

Behind the Script: Verbal and Nonverbal Pragmatic Features in Indonesian EFL Learner's Job Interview Simulations

Athiyah Salwa^{1,2*}, Rudi Hartono², Yuliaty²

1. Universitas Stikubank, Semarang Jawa Tengah

2. Universitas Negeri Semarang, , Semarang Jawa Tengah

*corresponding author e-mail: athiyahsalwa2@gmail.com

Article Info	Abstract
Keywords: Job interview Simulations; Nonverbal Communication; Pragmatic Competence; Pragmatic Failure; Speech Acts.	Job interview simulation is widely employed in English for business courses to develop communicative competence among Indonesian EFL learners in professionally meaningful contexts. However, when performed through memorized scripts, such simulations risk prioritizing surface fluency over pragmatic appropriateness. Scripted performances may also inadequately reflect learners' multimodal pragmatic readiness, particularly in nonverbal dimensions. While prior studies have examined EFL speaking performance in interview contexts, few have integrated verbal speech act analysis with nonverbal behavior as co-expressive pragmatic features in scripted simulations, particularly among non-English major learners in Indonesian higher education. This study investigates the pragmatic competence of non-English major students by examining their verbal speech acts and nonverbal behavior in scripted job interview simulations. Adopting qualitative multimodal discourse analysis, nine out of fifteen video recordings were analyzed using a speech act coding scheme alongside nonverbal features coded following Ho's multimodal interview competence framework, focusing on self-promotion, hedging, and clarification requests co-expressed through gesture, gaze, posture, and prosody. Findings reveal that participants successfully enacted genre structure across opening, body, and closing stages, with self-promotion as the dominant speech act (68.4%), followed by hedging (18.4%) and clarification requests (13.3%). However, pragmatic competence was predominantly formulaic and script-dependent, with verbal assertiveness routinely undermined by deferential nonverbal behavior, including intermittent downward gaze during self-promotion, flat prosodic delivery, and turn-final smiling following weakness disclosure. These performance shows systemic L1 sociocultural transfer of typical Javanese-Indonesian norms of modesty and the style across verbal and nonverbal channels. These findings argue that redesigning job interview simulations in EFL
Article history: Submitted June 22, 2026 Revised July 18, 2026 Accepted July 18, 2026	

curricula should cover both speaking tasks and explicitly integrated pragmatic awareness and multimodal communicative development. Pedagogical instruction for both channels of communication is essential to serve ad genuine preparation for authentic English-medium professional interaction.

Introduction

English proficiency nowadays is an essential asset for university graduates as Indonesia continues to strengthen its dominance across the global economic landscape ((Takaya, 2024)). As industries increasingly operate across linguistic (Chen, 2022) and cultural boundaries, Indonesian higher education institutions put English for Specific Purposes (ESP) curricula and integrated the aim of learning in the professional context (Marsakawati *et al.*, 2019). In non-English major, English for Business is now being taught in the early year and one of the requirements of the assignment must passed is job interview simulation. It is a high-stakes professional genre that employ more than surface level of grammatical correctness in a verbal context. To perform in job interview simulations properly, Indonesian EFL learners have to employ complex speech acts (Darong, 2024), manage interpersonal face dynamics, calibrate register, and navigate culturally specific communicative expectations (Yan, 2022) in real-time communication. Language competence in this instruction (Burns, 2019) includes not only the ability to communicate effectively with grammatical, textual, functional, and sociolinguistic knowledge but also socio-pragmatic competence (Leech, 1983). Critically, professional communication in an interview context is inherently multimodal in how it is delivered through gesture, facial expression, eye contact, and prosodic features such as intonation, stress, and pace (Ho, 2019).

Practically, a significant and continuing gap remains between the outcomes of Indonesian EFL speaking instruction and what professional communication demands. Speaking pedagogy at the tertiary level, especially for non-English major students, remains fluency-oriented on the surface over the pragmatic decision-making. It defines authentic professional interactions due to a lack of knowledge and experience with real professional contexts. Prominently, discourse competence is not only the composition of logical and coherent texts but also the knowledge of different discourse types and the rules of their formation in accordance with the communicative situations, including extra-linguistic features (Ageeva, 2016). The symptom of this pedagogical orientation is known as "scripted fluency". This term refers to the ability to produce polished and grammatically accurate English sentences within a rehearsed framework (Reeves, 2010) while remaining pragmatically underprepared and socially embedded in the demands of real professional encounters. This problem was subsequently compounded by the sociocultural dimension of Indonesian EFL learners' communicative repertoire (Deda, 2013) and adopting the Indonesian norms of the culture, such as indirectness, hierarchical deference, and communicative modesty. These sociocultural norms are employed not only in verbal choices but also in non-verbal behaviors (Takaya, 2024) such as downward gaze, closed

posture, and soft-volume delivery that undermine the illocutionary force of well-formed verbal utterances. In this phase, learners transfer pragmatic strategies of the L1 into English professional contexts.

