, a total of 52 questionnaires were distributed, yielding 49 valid responses. The respondents were diverse, including 28 males, 20 females, and 1 prefer not to disclose. The participants, with an average age of 23.4 ( $SD = 3.0$ ), represented various academic disciplines such as computer science, electronic engineering, biology, finance, statistics, forestry, and medicine. They spanned different grade levels and had experience with AOPs, with 15% having presented once, 20% twice, 40% three times, 15% four times, and 10% more than four times. Regarding preparation time for AOPs, 20.4% reported spending less than three days, 57.1% between three days and one week, 16.3% between one and two weeks, and 6.1% more than two weeks. In terms of challenges, 65.7% of students found AOPs to be difficult, and 42.9% expressed that their preparation time was insufficient. In terms of mentality, the majority of EFL students experienced varying feelings towards AOPs: 16.3% felt fear, 36.7% felt anxious, 18.4% were uninterested, 18.4% were confident, and 10.2% were excited. The issues most commonly identified during AOPs were lack of fluency (36.7%), flat intonation (28.6%), and absence of body language (12.2%). Notably, fluency was regarded as more crucial than pronunciation and grammar. Lexical gaps were identified as a challenge, leading to difficulties in word retrieval and conveying intended meaning effectively. Consequently, this posed a formidable obstacle for them to express themselves spontaneously and seamlessly, similar to native speakers.

In terms of supportive tools, the majority of participants (67.3%) had experience utilizing such tools to facilitate their AOPs. Among tool users, a significant majority (54.5%) employed these tools frequently. Impressively, the user experience with these supportive tools was generally positive, with 45.6% reporting a good experience, 51.5% describing it as mediocre, and only 3% indicating a negative experience. Participants who had suboptimal experiences highlighted challenges in understanding provided hints, locating required script snippets, excessive focus on the tools, and potential overdependence. Furthermore, the participants attributed several benefits to these tools, including facilitating script composition (57.1%), enhancing slide presentations (22.4%), boosting their confidence and mentality (20.4%), and aiding in controlling speech rate (18.4%). Interestingly, a significant majority (81.6%) perceived the necessity of supportive tools for AOPs. When asked about their hesitation in using such tools, cost of learning and instructional prerequisites emerged as the predominant concerns. Regarding anticipated features, participants expressed interest in all provided factors, with particularly prominent preferences for script display (73.5%), slide previews (61.2%), and support for controlling speech rate (42.9%). This underscores the value participants place on tools that assist in the delivery of effective presentations.

Turning to the **survey of instructors**, we received a total of 36 responses from instructors, with 24 males and 12 females, showcasing an average age of 45.3 ( $SD = 12.0$ ). This group of instructors

represented a variety of academic fields, including information science, physics, entrepreneurship and management, mathematics, and humanities. Their teaching experiences included organizing AOPs for EFL students throughout their careers, with 15.5% having done so once, 22.5% twice, 39.4% three times, 14.1% four times, and 8.5% more than four times. Interestingly, 83.3% of instructors expressed dissatisfaction with the AOP delivery of EFL students. Among all instructors, 44.4% felt weary, 38.9% were worried, 11.1% showed interest, and 5.6% were satisfied. The instructors identified the most common challenges that students face, namely the lack of preparation (55.6%) and poor delivery skills (30.6%). Additionally, a majority of instructors (83.3%) believed that incorporating delivery support tools into their courses wouldn't impact grading assessment. Among these instructors, 38.9% were open to allowing students to use such tools, and 30.6% supported encouraging students to utilize them. In response to queries about their reasons, most of them mentioned "*feeling awkward towards students' poor delivery*" and "*recognizing the benefits of improved delivery for both speakers and the audience*". On the other hand, a closer analysis of open-ended responses revealed that the remaining 30.6% of instructors maintaining a negative perspective were primarily concerned that these tools might hinder audience reception and foster excessive dependence among students.

## B DETAILS OF EXPERT INTERVIEW

LLMs, like ChatGPT, have opened up significant opportunities for human-AI collaboration and demonstrated impressive capabilities [94]. To assess the viability of employing ChatGPT for refining presentation scripts and offering delivery prompts, we conducted semi-structured interviews with five experienced experts (E1-5) in presentation instruction. These interviews aimed to tap into their expertise and gather insights on the feasibility of leveraging ChatGPT in this context. Additionally, we sought their professional perspectives on the outcomes of our initial exploratory surveys and design workshop, enriching our understanding of the potential applications and implications of the technology.

