E·S·P·Today
JOURNAL OF ENGLISH FOR SPECIFIC PURPOSES AT TERTIARY LEVEL

Fei-Wen Cheng*

National Chiayi University, Taiwan
chengfw@mail.ncyu.edu.tw

HOW EFL STUDENTS EXPERIENCE AND ADAPT TO ASYNCHRONOUS VIDEO INTERVIEWS: THE NEED FOR A MULTIMODAL ESP INSTRUCTION

Abstract

Asynchronous video interviews (AVIs) are being increasingly used in modern recruitment, yet little is known about how English as a Foreign Language (EFL) candidates navigate the unique multimodal, cognitive, and affective challenges of this format. This study qualitatively explored seven Taiwanese EFL students' experiences of completing a simulated AVI task, drawing on post-task one-on-one interviews to uncover their preparation strategies, performance struggles, and emotional responses. Thematic analysis identified four key themes: (1) preparation strategies, (2) challenges encountered, (3) anxiety and emotional strain, and (4) perceptions of the AVI platform's affordances and constraints. The findings highlight the heightened communicative complexity of AVIs for EFL learners, and underscore the need for English for Specific Purposes (ESP) training that integrates multimodal communication skills, affective strategy training, and familiarity with AVI platforms. This study contributes to understanding AVIs as a multimodal genre of professional communication, and offers practical insights for enhancing EFL students' digital interview preparedness and professional identity development.

Key words

asynchronous video interviews (AVIs), job interviews, multimodal communication, ESP genre.

* Corresponding address: Fei-Wen Cheng, No. 85, Wenlong Village, Minxiong Township, Chiayi County 621302, Taiwan.

1. INTRODUCTION

The job interview has long been a high-stakes genre in which candidates express their professional identities through both verbal and nonverbal communication for job-hunting purposes (Lipovsky, 2006; Posthuma et al., 2002). In recent years, this genre has been transformed ensuing the rapid development of digital recruitment technologies. Asynchronous video interviews (AVIs), in particular, have become increasingly prevalent across industries due to their operational efficiency and capacity for large-scale implementation (Arseneault & Roulin, 2024; Dunlop et al., 2025; Lukacik et al., 2022; Torres & Gregory, 2018). AVIs require candidates to submit pre-recorded responses to standardized prompts delivered either in written or video form without any real-time interaction with an interviewer. These responses are then evaluated by either human recruiters or artificial intelligence (AI) systems at a later date (Lewton & Haddad, 2024; Lukacik et al., 2022).

Despite the growing prominence of AVIs in the international job market, this emerging professional genre has received relatively limited scholarly attention within the field of English for Specific Purposes (ESP). Unlike traditional interviews, AVIs eliminate live communication, depriving candidates of adaptive cues such as interviewers' smiles, nods, or follow-up questions (Rizi & Roulin, 2023; Suen & Hung, 2024). Candidates must instead perform self-contained, camera-facing monologues that require the strategic coordination of verbal, vocal, and visual resources without interpersonal feedback (Langer et al., 2017; Lukacik et al., 2022; Suen et al., 2019). The transition from traditional face-to-face interviews to asynchronous formats introduces new communicative and performative challenges for applicants, fundamentally changing the skill set needed to succeed in professional selection processes (Dunlop et al., 2022).

These challenges can be particularly acute for English as a Foreign Language (EFL) learners. Although recent studies have begun to investigate general applicant perceptions of AVIs (e.g., Roulin et al., 2023; Zibarras et al., 2025), there remains a significant gap in understanding how EFL candidates orchestrate the complex interplay of linguistic fluency, multimodal coordination, and genre-specific expectations in constructing a credible professional identity within this genre.

To address this gap, the present study investigated EFL students' experience of participating in AVIs, and the challenges and strategies that emerged throughout their preparation, performance, and self-presentation processes. Drawing on post-task interviews conducted with students who completed an AVI assignment, this study offers nuanced insights into the linguistic, multimodal, cognitive, and affective dimensions of AVI engagement. The findings not only contribute to a richer understanding of how EFL learners navigate this emerging genre, but also yield significant pedagogical implications for incorporating AVI training into ESP curricula.

2. LITERATURE REVIEW

2.1. Conceptual framework: ESP multimodal spoken genres

Job interviews, whether conducted face-to-face or via technology-mediated platforms, are inherently multimodal communicative events in which candidates must strategically integrate verbal content, vocal delivery, and visual presentation (Fortanet-Gómez & Beltrán-Palanques, 2022). These performances rely on the coordination of multiple semiotic modes: linguistic (spoken language), paralinguistic (intonation, pitch, pacing), and visual (facial expression, gesture, appearance), each contributing to meaning-making in the interaction. This multimodal orchestration, encompassing what is said, how it is said, and how one appears while saying it, plays a pivotal role in impression management, shaping perceptions of professional competence, confidence, and interpersonal appropriateness (Fortanet-Gómez & Beltrán-Palanques, 2022; Lee, 2023, 2024; Lipovsky, 2006, 2008; Van De Mieroop, 2019). At the heart of this process is the achievement of *intersemiotic coherence*, that is, the harmonious alignment of multiple semiotic modes to create a cohesive and convincing self-presentation. When this alignment breaks down, it can reduce perceived authenticity and weaken the candidate's professional credibility (Roulin et al., 2023).

Although empirical studies on job interviews from a multimodal perspective remain limited, emerging research from adjacent ESP-related genres provides a valuable conceptual foundation (e.g., Lindenberg, 2023; Ruiz-Madrid, 2021). Studies of online presentations and spoken performance genres consistently demonstrate that speakers who exhibit intersemiotic alignment, synchronizing verbal messages with congruent vocal tone, gaze, and gesture, are more likely to be evaluated as confident, credible, and persuasive (e.g., Lindenberg, 2023; Valeiras-Jurado & Ruiz-Madrid, 2020).

Lindenberg (2023) found that students delivering online presentations were rated more favorably when they combined clear articulation with a steady gaze, expressive intonation, and natural body language. In contrast, speakers who read monotonously from a script appeared disengaged and less credible. Similarly, Ruiz-Madrid (2021), analyzing academic pitch videos, observed that combinations of gaze control and gestural coordination often exerted a stronger influence on audience evaluations than verbal content alone. Similar patterns have been observed in entrepreneurial pitch contexts, where alignment between verbal content and positive nonverbal cues, such as an open posture and enthusiastic tone, enhances the speaker's credibility (Valeiras-Jurado & Ruiz-Madrid, 2020). Their findings suggest that persuasive impact stems not only from what is said, but also from how it is physically delivered.