**Participants.** The individuals we interviewed comprise five experts (2 males and 3 females) affiliated with the language teaching group at a local university. Their collective background includes substantial experience in presentation instruction and proficiency in delivering presentations themselves. All these experts possess substantial theoretical knowledge and a wealth of practical experience in the domain of presentations. Their collective expertise (Figure 7A) spans a range of 2 to 7 years, with an average duration of 4 years ( $SD = 2.78$ ). Notably, all of these experts have taught presentation courses to EFL students across multiple semesters ( $Mean = 6.30$ ,  $SD = 3.01$ ).

**Preparation.** We selected two prominent LLMs, GPT-3.5-turbo and GPT-4 (both March 14 versions), as our candidates for evaluation. Our process commenced by collecting 10 practical presentation scripts from EFL students through social networks. These scripts covered a range of topics including HCI, machine learning, sociology, etc. To devise effective prompts for ChatGPT that would yield desired outputs, we initiated the process with a preliminary draft. We iteratively refined this draft by interacting with both GPT-3.5-turbo and GPT-4, ensuring that the generated content and output

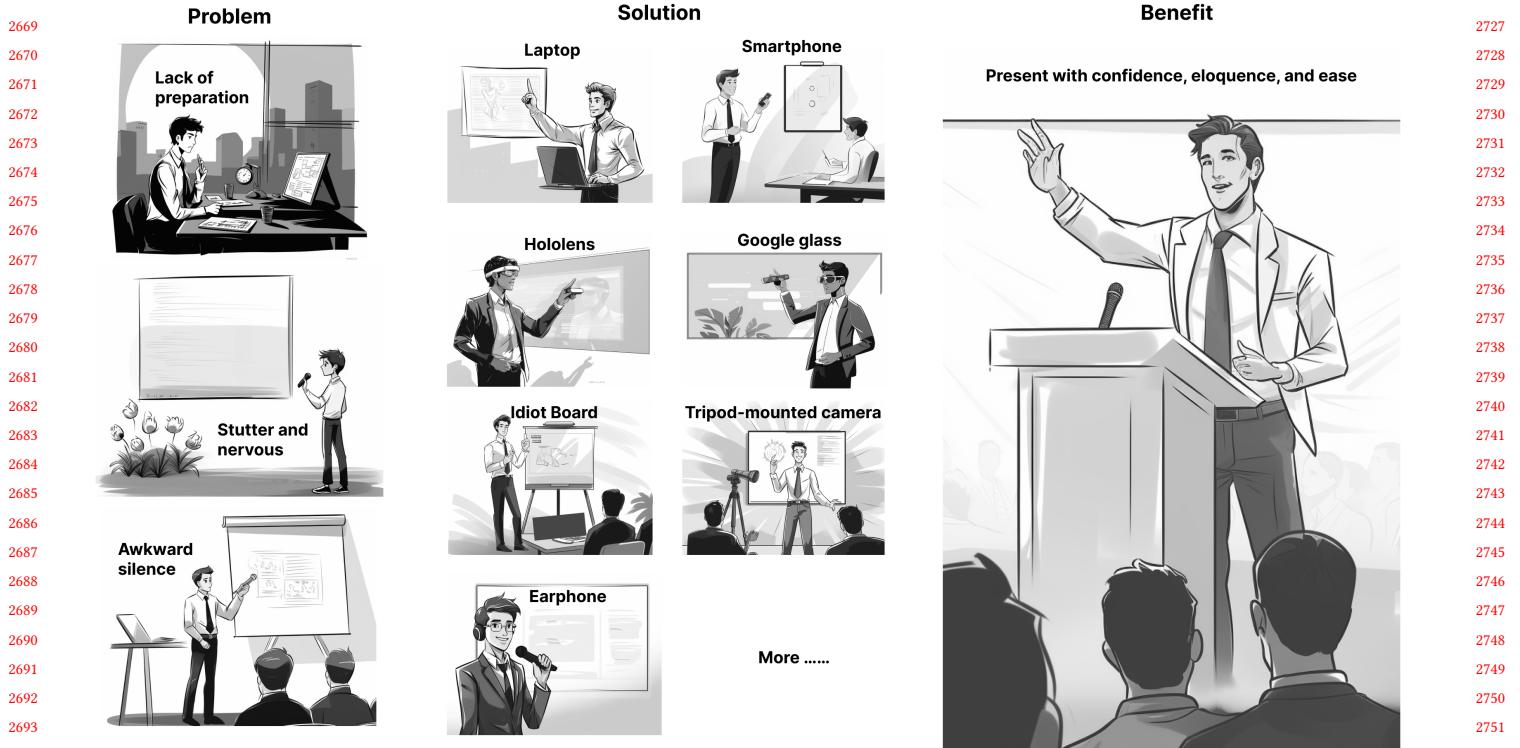


Figure 20: Summarized version of the Storyboards used in our design workshop.

format aligned with our expectations. Following the completion of this iterative process, we employed the final GPT prompts on both GPT-3.5-turbo and GPT-4 to further polish all the scripts. Additionally, in-line delivery prompts were incorporated into the scripts to enhance their presentation quality and effectiveness.

**Interview Protocol.** Each interview session had a duration of approximately 40 minutes. Our interview process involved formulating questions based on 7-point Likert scale to evaluate the quality of output scripts and delivery prompts from LLM candidates (as depicted in Figure 7B). These questions were tailored for individual expert interviews. Simultaneously, open-ended questions were crafted to capture insights into their presentation instruction experiences, suggestions on addressing users' needs, and their perceptions of the outcomes from our exploratory survey and design workshop. For example, questions like “*What strategies do you typically employ to enhance the effectiveness of presentations?*”; “*what are your suggestions if a LLM-based tool were designed to aid EFL students in improving their presentation delivery on the fly?*”; and “*Did any particular findings from the survey results stand out to you?*” Throughout the interviews, we also asked the experts follow-up questions according to their responses and shared specific instances of challenges encountered by EFL students in their presentation deliveries.

**Results and Analysis.** The semi-structured interviews with the expert panel yielded valuable insights regarding the potential of using LLMs, specifically GPT-3.5-turbo and GPT-4, for improving presentation scripts and providing delivery prompts. First, in terms of the quality of output scripts generated by the LLMs, the experts generally provided positive ratings (Figure 7B). Specifically,

the experts reached a consensus that both LLMs excelled in terms of structural completeness and keyword identification. However, it was noted by E1 and E3 that GPT-3.5-turbo tended to employ overly complex wording and sentence structures, rendering the generated script less suitable for practical uses. In terms of fluency and maintaining meaning, ratings indicated a generally favorable reception for both LLM candidates, with GPT-4 slightly outperforming GPT-3.5-turbo according to the experts' assessments. Qualitative feedback highlighted the strengths of these LLMs in enhancing the clarity, coherence, and overall organization of presentation scripts. E2 illustrated that LLMs have “*demonstrated an impressive capability in enhancing presentation script.*” Second, the effectiveness of the in-line delivery prompts provided by the LLMs also received positive ratings from the experts. They generally affirmed the efficacy of GPT-4 in factor judgment and modulation during presentations. In general, the experts agreed that GPT-4 outperformed GPT-3.5-turbo across all metrics. However, concerns were raised regarding the density of these delivery prompts. E4 commented that “*modulation techniques in presentation training usually focus on details at the phrase level or word level. While these (delivery) prompts are reasonable, I can imagine that such dense prompts might be overwhelming for students if you intend to provide on-the-fly support.*” Conversely, open-ended responses from the experts indicated that these prompts could assist students in maintaining a smooth flow of speech, adhering to the topic, and effectively engaging the audience. For instance, E3 mentioned that “*these prompts could serve as a delivery template for students, allowing them to learn from their own presentation scripts.*”

2669  
2670  
2671  
2672  
2673  
2674  
2675  
2676  
2677  
2678  
2679  
2680  
2681  
2682  
2683  
2684  
2685  
2686  
2687  
2688  
2689  
2690  
2691  
2692  
2693  
2694  
2695  
2696  
2697  
2698  
2699  
2700  
2701  
2702  
2703  
2704  
2705  
2706  
2707  
2708  
2709  
2710  
2711  
2712  
2713  
2714  
2715  
2716  
2717  
2718  
2719  
2720  
2721  
2722  
2723  
2724  
2725  
2726  
2727  
2728  
2729  
2730  
2731  
2732  
2733  
2734  
2735  
2736  
2737  
2738  
2739  
2740  
2741  
2742  
2743  
2744  
2745  
2746  
2747  
2748  
2749  
2750  
2751  
2752  
2753  
2754  
2755  
2756  
2757  
2758  
2759  
2760  
2761  
2762  
2763  
2764  
2765  
2766  
2767  
2768  
2769  
2770  
2771  
2772  
2773  
2774  
2775  
2776  
2777  
2778  
2779  
2780  
2781  
2782  
2783  
2784