Together, these studies emphasize the importance of intersemiotic coherence in shaping audience perceptions across a range of high-stakes spoken genres. How

EFL students orchestrate verbal, vocal, and visual modes to convey credibility and engagement, along with verbal skill warrants further research.

2.2. Challenges encountered in asynchronous video interviews

A growing body of literature highlights the multifaceted challenges that AVIs present for job applicants (Lukacik et al., 2022; Rizi & Roulin, 2023; Suen & Hung, 2024). At the interactive level, AVI platforms eliminate the real-time exchange that characterizes face-to-face communication. The interface typically provides *no live feedback*: candidates speak to a camera or avatar with no indication of how their answers are being received. The AVI structure thus deprives candidates of conversational resources that normally assist in self-presentation and impression management in a traditional job interview (Lipovsky, 2006; Van De Mieroop, 2019). Without an interactive dialogue, the responsibility falls to the interviewee to interpret questions accurately, formulate structured responses, and deliver them fluently while simultaneously monitoring their tone, facial expressions, gaze, and timing. This multitasking can impose substantial cognitive load on candidates. Roulin et al. (2023) reported that even well-prepared individuals may struggle to coordinate multiple communicative modes under the one-way pressure of an AVI.

The affective demands of AVIs further complicate the experience. This absence of direct human presence can produce a sense of detachment and uncertainty, often described as speaking into a void (Rizi & Roulin, 2023; Suen & Hung, 2024). Research also indicates that knowing one's recorded answers can be replayed and scrutinized by multiple people (or algorithms) tends to heighten anxiety and self-consciousness. Lewton and Haddad (2024) noted that candidates often become hyper-aware of micro-level behaviors, posture, intonation, and so on, which can lead to stiff or unnatural performances. In Roulin et al.'s (2023) study, many participants felt that AVI formats reduced their ability to "be themselves," replacing a genuine exchange with a performative script. Indeed, the absence of social presence cues (no smiling interviewer, no nodding head) can make it difficult for candidates to project warmth or enthusiasm confidently.

At a technological level, the camera framing typical of AVI platforms, usually limited to head-and-shoulders, restricts the use of full-body gestures and natural posture shifts, forcing candidates to rely more heavily on facial expressions, eye contact with the lens, and vocal modulation to convey engagement. Candidates must also manage the technical setup themselves, adjusting the camera position, lighting, audio quality, and background to ensure a professional appearance. Any deficiencies in these areas may adversely influence evaluators' impressions, regardless of the candidate's qualifications (Dunlop et al., 2022; Lukacik et al., 2022). Proficiency in these technical aspects is now essential for achieving a professional communicative impression.

As noted above, the reviewed literature highlights the transformative impact of AVIs on recruitment practices, the intricate multimodal rhetorical demands of professional identity construction, and the compounded technological, cognitive, and affective challenges faced by candidates. Preparing EFL learners to handle such demands is increasingly important for ESP practitioners, as traditional interview training may not address the multimodal and self-directed nature of AVI tasks (Blacksmith et al., 2016; Roulin et al., 2023). This study contributes to that discussion by providing empirical insights into how EFL students navigate and sometimes struggle with the complexities of AVIs in practice.

3. METHOD

3.1. Participants and contexts

The study was conducted at a public university in southern Taiwan in 2022, focusing on senior undergraduates in the Department of Foreign Languages. Participants were initially invited to volunteer through an in-class announcement made in a Business Communication elective course taught by the researcher. Seven fourth-year students (five females and two males) were purposively selected from the volunteer pool based on two criteria: English proficiency and immediate postgraduation career plans. Specifically, all selected participants had at least a B2 level of English proficiency (upper-intermediate), as evidenced by standardized language test scores or course performance, and all intended to enter the job market directly after graduation rather than pursue further studies.

These selection criteria ensured a sample of EFL learners who showed sufficient English ability to participate in an English-language interview and who had a vested interest in developing job-hunting skills. Although the Business Communication course covered workplace genres, it did not include job interview preparation, making the AVI task a novel communicative genre for the participants. All seven selected students (aged approximately 21-22) received detailed information about the research purpose and procedures, and each provided informed consent prior to participation. Pseudonyms (Student 1 through Student 7) are used to maintain confidentiality. The study protocol was reviewed and approved by the ethics committee of a government-accredited university in Taiwan authorized to evaluate research ethics compliance.

3.2. AVI tool

The AVI task was carefully designed to simulate a realistic job screening scenario using InterviewStream, a commercial AVI platform widely adopted in both educational and professional recruitment settings (e.g., InterviewStream, 2024;

Torres & Gregory, 2018). Designed to support asynchronous interviewing, InterviewStream allows recruiters or researchers in this context to customize interview experiences by selecting questions from a vast library of over 7,000 industry-specific prompts, determining the number of allowed attempts, and setting time limits for responses. The platform's ability to present a virtual recruiter in the form of a human-like avatar instead of a static text prompt enhances the realism of the simulation.

Additionally, InterviewStream supports mock interview practice, allowing users to rehearse and refine their responses in a low-pressure environment. Participants can record, review, and re-record their answers to improve their performance. Moreover, it enables the researcher to set parameters, such as the number of recording attempts per question and to impose time limits, thereby balancing standardization with flexibility in performance. Participants could also delete and re-record answers to improve clarity or confidence.

3.3. AVI task design

In preparation for the AVI, participants first attended a 20-minute instructional workshop in Chinese led by the researcher. This workshop introduced the InterviewStream platform's interface, technical navigation procedures, and key features such as time limits, recording functions, and options for multiple attempts. Following the workshop, students were provided with a written guide summarizing the procedural steps for completing the AVI, tips for managing both content and delivery, and a reminder that the task should be treated with the seriousness and professionalism of an actual job interview, despite it not being a course assignment. It should be emphasized that while technical guidance on platform navigation was provided, no explicit instruction regarding interview performance strategies or AVI-specific communicative techniques was offered. This decision was made to ensure that participants' interactions with the AVI task reflected the authentic experiences of novice users encountering this format for the first time.

Each participant was tasked with responding to a set of 10 pre-selected interview questions (see Appendix A), selected from the platform and reflecting standard themes in job recruitment processes. Participants were allotted 2 minutes to respond to each interview question using the InterviewStream platform, which allowed up to five recording attempts per question. While students had a maximum of one day to complete the recording part of the AVI task, a two-week window was provided for the overall assignment. This structure was designed to strike a balance between allowing deliberate preparation and multiple practice opportunities, while maintaining time constraints that simulated the conditions of real-world asynchronous video interviews.