Furthermore, during the interviews, our experts shared invaluable insights into their teaching experiences, shedding light on various strategies employed to enhance presentations effectiveness. These strategies included techniques for maintaining eye contact, using body language, and managing nervousness, among others. For instance, E2 mentioned that “*simply encouraging individuals to keep their lower body motionless can significantly enhance their stage presence.*” Additionally, the experts provided insightful suggestions for the design of an LLM-based tool to support EFL students in real-time presentation delivery. One innovative approach they proposed involved integrating multimodal feedback to effectively guide students’ attention and behavior. To build upon this idea, E1 noted “*using subtle auditory cues, like a beep or pitch, can signal students to adjust their speech volume or speed. Moreover, incorporating somatosensory cues, such as slight vibrations synchronized with key parts of the speech, can provide haptic feedback when emphasizing specific points.*” Building upon this concept, E3 emphasized the potential to incorporate visual cues such as arrows or highlights into the tool, acting as a navigational guide to ensure students focus on relevant slides or audience members.

Regarding their perspectives on the results obtained from our exploratory surveys and design workshop, the experts highlighted several noteworthy findings. These included self-reported presentation challenges among EFL students, preferences for presentation aid features, and reactions to prototype design concepts. For example, the experts recognized that the discrepancy between verbal and non-verbal communication challenges may be attributed to differences in EFL students’ exposure and practice. As E4 pointed out, “*EFL students tend to acquire and practice non-verbal communication skills through daily experiences like watching videos, playing games, or engaging in interpersonal interactions. However, opportunities to enhance their verbal communication skills, particularly in academic or professional contexts, are more limited, which is where they often face the greatest difficulties.*” E5 also noted, “*Especially for them (EFL students), the pronunciation of unfamiliar or polysyllabic words can pose significant challenges during presentations.*”

The feedback provided by the experts on these findings will play a pivotal role in refining the development of our tool, ensuring it aligns more closely with user needs. To sum up, our expert interviews have contributed valuable insights into the application of LLMs for presentation enhancement and delivery support. The positive ratings and feedback regarding the quality of output scripts and delivery prompts from both LLMs, especially GPT-4, suggest promising potential for integrating LLMs into practical settings to improve presentation delivery.

## C MINOR FINDINGS FROM USER STUDY

**Minor F1:** Presenters verify the polished script against their original manuscript. Trinity generates a polished script from the presenters’ manuscript. We noticed that 10 out of 11 presenters switched between the manuscript and the polished script for confirmation, with 8 out of 11 presenters reviewing the entire polished script:

“*GPT is indeed powerful, but after all, it can’t guarantee to fully align with your intentions, so I definitely need to double-check.*”  
(S2, male, age: 21)

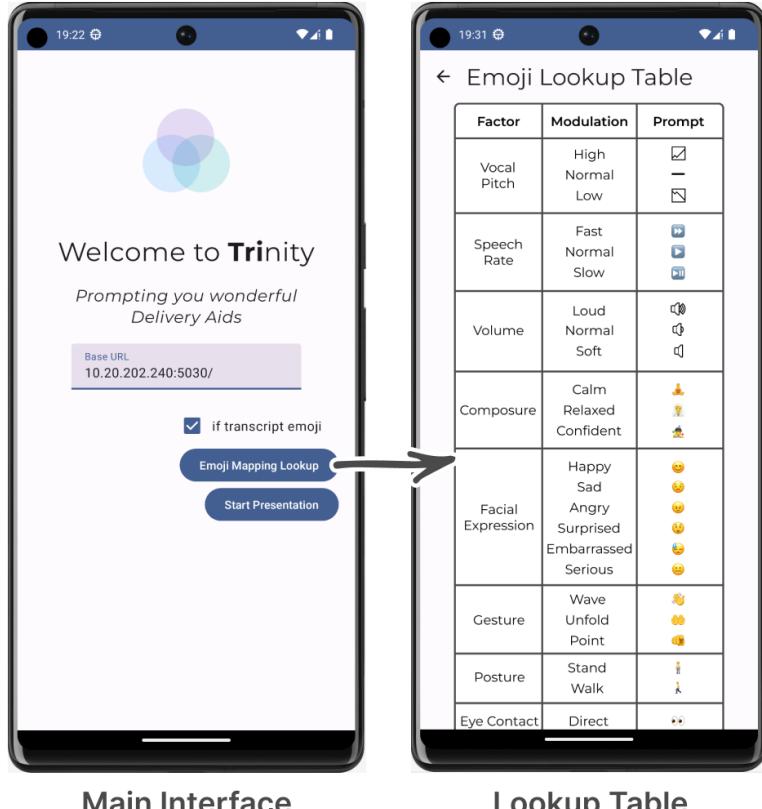
**Minor F2: Presenters frequently check their smartphones during presentations.** Analyzing how presenters use their smartphones during the presentation is crucial. We verified these usage patterns by co-reviewing video replays with student presenters. In general, presenters in both the *Trinity* and *OfficeRemote* conditions frequently glanced at their smartphones, whereas those in the *Inteliprompter* condition primarily focused on their laptop screens. This suggests that *Trinity* did not substantially alter their presentation approach, which primarily relies on scripts:

“*I know it’s (referring scripts) not the best way, but scripts are essential for my speeches because if I speak in English, I might not be able to talk coherently without relying on them.*” (S3, male, age: 22)

**Minor F3: Students generally trust the system but not completely.** All presenters in the *Trinity* condition rated their trust above the neutral point (i.e., 4) on the 7-point Likert scale. They expressed trust in the system because there were no compelling reasons to doubt it and no obviously questionable or unexpected information, which aligns with findings from psychological literature known as the *Truth-Default Theory* [50]. However, their trust in the system was constrained by their limited knowledge of best practices and professional certification.

“*The polished script and provided delivery prompts seem to make sense, but I’m not sure if this is the best way to express or to do. So, I would trust the system more if it’s endorsed by professionals.*” (S3, male, age: 22)

2901  
2902  
2903  
2904  
2905  
2906  
2907  
2908  
2909  
2910  
2911  
2912  
2913  
2914  
2915  
2916  
2917  
2918  
2919  
2920  
2921  
2922  
2923  
2924  
2925  
2926  
2927  
2928  
2929  
2930  
2931  
2932  
2933  
2934  
2935  
2936  
2937  
2938  
2939  
2940  
2941  
2942  
2943  
2944  
2945  
2946  
2947  
2948  
2949  
2950  
2951  
2952  
2953  
2954  
2955  
2956  
2957  
2958



Main Interface

Lookup Table

**Figure 21: Main interface and built-in emoji lookup table of *Trinity* App. Users have access to the emoji lookup table by tapping the “Emoji Mapping Lookup” button in the main interface.**

2959  
2960  
2961  
2962  
2963  
2964  
2965  
2966  
2967  
2968  
2969  
2970  
2971  
2972  
2973  
2974  
2975  
2976  
2977  
2978  
2979  
2980  
2981  
2982  
2983  
2984  
2985  
2986  
2987  
2988  
2989  
2990  
2991  
2992  
2993  
2994  
2995  
2996  
2997  
2998  
2999  
3000  
3001  
3002  
3003  
3004  
3005  
3006  
3007  
3008  
3009  
3010  
3011  
3012  
3013  
3014  
3015  
3016

**Table 6: Questionnaires for student presenters in 7-point Likert scale (1: Not at all/very bad, 7: Very Much/very good)**

Category	Question	
Usability	(1) How easy was the system to use? (2) How helpful was the system in presenting? (3) How distracting was the system in presenting? (4) How satisfying was the system? (5) What is the likelihood of future use?	3075 3076 3077 3078 3079
Effectiveness	(6) How much could the system make the speech more fluid, coherent, and modulated? (7) How much could the system facilitate good eye contacts with audience? (8) How much could the system facilitate appropriate facial expressions? (9) How much could the system facilitate natural gestures? (10) How much could the system enhance the easiness of visual control? (11) How much could the system enhance the effectiveness of slides? (12) How much could the system enhance control of the presentation progress? (13) How much could the system make the presentation more engaging?	3080 3081 3082 3083 3084 3085 3086 3087 3088 3089 3090
Emotion	(14) Did you feel noticed? (15) Did you feel excited? (16) Did you feel exhausted? (17) Did you feel frustrated? (18) Did you feel happy? (19) Did you feel hopeful? (20) Did you feel overwhelmed? (21) Did you feel sense of safety? (22) Did you feel nervous? (23) Did you feel anxious? (24) Did you feel confident?	3091 3092 3093 3094 3095 3096 3097 3098 3099 3100 3101
Load	(25) How is your cognitive load? (26) How is your attentional load? (27) How is your workload? (28) How is your self-perceived performance?	3102 3103 3104 3105
Usefulness	(29) How do you perceive the usefulness of the system? (30) How do you perceive the usefulness of keyword highlighting? (31) How do you perceive the usefulness of remote visual control? (32) How do you perceive the usefulness of underpainting? (33) How do you perceive the usefulness of progress bars? (34) How do you perceive the usefulness of emoji? (35) How do you perceive the usefulness of participle? (36) How do you perceive the usefulness of script polishing? (37) How do you perceive the usefulness of speech tracking?	3106 3107 3108 3109 3110 3111 3112 3113 3114
Accuracy	(38) How do you perceive the accuracy of the whole system? (39) How do you perceive the trust level of the system? (40) How do you perceive the trust level of keyword highlighting? (41) How do you perceive the trust level of remote visual control? (42) How do you perceive the trust level of underpainting? (43) How do you perceive the trust level of progress bars? (44) How do you perceive the trust level of emoji? (45) How do you perceive the trust level of participle? (46) How do you perceive the trust level of script polishing? (47) How do you perceive the trust level of speech tracking?	3115 3116 3117 3118 3119 3120 3121 3122 3123 3124
Trust		3125 3126 3127 3128 3129 3130 3131 3132