3.4. Data collection procedures

Data collection proceeded in four interconnected phases (see Figure 1) to capture both students' observable performance and their retrospective reflections on the AVI experience. Before undertaking the AVI task, participants completed tailored written résumés as a course assignment in response to a self-selected job advertisement, simulating an authentic application scenario (Phase 1). While the résumé was not the primary data source for this study, its close relevance to the AVI task positioned it as a preparatory pre-AVI activity.

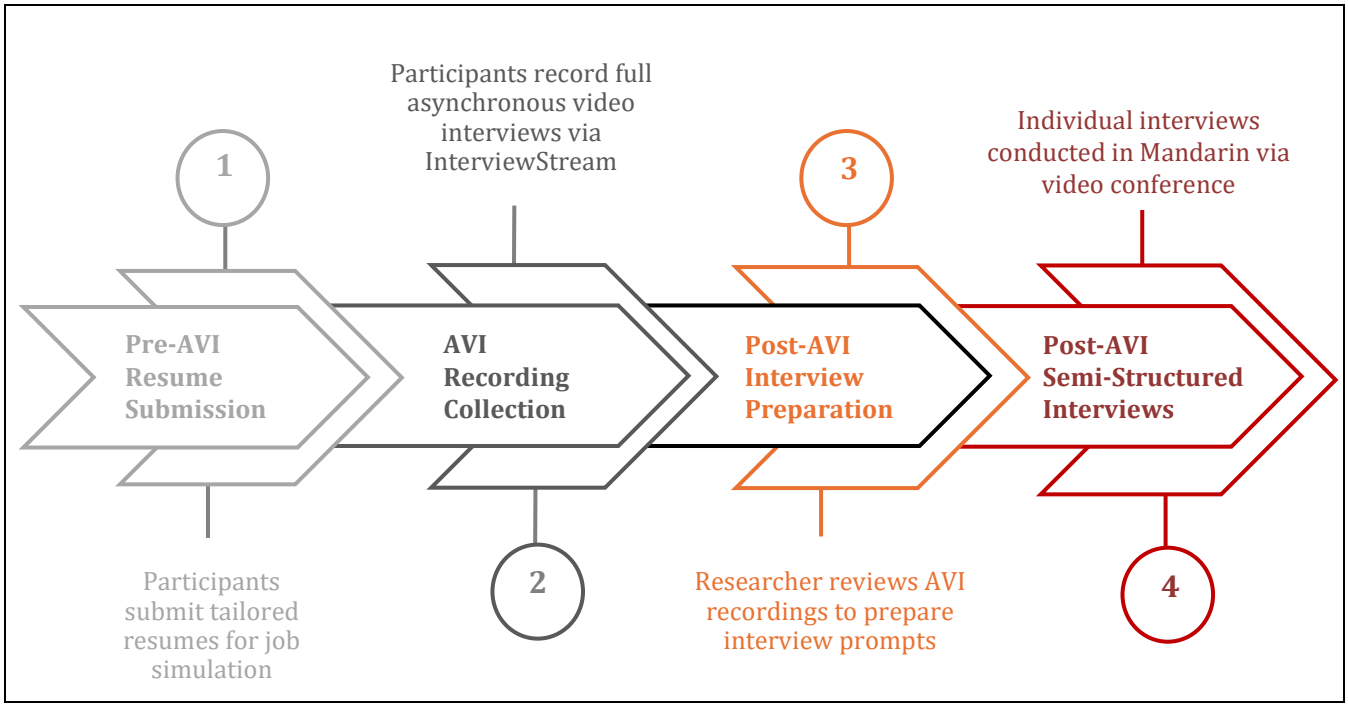

Figure 1. Data collection process

In the second phase, participants' full AVI recordings in English were collected via the InterviewStream platform. These recordings served as the primary observational data, allowing the researcher to gain an informed understanding of students' multimodal self-presentation strategies, including verbal fluency, content coherence, paralinguistic delivery (e.g., tone, pacing, and intonation), visual behaviors (e.g., eye contact, facial expressions, posture, and gestures), and technical framing choices (e.g., camera positioning and background management). Although a systematic analysis of the AVI recordings was beyond the scope of the present study, the video data provided important contextual insights that enriched the subsequent post-task semi-structured interviews.

Next, following the completion and submission of their video-recorded interviews, the researcher reviewed each AVI recording to develop a preliminary understanding of the participants' performance and engagement with the task.

In the final phase, post-task semi-structured interviews was conducted individually with each participant within two weeks of completion of the AVI task. These interviews were held via video conferencing platforms (Google Meet or Microsoft Teams), depending on participants' availability and platform familiarity. To enable participants to express themselves fully and with nuance, interviews were conducted in Mandarin Chinese, their first language. The interview protocol was semi-structured in format, with guiding questions adapted from Li and Deng (2021). Selected questions used in the interview process are provided in Appendix B.

Each interview lasted approximately 60 to 75 minutes and was video-recorded with participants' prior consent. All interviews were transcribed verbatim, yielding a total of 72,767 words of transcribed text. The semi-structured interview protocol was designed to elicit participants' reflections on a range of topics, including their preparation strategies (e.g., script drafting, rehearsal practices, and time management approaches), technological challenges encountered (e.g., platform navigation, audiovisual setup, and troubleshooting experiences), cognitive and affective experiences during the AVI task (e.g., feelings of anxiety, confidence, or cognitive overload), and perceived difficulties in multimodal communication (particularly in simultaneously managing verbal, vocal, and visual modes). To facilitate a deeper exploration of their multimodal self-presentation strategies, participants were also invited to review selected segments of their own AVI recordings during the interview.

3.5. Data analysis

Data analysis followed the six-phase thematic analysis framework proposed by Braun and Clarke (2021), with particular attention paid to the multimodal, cognitive, and affective dimensions of AVI experiences (see Figure 2). First, the researcher engaged in repeated readings of the AVI recordings and interview transcripts to achieve deep familiarization with the data, allowing emerging patterns, affective nuances, and multimodal behaviors to become salient.

In the second phase, an initial coding process was undertaken using a hybrid inductive-deductive approach. Deductive codes were informed by the theoretical focus on multimodal communication strategies and affective-cognitive challenges, while inductive codes were derived directly from participants' narratives, capturing emergent patterns and previously unreported strategies. Subsequently, related codes were organized into broader thematic categories. This process yielded four preliminary domains: preparation and rehearsal behaviors; multimodal self-presentation and cognitive challenges; emotional reactions to AVI performance; and technological management and platform perceptions.

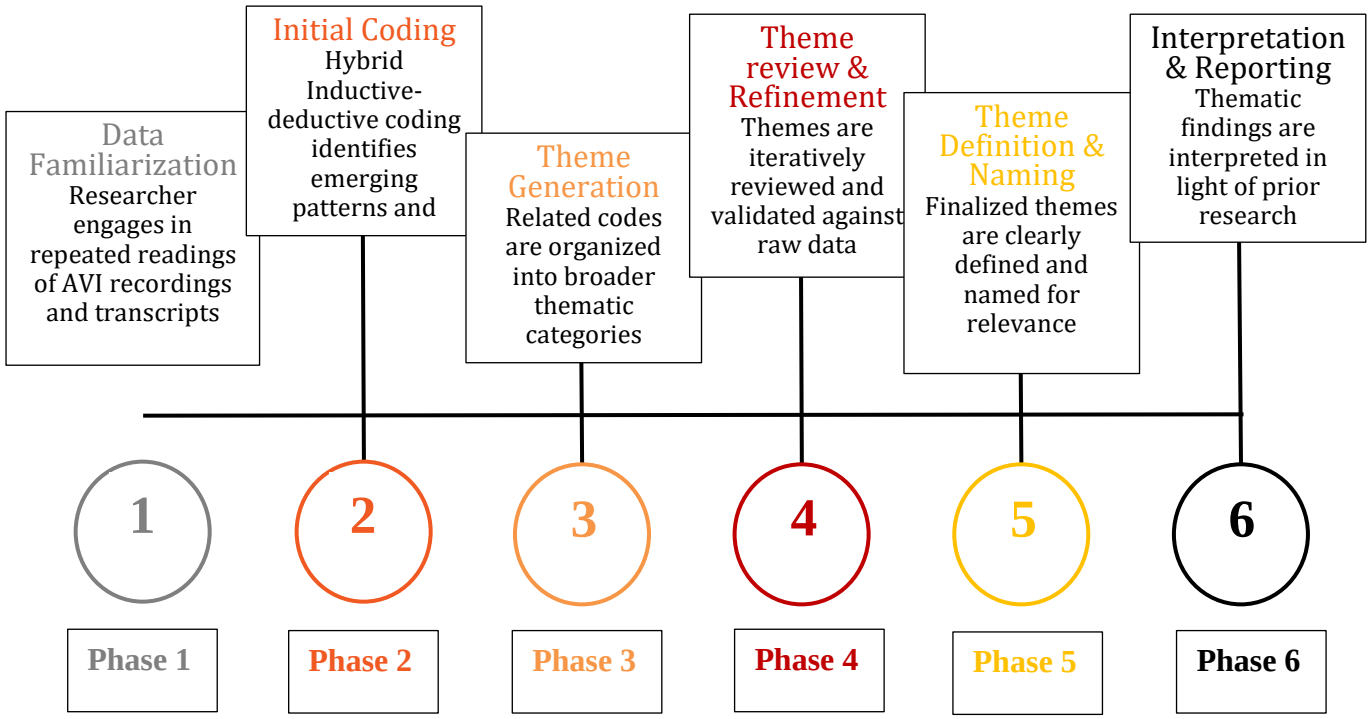

Figure 2. Data analysis stages

Themes were reviewed and refined through an iterative process to ensure internal coherence and distinctiveness, by continuously referring to the raw data to validate interpretations and minimize overgeneralization. This involved revisiting participants' original verbal responses and, where relevant, their nonverbal behaviors captured in the AVI recordings, to ensure the themes authentically reflected both the spoken and multimodal dimensions of their experience. Final themes were clearly defined and labelled to emphasize their theoretical relevance to the research questions, particularly in relation to multimodal discourse frameworks and existing AVI literature.

Nvivo's AI-assisted features supported data analysis by enhancing coding consistency and transparency. All coding outputs were subsequently reviewed and verified by the researcher to maintain reliability and consistency with the research objectives. This triangulated approach reinforced methodological rigor and analytic depth in interpreting participants' AVI experiences.

4. RESULTS

The findings of the study showed four identified themes that characterize students' experiences with AVIs: (1) preparation strategies, (2) challenges encountered, (3)

anxiety and emotional strain, and (4) perceptions of the implemented AI-based interview platform. Together, these themes illuminate the multifaceted demands that AVIs place on EFL learners, encompassing linguistic, multimodal, cognitive, and affective dimensions. The findings highlight participants' strategies for negotiating the technological, rhetorical, and emotional challenges embedded in AVI performance. All participant quotations were translated from the original Chinese with careful attention to preserving their intended meanings, ensuring the integrity and authenticity of the interview data.

4.1. Preparation strategies: From script to screen

All seven students undertook extensive preparation for the AVI, engaging in a range of strategies to maximize their performance. A consistent pattern was the significant time investment in scripting and organizing responses before ever hitting the "record" button. Participants reported devoting substantial time to preparing for the interview, which involved drafting responses to anticipated questions, refining their language for grammatical accuracy and fluency, and organizing their answers into coherent, logically structured narratives. *"I probably spent four or five afternoons (around 50 hours) just writing my answers,"* admitted Student 4, emphasizing the heavy commitment to crafting well-formulated responses. Student 1 similarly described a lengthy drafting process, stating, "*I think it took me about 10 days. I wrote down things I personally considered important and kept a record.*" Student 7 also echoed the importance of scripting, explaining, "*I first wrote out all the answers and reviewed them carefully.*"

Several students treated the preparation almost like writing an essay or exam: they carefully chose words and examples, and some even applied formal frameworks. For instance, Student 3 used the STAR method (Situation, Task, Action, Result) to structure his answers: *"I answered the questions using the STAR method… For example, if a question asked about how I define success, I would first explain my view on success, then bring in my personal experience and use the STAR method to elaborate."* By employing such frameworks, students aimed to ensure their content was thorough and in accordance with professional interview conventions suited to an interview context, even without an interviewer present.

Beyond writing out their answers, participants sought external input and models to improve their content. Sensing gaps in their experience and uncertainty about what recruiters might expect, many students turned to online resources. "*This kind of interview had many new questions we'd never seen in casual part-time job interviews, so I had to look up how to answer them,*" explained Student 6, who searched the internet for sample answers and tips. Student 7 described a similar strategy, "*I referred to the résumé we wrote, and also searched online to see how others answered such questions.*" She further explained, "*I just pasted the question into Google,*" indicating how she adapted available answers into her own personal

narratives. While looking up model answers is not unique to AVIs, it underscores a broader challenge overlapping with face-to-face interviews. Students' reliance on online exemplars indicates a need for rhetorical support: they wanted to understand what *kinds* of answers would sound credible and how to *frame* their own stories effectively.