	CATEGORY POINTS	RATINGS			
		1	2	3	4
Eye contact	Does not attempt to look at audience	Focuses attention on one particular part of the room, does not scan audience	Regularly makes eye contact with someone or a group; occasionally scans audience	Good eye contact with entire audience; regularly scans room	
Notecards / Notes	Presentation read like a script (uses printed PowerPoint slides or notebook)	Frequently reads note cards	Note cards used on regular basis	Refers to note cards occasionally or no note cards used	
Facial expression	Has either a deadpan expression or shows a conflicting expression during entire presentation	Occasionally displays both a deadpan and conflicting expression during presentation	Occasionally demonstrates either a deadpan or conflicting expression during presentation	Gives audience clues to what the content of presentation is about; Appropriate expression, never notice a deadpan or conflicting expression	
Compose	Obvious anxiety leading to long pauses or continuously confusing material	Anxiety that affects presentation or speech	Fairly at ease with little evidence of anxiety	At ease speaker, enjoys audience interaction, should become a teacher!	
Gestures / Distracting mannerisms	No gestures are noticed; Significant distracting mannerisms	Regular distracting mannerisms	Some distracting mannerisms noted	Natural hand gestures demonstrated; no distracting mannerisms	
Posture	Continually leaning on podium or obvious shifting which affects presentation	Slumps or shifts regularly	Occasionally slumps or shifts during presentation	Stands straight up, both feet on the floor	
Enthusiasm / Vocal pitch	Shows absolutely no interest in topic presented or negativity toward topic	Absolute monotone	Generally shows positive feelings about topic; some pitch variance	Demonstrates a strong positive feeling about topic during entire presentation; uses voice efficiently to emphasize points	
Articulation / Vocalized pauses / Pronunciation of terms	"Vocalized pauses" used continuously throughout presentation; distracting; commonly mispronounced words	>20 vocalized pause; mispronounced terms	10-20 vocalized pauses; minor mispronunciations of terms	< 10 vocalized pauses, rarely mispronounced terms	
Rate of speech	So fast or slow that the talk cannot be understood or the audience cannot be kept awake	Definite tendency for either too fast or slow, such that presentation is difficult to understand	Fast or slow delivery but minimally affects ability to follow presentation	Appropriate rate for audience understanding and attention	
Volume	So poorly heard that there is no variation in intensity, and the presentation points are lost	Significant difficulty in hearing the presentation, with little slight variation in intensity	Some difficulty in hearing presentation with some variation in intensity	Speaker is easily heard with a well-ordered variation in intensity	
Slides effectiveness	Slides so poorly constructed that they detract from presentation - too many word lines or sentences; graphs/tables not described	Many slides ineffective-too wordy lack of variety (ex. all bullet lists w/ no graphs/tables); graphs/tables not described	Too many or too little slides, poor color, font selection or other problems; graphs/tables not always described	Effective slides which enrich the presentation and are easily read; graphs/tables used and described / explained throughout	
Slides spelling / Grammar	>10 spelling/grammar errors	6-10 spelling/grammar errors	1-5 spelling/grammatical errors	No spelling/ grammatical errors	
References slide	No references listed on slides (ex.references used as slide titles)	References listed inappropriately (ex.references used as slide titles)	Occasional reference missing / inappropriate format	References formatted appropriately throughout	