Another major aspect of preparation was rehearsal, particularly oral and multimodal practice. They tried to monitor their nonverbal behaviors such as maintaining a pleasant facial expression and good posture. Student 3 recounted, "*At first, I didn't use the assigned software; instead, I used my own computer to record myself facing the camera and simulated an interview. During this practice, I paid attention to whether I was smiling and whether my eye contact remained consistently on the camera.*" Similarly, Student 4 said, "*I also recorded each question two or three times… I'd review my performance and if my eye contact drifted or I spoke too fast, I re-recorded it.*" Student 1, by contrast, took a mixed approach to rehearsal, stating, "*Some questions I answered by writing a script first; others I responded to more spontaneously, adjusting my answer while recording.*" This reveals an attempt to balance memorization and naturalness.

The platform's affordances also influenced preparation behaviors. InterviewStream (like many AVI platforms) allowed multiple attempts for each question, and students took full advantage of this in their actual recordings. This led to a "perfectionist" approach. Student 5 reflected on this during her interview: *"I thought it would be quick, but it took me about 6 hours of recording. Whenever I stumbled or forgot a word, I felt I had to start over."* She had not anticipated needing so many retakes, but knowing she *could* retry made her unwilling to settle for an imperfect response. Student 2 likewise noted, *"If I wasn't satisfied, I would just redo it from the start."* This behavior highlights a significant shift. On the one hand, the ability to redo gave them a sense of control over quality; on the other hand, it led to obsessive self-critiquing. This phenomenon echoes Suen and Hung's (2024) observation that AVI environments tend to heighten candidates' self-monitoring and encourage cycles of perfectionism. In a live interview, a candidate cannot "restart" an answer midway; the conversation flows on, and minor mistakes are often overlooked. In AVIs, every recorded answer could be scrutinized, so these students pushed themselves to eliminate mistakes entirely, which demanded enormous effort.

4.2. Challenges encountered in AVI

Despite their thorough preparation, students encountered numerous difficulties during the AVI performance itself. These challenges focused on managing nonverbal communication in the absence of feedback and coping with cognitive pressure as they tried to deliver their answers in English on camera.

4.2.1. Nonverbal communication and visual performance

One common difficulty the students discovered was a mismatch between how they thought they appeared and how they actually appeared on camera. In the post-task interviews, many reflected that when they watched their AVI video, they noticed their expressions seemed tense or blank even when they believed they had been smiling or looking confident. "*I thought I was smiling, but on camera it didn't really look like a smile,*" Student 1 admitted. She elaborated, "*I might have been smiling, but the smile needed to be stronger.*" "*Probably because I was concentrating so much on my script and the camera, I forgot to smile*," added Student 4, recognizing that her focus on the content made her facial expression unintentionally serious. Similarly, Student 5 noted, "*I realized I kept frowning; it looked strange.*" These realizations indicate that in the moment of recording, participants were not fully aware of their nonverbal presentation. They felt they were conveying a pleasant demeanor, but without the normal feedback cues, their internal sense did not match the external reality. The discrepancy between felt emotion and visible expression underscores the increased self-regulation AVIs demand. Candidates must not only manage what they say, but also consciously project engagement and positivity through their nonverbal behavior, something that typically happens more intuitively in interactive settings.

Another pervasive issue was maintaining eye contact with the camera. Eye contact is a key component of conveying confidence and sincerity in an interview, but in an AVI it translates to looking into the webcam lens for an extended period, which can be an unnatural act for most people. Several students admitted their gaze wandered during their responses. Student 1 acknowledged, "*The most important thing in an online interview is expressing yourself toward the camera, maintaining eye contact and remembering to smile*." "*I noticed my eyes were looking down in the video. I adjusted later so that I was looking into the camera lens*," said Student 3, who caught herself on playback and tried to correct it in subsequent answers. Student 5 shared, "*Maintaining continuous eye contact with the webcam during the entire interview process is a challenge for me. This is why I often find myself looking away.*" Apparently, it simply did not feel normal to stare into a tiny lens non-stop. Student 6 admitted, "*I'm not used to looking at the camera when speaking, even in video chats with friends, I usually turn the camera off*." Student 7 shared a similar reflection, "*I tend to get quite nervous, which causes me to frequently glance elsewhere. Fixing my gaze on the webcam is not easy for me, possibly because I'm not accustomed to speaking directly to the camera.*" These comments reveal that even in more casual online communication, Students 6 and 7 were not used to maintaining visual presence, a habit that carried over into the AVI, where turning off the camera was not an option. These struggles with gaze confirm findings by Lukacik and Bourdage (2024) that sustaining eye contact with a camera is particularly difficult in AVI contexts. Although the participants were aware that they *needed* to look at the camera, doing so felt artificial and was easy to forget when they were concentrating on recalling their answers or controlling their anxieties.

Gestures and body language posed an additional challenge. As the platform primarily recorded candidates from the chest or midsection upward, many students deliberately limited their hand movements, concerned that their gestures might appear awkward or be partially cropped by the camera frame. *"I find it somewhat odd to incorporate body language during an AVI. Since we're recording our answers while seated in front of the computer with a relatively static posture, it can be challenging to effectively convey gestures,"* explained Student 4. Student 5 had a similar experience: *"I found it tricky to use body language… with only my head and shoulders in frame, any big hand movement felt weird, so I basically kept gestures small or avoided using them."* Essentially, the constraints of the medium led them to minimize gesturing, even if that made them appear a bit stiff. They were unsure how much was "too much" on camera and chose to do less.

Notably, the absence of real-time feedback in the AVI placed significant demands on students' ability to self-manage their nonverbal communication. They had to generate the sense of an interactive presence through deliberate effort: maintaining eye contact with a faceless camera, consciously smiling or conveying enthusiasm, and ensuring their gestures and surroundings aided rather than hindered their professional image. Many of them expressed that they had never thought about these details before this task, indicating that AVI-specific training (e.g., camera gaze or on-screen persona) could be highly beneficial for students like them.

4.2.2. Cognitive load and fluency breakdown

In addition to nonverbal issues, the students struggled with cognitive and linguistic challenges during the AVI. Despite extensive rehearsal, several found that once the camera was recording, they experienced unexpected memory lapses and loss of fluency. Student 3 confessed, "*Even though I wrote everything down beforehand, I couldn't say it smoothly when the camera was on. I forgot parts and got stuck.*" This highlights a classic pitfall: having a script in mind is not the same as delivering it under pressure. The act of delivering spoken responses without a script, even after rehearsal, under high-pressure conditions strained her cognitive resources and led to brief disfluencies. Student 4 shared a similar sentiment, "*I'm good at writing, but speaking without a script is much harder. I kept repeating myself.*" Student 2 added, "*I kept forgetting words. If I stumbled even a little, I'd feel the whole recording wasn't good enough and redo it*." Perhaps most strikingly, Student 1 said, "*I memorized my lines, but I forgot them during recording. Then I just froze.*" Despite having prepared thoroughly, the stress of recording caused her mind to blank out, resulting in a block where she could not continue. These scenarios demonstrate how performance pressure can disrupt cognitive processes, especially in an AVI where the EFL candidates are acutely aware that the "take" is being recorded for judgment.

These observations support the notion of cognitive overload in AVIs. When candidates must manage content (what to say), language (how to say it correctly in English), and delivery (speaking clearly and naturally while on camera)

simultaneously, it creates a heavy mental burden. Thus, some mental "bandwidth" issues occurred, resulting in hesitations or blanks. Roulin et al. (2023) emphasized that even well-prepared individuals can struggle under these conditions, as the lack of interactive support and the awareness of being recorded can hinder performance. These EFL participants who had fluently practiced their answers alone still encountered fluency breakdowns when it was "for real" on the AVI. As shown, these cognitive demands highlight that thorough preparation did not guarantee flawless performance. Moreover, disfluencies or brief mental blocks become more detrimental, as there was no real-time interaction to mitigate them. The format's lack of immediate feedback and high-pressure recording environment introduced a risk of delivery breakdowns that even the good students were not immune to.

4.3. Students' emotional experiences with AVI

Students' emotional reactions to the AVI experience were complex and sometimes contradictory, reflecting the distinctive psychological challenges posed by asynchronous interviewing, particularly for EFL learners.

One prominent dimension of students' emotional experiences concerned their reactions to the AVI format itself. For some participants, the asynchronous environment offered a sense of relief. The ability to prepare, rehearse, and record responses without the immediate presence of a live interviewer appeared to mitigate social pressure. For instance, Student 6 reflected positively: *"Not having the interviewer right in front of me made it a lot less nerve-racking; I actually wasn't as nervous."* Similarly, Student 4 appreciated the opportunity to pre-script and control responses: *"Compared to an in-person interview, I was less nervous because I could script out what to say beforehand."* These experiences echo Lewton and Haddad's (2024) findings that AVI practice increased candidates' confidence and reduced interview-related stress.

In contrast, Student 5 acknowledged that, *"I was extremely nervous; even though it was pre-recorded, I was still so nervous. I probably need to adjust my mindset."* Likewise, Student 6 remarked, "*I actually wanted to smile… I re-recorded each one many times… because I get nervous easily and tend to fumble my words.*" Their accounts affirm findings in AVI research indicating that one-way interviews intensify performance anxiety and emotional self-monitoring (Rizi & Roulin, 2023; Roulin et al., 2023; Suen & Hung, 2024). Despite these stressors, some students demonstrated proactive coping strategies to manage their anxiety during the AVI process. For example, Student 5 attempted mindset adjustment: *"I kept telling myself this is just a recording and I can always do it again, so there's no need to get too nervous."* This strategy is consistent with Rizi and Roulin's (2023) suggestion that reframing the AVI experience as a form of practice rather than a final evaluation can help mitigate performance anxiety.

Another striking emotional strain was the awkwardness of speaking to a camera without any real-time audience. *"It felt like I was just talking to myself. I couldn't tell if my answer made sense or if I should add more,"* said Student 2, describing the unease caused by the lack of feedback. Student 2 further reflected, "*The screen was all black (while recording), and I just felt a bit insecure, not knowing what was going on*…" Student 3 echoed this: *"The hardest part was that there was no one reacting to what I said. In a real interview, you adjust based on the interviewer's expression or questions."* Without those signals, she was unsure if she was hitting the mark. Student 7 put it bluntly: *"Honestly, I'm not used to it, talking to a computer without a response."* These comments illustrate what can be described as "speaking to a void" (Suen & Hung, 2024, p. 7), a phenomenon reflecting their discussion of how AVIs, which lack real-time feedback and social presence, can make speaking feel strangely context-less. The students felt disconnected and even a bit awkward at times, as if their words were floating aimlessly. This lack of reciprocal interaction not only heightened their anxiety but also made it hard to maintain a natural delivery.

4.4. Perceptions of the AVI platform

Students' perceptions of the InterviewStream platform revealed both an appreciation of its structured practice environment and frustration with its lack of interactivity and authenticity. Their mixed evaluations mirror broader debates in the literature about the affordances and limitations of AVI platforms for developing job interview skills (Lewton & Haddad, 2024; Rizi & Roulin, 2023).

First of all, students expressed gratitude for the platform's role in providing exposure to diverse question types and opportunities for repeated practice without the immediate pressure of live interviews. Student 4 highlighted the value of the platform's extensive question database: "*It provides so many different questions that really helped someone like me, a newbie, to see what might be asked*." Student 6 also praised the structured nature of the platform, noting that: "*The system's organized design made it easier to practice, and reduced some anxiety compared to real-time interviews*." These perspectives support the findings of Lewton and Haddad (2024), who noted that AVI platforms can help students develop initial interview readiness, particularly for novices with little prior experience. The participants credited the platform with giving them exposure and experience: by practicing on InterviewStream, they got a sense of real interview formats and were able to rehearse without the immediate pressure of facing an interviewer. In essence, they treated it as a learning tool.

However, most students voiced concerns about the platform's lack of interpersonal interaction. The absence of any real-time reactions or adaptive follow-up questions contributed to a sense of detachment, weakening the authenticity of the experience. As Student 2 expressed: "*In a normal video call or in-person interview*

you can see the other person, or at least see yourself, but here it's just a black screen. It gave me no sense of feedback or security." Her use of a "black screen" is telling as, during recording, the platform hides the self-view to avoid distraction, which made it hard to gauge her own demeanor in the moment. Similarly, Student 5 commented on the impersonal nature of the interaction: "*Although it was helpful to practice answering questions, I felt strange talking into a void without seeing anyone's reaction*." This perceived sense of isolation corresponds with Rizi and Roulin's (2023) finding that AVI platforms using static, text-based prompts or non-interactive avatars can intensify candidates' feelings of social isolation and communicative uncertainty.

Students also criticized the platform for failing to replicate the dynamic nature of real interviews, where interviewer reactions guide the conversation. Student 3 commented, "*It would be great if, based on my answer, it could ask a second question – more like a real interview*." Likewise, Student 6 lamented: "*It feels unnatural*. *The interviewer just asks one question and then it's all you talking.*" These reflections highlight what Suen and Hung (2024) regard as the monologic constraint of AVIs: candidates must deliver complete, self-contained responses without conversational scaffolding, increasing the cognitive burden and distancing the task from authentic workplace communication. Given the conversation is not reciprocal but a one-way delivery, this was viewed as a negative.