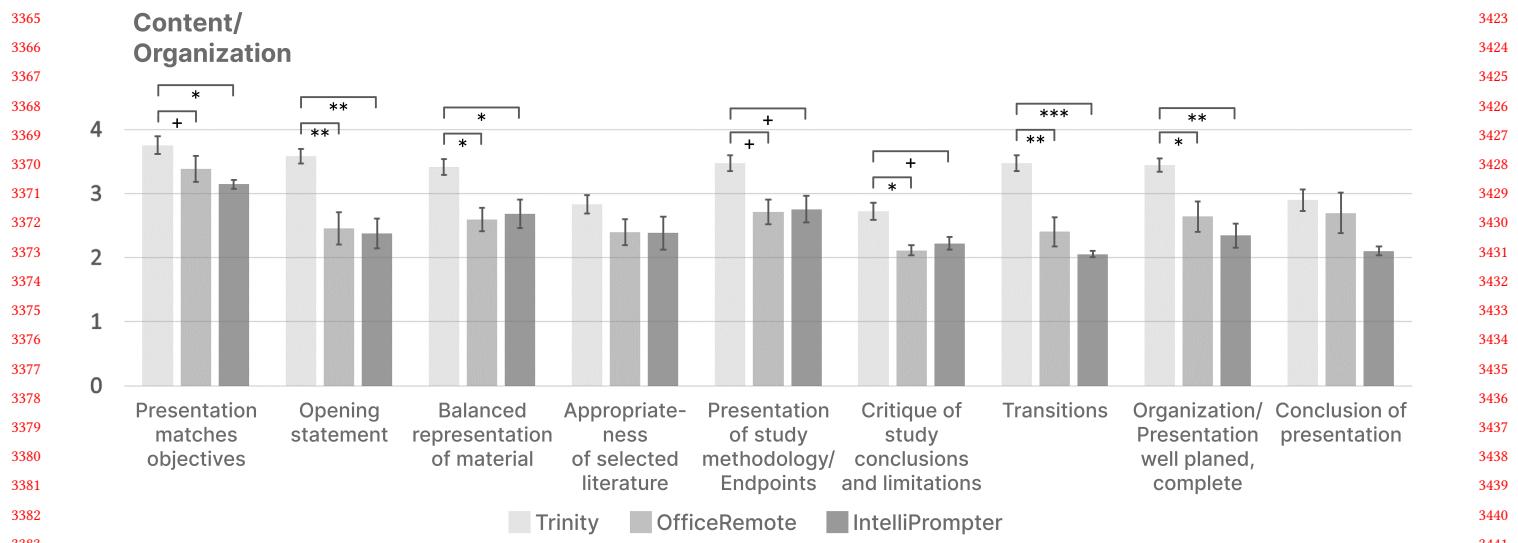
Figure 22: Evaluation form used in the user study.

3133  
 3134  
 3135  
 3136  
 3137  
 3138  
 3139  
 3140  
 3141  
 3142  
 3143  
 3144  
 3145  
 3146  
 3147  
 3148  
 3149  
 3150  
 3151  
 3152  
 3153  
 3154  
 3155  
 3156  
 3157  
 3158  
 3159  
 3160  
 3161  
 3162  
 3163  
 3164  
 3165  
 3166  
 3167  
 3168  
 3169  
 3170  
 3171  
 3172  
 3173  
 3174  
 3175  
 3176  
 3177  
 3178  
 3179  
 3180  
 3181  
 3182  
 3183  
 3184  
 3185  
 3186  
 3187  
 3188  
 3189  
 3190  
 3191  
 3192  
 3193  
 3194  
 3195  
 3196  
 3197  
 3198  
 3199  
 3200  
 3201  
 3202  
 3203  
 3204  
 3205  
 3206  
 3207  
 3208  
 3209  
 3210  
 3211  
 3212  
 3213  
 3214  
 3215  
 3216  
 3217  
 3218  
 3219  
 3220  
 3221  
 3222  
 3223  
 3224  
 3225  
 3226  
 3227  
 3228  
 3229  
 3230  
 3231  
 3232  
 3233  
 3234  
 3235  
 3236  
 3237  
 3238  
 3239  
 3240  
 3241  
 3242  
 3243  
 3244  
 3245  
 3246  
 3247  
 3248