In light of these limitations, several students proposed improvements to make the platform more interactive and pedagogically supportive. As Student 3 envisioned, "*It would be more like a real interview if the system could continue the conversation based on what I said*." Essentially, adding a layer of simulated social presence could mitigate that "detachment" issue and help candidates remain engaged. They also desired more immediate and formative feedback mechanisms. Although InterviewStream offered basic playback and self-assessment options, some participants suggested that integrating AI-generated feedback (e.g., comments on speech pace, use of fillers, gaze direction) or instructor-modeled examples would significantly enhance the learning experience. Student 5 suggested: "*It would help if the platform could give some suggestions after recording, like if I spoke too fast or if my eye contact was bad*." This resonates with findings by Lewton and Haddad (2024), which indicated that students practicing with AI-enhanced AVI platforms valued instant performance feedback, such as indicators of speech pace, filler words, or gaze direction. It is worth noting, however, that given the rapid development of AI tools, certain paid versions of AVI tools may already offer upgraded functions capable of addressing these limitations.

5. DISCUSSION

The purpose of this study was to uncover how EFL students experience an evolving job-hunting genre, the AVI task. This genre, in line with previous research on

multimodal job application genres, such as traditional interviews and video résumés, serve as key spaces for self-branding and identity negotiation (Fortanet-Gómez & Beltrán-Palanques, 2022; Van De Mieroop, 2019). The four themes identified in the current findings reveal the complex and multifaceted demands that asynchronous video interviews impose on EFL learners across linguistic, multimodal, cognitive, and emotional dimensions. Although students demonstrated strong commitment to preparation, and gained important insights into professional self-presentation, the task simultaneously highlighted critical areas where pedagogical support is needed, particularly in scaffolding multimodal communication strategies, managing performance anxiety, and facilitating strategic identity construction.

The difficulties that participants faced in managing eye contact, facial expression, and body language point to a critical need for explicit multimodal communication training in ESP curricula. Traditional interview preparation often emphasizes *what to say* (linguistic content), yet this study underscores that *how it is delivered* (paralinguistic and nonverbal elements) is equally critical. In AVIs, where face-to-face interaction is absent, visual and vocal cues become the principal channels through which rapport is established. As several ESP studies have emphasized (Fortanet-Gómez & Beltrán-Palanques, 2022; Lee, 2023, 2024; Lindenberg, 2023; Ruiz-Madrid, 2021; Valeiras-Jurado & Bernad-Mechó, 2022), successful professional communication requires the integrated orchestration of verbal and nonverbal modes. The participants experienced difficulty integrating these modes, indicating a lack of sufficient training to manage them in a coherent manner. One key pedagogical implication is that ESP instructors should incorporate modules aimed at developing multimodal literacy (Fortanet-Gómez & Beltrán-Palanques, 2022), including practical training on maintaining camera eye contact, managing facial expressions, and using gestures effectively within the camera frame. The goal is to help learners internalize that content and delivery are intertwined: a confident answer is not just well-phrased but also well-delivered through posture, gaze, and voice. This approach reflects the notion of genre-specific communicative competence, recognizing AVIs as an emerging genre that requires distinct multimodal skills in ESP instruction.

Moreover, the results highlight the importance of structured, platform-specific AVI practice within ESP courses. Beyond general interview training, it is critical to simulate AVI environments by providing students with authentic timed recording tasks, exposure to typical platform features, and opportunities for repeated practice under AVI-specific conditions (Lewton & Haddad, 2024). Such targeted preparation would not only reduce students' technological anxiety, but also better equip them to manage time constraints, self-monitor multimodal performance, and deliver coherent responses without real-time interviewer support (Lewton & Haddad, 2024). In short, students need guidance not only in *what to perform* but also in *how the medium works*. This will better equip EFL candidates to handle AVI contexts confidently and strategically.

Another pedagogical implication here is that preparing students for AVIs (and interviews in general) must address the emotional dimension, not solely the content and skills. ESP pedagogy must include affective scaffolding. Instructors should mitigate AVI-related anxiety by embedding stress management and step-by-step practice sessions designed to strengthen students' confidence and emotional control (Roulin et al., 2023).

6. CONCLUSION

This study has provided new insights into how EFL students experience asynchronous video interviews, illuminating the linguistic, multimodal, cognitive, and emotional challenges they face in navigating this increasingly prevalent hiring genre.

Several limitations of this research must be acknowledged and their implications for future research were also discussed. First, the small sample size and single institutional context limit the generalizability of the findings. The experiences of seven Taiwanese students, while rich in detail, may not represent the full diversity of EFL learners' experiences with AVIs. Future studies should include larger and more diverse samples to capture a broader range of strategies and challenges. Second, this study focused on one particular AVI platform and a simulated context. Different platforms with varied interface features (e.g., avatar vs. text prompts, different time constraints) could yield different user experiences. Future research should compare multiple AVI systems and settings to determine how specific design features impact EFL candidates' communicative strategies and stress levels. Third, the data were collected from immediate post-task interviews. A longitudinal approach would provide a more comprehensive view of learning gains and behavioral changes, following students through repeated AVI practices or actual job applications to see how their competencies evolve over time.

With AVIs now an integral part of recruitment practices, it remains essential for ESP educators and researchers to continue investigating this dynamic and evolving genre. This study offers an initial contribution by illuminating both the challenges and developmental opportunities that AVIs present for EFL learners. Targeted pedagogical interventions are integral to empowering learners to communicate effectively and authentically within the increasingly complex digital contexts of the 21st-century workplace.

[Paper submitted 22 Aug 2025]
[Revised version received 21 Oct 2025]
[Revised version accepted for publication 28 Oct 2025]

Acknowledgement

This research was funded by the Ministry of Science and Technology in Taiwan (Grant No. MOST 110-2410-H-415-013). I would like to acknowledge the use of artificial intelligence tools to enhance the clarity and coherence of the manuscript in the editing of this paper. However, all final revisions and content adjustments were made by me.

References

Arseneault, R., & Roulin, N. (2024). Investigating impression management use in asynchronous video interviews across 10 countries. *International Journal of Selection and Assessment*, *32*(3), 461–477. https://doi.org/10.1111/ijsa.12476

Blacksmith, N., Wilford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions*, *2*(1), Article 2. https://doi.org/10.25035/pad.2016.002

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.

Dunlop, P. D., Holtrop, D., & Wee, S. (2022). How asynchronous video interviews are used in practice: A study of an Australian-based AVI vendor. *International Journal of Selection and Assessment*, *30*(3), 448–455. https://doi.org/10.1111/ijsa.12372

Dunlop, P. D., Hickman, L., Holtrop, D., & Powell, D. M. (2025). Asynchoronous video interviews in recruitment and selection: Lights, camera, action! *International Journal of Selection and Assessment*, *33*(2), Article e70010. https://doi.org/10.1111/ijsa.70010

Fortanet-Gómez, I., & Beltrán-Palanques, V. (2022). Enhancing multimodal communicative competence in ESP: The case of job interviews. *Multimodal Communication*, *11*(1), 17–29. https://doi.org/10.1515/mc-2020-0032

InterviewStream. (2024). https://interviewstream.com/

Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, *25*(4), 371–382. https://doi.org/10.1111/ijsa.12191

Lee, J. C. P. (2023). Gaze and facial expression in engineering student presentations: A comparative case study of a high- and low-performing presenter. *ESP Today*, *11*(1), 6–30. https://doi.org/10.18485/esptoday.2023.11.1.1

Lee, J. C. P. (2024). Co-speech gestures in enhancing engineering proposal presentations: A multimodal analysis of ESP spoken discourse. *ESP Today*, *12*(2), 319–348. https://doi.org/10.18485/esptoday.2024.12.2.6

Lewton, J., & Haddad, A. R. (2024). Pharmacy student perceptions of using AI-based interview software (Big Interview) for career preparation. *American Journal of Pharmaceutical Education*, *88*(9), Article 101263. https://doi.org/10.1016/j.ajpe.2024.101263

Li, Y., & Deng, L. (2021). Disciplinarily capable and personally unique: Voicing disciplinary identity in personal statement writing. *Journal of English for Academic Purposes*, *50*, Article 100949. https://doi.org/10.1016/j.jeap.2020.100949

Lindenberg, D. (2023). Modes and intersemiotic cohesion in student presentations performed online: An SF-informed multimodal discourse analysis. *English for Specific Purposes*, *69*, 67–79. https://doi.org/10.1016/j.esp.2022.10.002

Lipovsky, C. (2006). Candidates' negotiation of their expertise in job interviews. *Journal of Pragmatics*, *38*(8), 1147–1174. https://doi.org/10.1016/j.pragma.2005.05.007

Lipovsky, C. (2008). Constructing affiliation and solidarity in job interviews. *Discourse & Communication*, *2*(4), 411–432. https://doi.org/10.1177/1750481308095938

Lukacik, E.-R., & Bourdage, J. S. (2024). Does design matter? The (limited) effects of six asynchronous video interview design features on impression management, reactions, and evaluations. *International Journal of Selection and Assessment*, *33*(1), Article e12511. https://doi.org/10.1111/ijsa.12511

Lukacik, E.-R., Bourdage, J. S., & Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interview. *Human Resource Management Review*, *32*(1), Article 100789. https://doi.org/10.1016/j.hrmr.2020.100789

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, *55*(1), 1–81. https://doi.org/10.1111/j.1744-6570.2002.tb00103.x

Rizi, M. S., & Roulin, N. (2023). Does media richness influence job applicants' experience in asynchronous video interviews? Examining social presence, impression management, anxiety, and performance. *Journal of Selection and Assessment*, *32*(1), 54–68. https://doi.org/10.1111/ijsa.12448

Roulin, N., Pham, L. K. A., & Bourdage, J. S. (2023). Ready? Camera rolling… action! Examining interviewee training and practice opportunities in asynchronous video interviews. *Journal of Vocational Behavior*, *145*, Article 103912. https://doi.org/10.1016/j.jvb.2023.103912

Ruiz-Madrid, N. (2021). A multimodal discourse approach to research pitches. *Journal of English for Academic Purposes*, *52*, Article 101003. https://doi.org/10.1016/j.jeap.2021.101003

Suen, H.-Y., & Hung, K.-E. (2024). Revealing the influence of AI and its interfaces on job candidates' honest and deceptive impression management in asynchronous video interviews. *Technological Forecasting and Social Change*, *198*, Article 123011. https://doi.org/10.1016/j.techfore.2023.123011

Suen, H.-Y., Chen, M. Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior*, *98*, 93–101. https://doi.org/10.1016/j.chb.2019.04.012

Torres, E. N., & Gregory, A. (2018). Hiring manager's evaluations of asynchronous video interviews: The role of candidate competencies, aesthetics, and resume placement. *International Journal of Hospitality Management*, *75*, 86–93. https://doi.org/10.1016/j.ijhm.2018.03.011

Valeiras-Jurado, J., & Bernad-Mechó, E. (2022). Modal density and coherence in science dissemination: Orchestrating multimodal ensembles in online TED talks and YouTube science videos. *Journal of English for Academic Purposes*, *58*, Article 101118. https://doi.org/10.1016/j.jeap.2022.101118

Valeiras-Jurado, J., & Ruiz-Madrid, N. (2020). When pointing becomes more than pointing: Multimodal evaluation in product pitches. *Language & Communication*, *74*, 74–86. https://doi.org/10.1016/j.langcom.2020.06.006

Van De Mieroop, D. (2019). Implying identities through narratives of vicarious experience in job interviews. *Journal of Pragmatics*, *152*, 61–75. https://doi.org/10.1016/j.pragma.2018.01.006

Zibarras, L. D., Castano, G., & Cuppello, S. (2025). Applicant perceptions of selection methods: Replicating and extending previous research. *International Journal of Selection and Assessment*, *33*(2), Article e70007. https://doi.org/10.1111/ijsa.70007

FEI-WEN CHENG is a Professor in the Department of Foreign Languages at National Chiayi University, Taiwan. Her teaching focuses on advanced academic literacy and English for Specific Purposes (ESP). Her research interests include academic discourse, genre analysis, and multimodality, with an emphasis on how these areas inform language learning and professional communication. Her recent work has been published in peer-reviewed journals, such as the *Journal of English for Academic Purposes*, *Ibérica*, *ESP Today*, and *System*. ORCID: 0000-0001-9756-4918

Appendix A

Video interview questions

1. Tell me about yourself.
2. How would you define success?
3. Describe an ideal working environment.
4. What are 2-3 weaknesses you need to improve?
5. What are your long-term career goals?
6. What are your salary expectations?
7. What are your strengths?
8. What is your greatest achievement and why?
9. Why do you want to work here?
10. Why should I hire you?

Appendix B

Exemplary post-task interview questions
(adapted from Li & Deng, 2021, pp. 13–14)

1. What did you do before you started the video interview? Did you turn to anyone or any resources for help?
2. Can you describe your interviewing process? (What did you pay attention to most while you were working on it: the content of your response, language use, voice, or body language? What were the most difficult question(s) to respond to? What challenges did you encounter when using this interview format?)
3. What did you intend to highlight in your video interview?
4. Did you have a target recruiter in mind when performing the video interview?
5. What kind of image did you want to create in your video interview?