	CATEGORY POINTS	RATINGS			
		1	2	3	4
3249 3250 3251 3252 3253 3254 3255 3256 3257 3258 3259 3260 3261 3262 3263 3264 3265 3266 3267 3268 3269 3270 3271 3272 3273 3274 3275 3276 3277 3278 3279 3280 3281 3282 3283 3284 3285 3286 3287 3288 3289 3290 3291 3292 3293 3294 3295 3296 3297 3298 3299 3300 3301 3302 3303 3304 3305 3306	Delivery Consistency				
Verbal - Nonverbal	The speaker's verbal and nonverbal cues frequently contradict each other, leading to confusion	There are moments of alignment between verbal and nonverbal cues, but inconsistencies are common	Verbal and nonverbal cues align most of the time, but occasional inconsistencies occur	Verbal and nonverbal cues are consistently aligned, effectively conveying a unified message and emotion	
Nonverbal - Visual	Nonverbal elements frequently conflict with or undermine the visuals, causing confusion	There are sporadic instances of nonverbal-visual misalignment, affecting comprehension	Nonverbal elements generally align with visuals, but occasional discrepancies exist	Nonverbal elements consistently complement the visuals, enhancing understanding and engagement	
Visual - Verbal	Visual aids and spoken content frequently diverge, leading to a lack of clarity	There are sporadic instances where visual aids and spoken content do not align perfectly	Visual aids generally support spoken content, with occasional minor discrepancies	Visual aids and spoken content are consistently aligned, effectively conveying the same information	
Presentation matches objectives	Presentation was not related to announced purpose of session	Some objectives addressed	Most objectives addressed	Presentation matched announced purpose and met all objectives	
Opening statement / Relevance to audience	No useful introduction to presentation; audience has no idea what the presentation is on	Minimal opening statement with little mention of relevance of topic to audience	Introduction present, may state how topic impacts audience	Effective opening which states what the presentation will be covering and/or how topic impacted the presenter and audience	
Balanced representation of material	Presentation heavy in introduction/ background material with little emphasis on studies and application	Presentation of material is one-sided or biased; bit too much emphasis on introduction / background	Balanced presentation of Introduction/ background	Balanced presentation of Introduction / background, presentation of literature, and application/ conclusion;	
Appropriateness of selected literature / Data in scope of presentation	Selected literature does not support theme of presentation	Significant gaps in literature presented or selected literature appeared to represent only one side of issue.	Missing some important portion of the literature without stating the limited scope of the presentation	Selected literature supported theme of the presentation and was well balanced	
Presentation of study methodology / Endpoints; includes an interpretation of statistics as appropriate	Studies poorly outlined leading to confusion and / or inappropriate endpoints/ outcomes presented and discussed	Study outline too comprehensive or superficial thus relevant information hard to discern (ie, too much/little data provided) or not emphasized	Outline of study method/results appropriate most relevant data and endpoints described	Studies outlined succinctly and but thorough and pertinent data and endpoints emphasized	
Critique of study conclusions and limitations	Inappropriate conclusions from presented data and/or no critiques or limitations included	Superficial conclusions and critique regarding study/limitations provided (ie."small n": industry funded"etc)(with no explanation of impact on conclusions	Study conclusions could be more thorough; attempted to provide critique with limitations and explanation	Conclusions thorough and appropriate for studies and placed into context with similar literature; thoughtful critique/limitations provided	
Transitions	No transitions	Transitions exist and are obvious and not creative	Good transitions exist, but are not seamless	Excellent transitions throughout which are seamless	
Organization /	Organization / Presentation well planned, complete / coherent	Many points left out; talk was completely disorganized	Majority of points glossed over; insufficient depth: difficult to follow talk due to disorganization	Majority of points covered in depth; some important points may be unclear; some organization issues	
Application / Conclusion of presentation (ie. take home message valid / clear)	No application to practice nor is a conclusion presented	Opinions on application to practice or a conclusion presented but they are not supported by data presented	Superficial conclusions or opinions presented with limited reference to support from data	Valid conclusion presented which were supported by data	

Figure 23: (continued) Evaluation form used in the user study.

3307  
3308  
3309  
3310  
3311  
3312  
3313  
3314  
3315  
3316  
3317  
3318  
3319  
3320  
3321  
3322  
3323  
3324  
3325  
3326  
3327  
3328  
3329  
3330  
3331  
3332  
3333  
3334  
3335  
3336  
3337  
3338  
3339  
3340  
3341  
3342  
3343  
3344  
3345  
3346  
3347  
3348  
3349  
3350  
3351  
3352  
3353  
3354  
3355  
3356  
3357  
3358  
3359  
3360  
3361  
3362  
3363  
3364



**Figure 24: Audience's ratings on content and organization of students' presentations. The error bars indicate standard errors.** (+ :  $p < .1$ ; \* :  $p < .05$ ; \*\* :  $p < .01$ ; \*\*\* :  $p < .001$ )