Collectively, several studies affirm that job interviews are high-stakes communicative events in which success depends on both linguistic proficiency and the capacity to navigate cultural and pragmatic norms. The question then arises “to what extent do instructional approaches account for simultaneous of verbal and nonverbal pragmatic features in EFL learners’ performance?

Some of the previous studies has examined how verbal and nonverbal dimensions intersect in students’ speaking performance, especially in job interview simulations as the channel for speaking assessment in higher education level. At the secondary level, Kamaliah *et al.* (2018) investigated the effectiveness of the Three-Step Interview technique as a cooperative learning strategy for improving ESL speaking skills among high school students. Their findings demonstrated that structured interpersonal interaction through this technique significantly improved students’ performance and generated positive motivation beyond outcomes that are associated with teacher-centered instruction (Kamaliah *et al.*, 2018). At the university level, Krishnan *et al.* (2021) examined how professional communicative competence and lexical choices of Malaysia’s graduates directly influenced their success in job interviews. The researchers found that successful candidates articulated a broad range of employability skills and identified limited English proficiency as the primary barrier to employment. It advocated for revamped educational modules that foreground professional vocabulary instruction (Krishnan *et al.*, 2021).

To design pedagogical framework intensively, Fortanet-Gomez and Betran-Palanques (2022) proposed a multimodal communicative competence approach integrated into the instruction of job interview preparation from the view that professional success requires the mastery of multimodal ensembles by performing gestures, eye contact, pitch, and facial expression. Through multimodal instructional analysis, their study showed how specific nonverbal cues function to establish rapport and convey professional credibility (Fortanet-Gómez & Beltrán-Palanques, 2022). Moreover, no study has investigated how Indonesian EFL learners move beyond memorized responses to deploy verbal and nonverbal pragmatic features adaptively within simulated professional encounters.

The recognition that professional communication cannot be reduced to its verbal dimension has generated a growing body of scholarship on multimodal interaction in both professional and educational settings. By engaging Norris’s foundational framework of multimodal interaction analysis, researchers have demonstrated that nonverbal behavior carries independent illocutionary weight rather than supplementing verbal utterances (Norris, 2004). It establishes the assumption that all human interaction is inherently multimodal in its constitution. This understanding has contributed significantly to how scholars theorize by which speakers construct and shape communicative events, engaging interlocutors through varied and simultaneously deployed communicative modes (Fortanet-Gómez & Beltrán-Palanques, 2022).

In higher education settings, multimodal engagement has been shown to support participants' active involvement in academic discourse and to illuminate the meaning-making potential of diverse semiotic resources, including gesture, gaze, and spatial positioning (Morell, 2018). These modal combinations do not operate in isolation (Peng *et al.*, 2017), but rather their communicative effectiveness is conditioned by the coordinating of timing, rhythm, and socially shared norms that govern language use in context (Alviar *et al.*, 2023). It acquires heightened significance in professional speaking genres where interactional alignment carries evaluative consequences. Critically, however, this body of scholarship has been developed primarily within European ESP contexts and has not been applied to the performance of Indonesians. EFL learners in scripted speaking assessment conditions, where additional sociocultural variables bear directly on how multimodal choices are made and interpreted.

Another constraint students face in transforming their L1 into L2 during job interview simulations is sociocultural transfer. Sociocultural transfer refers to the use of one's own cultural ways of speaking to a learned language (Al-Issa, 2003). A prominent issue is that sociocultural forces shape how EFL learners make pragmatic choices, both verbal and nonverbal in professional speaking tasks. It may happen because they do not understand the cultural differences between particular speech events in their L1 and L2; learners may inadvertently project a negative image to potential employers, a phenomenon labeled as "cross-cultural pragmatic failure" (Otlowski, 2008). In the Indonesian sociocultural context, this transfer exists through communicative values embedded in hierarchical social structures such as indirectness as a face-saving strategy, deference toward authority figures, and communicative modesty that resists unambiguous self-promotion (Deda, 2013). These norms extend beyond lexical choice to nonverbal behavior (Zarrinabadi & Khodarahmi, 2017), such as downward gaze as a marker of respect, closed posture as a signal of deference, and subdued vocal delivery as an expression of modesty (Wilmes & Siry, 2021). All of which can undermine the illocutionary force of otherwise well-formed English utterances in a professional interview context because L2 pragmatic competence often focuses on learners' speech act behavior, primarily by contrasting nonnative with native performance (M. Yu, 2011). Consequently, scripted fluency in job interviews performed by Indonesian EFL students may produce grammatically correct utterances. However, students who memorize scripts remain pragmatically and multimodally unprepared for authentic professional interaction.

Drawing on speech act theory (Blake *et al.*, 2011; Sbisa, 2002) and multimodal interactional analysis (Stivers & Sidnell, 2005; Wilmes & Siry, 2021; C. Yu, 2013), the present study examines the scripted yet pragmatically revealing performances of Indonesian EFL learners in job interview simulations. Rather than treating verbal output in isolation, this study adopts an integrated analytical lens for the co-occurrence and the mutual reinforcement of linguistic and nonverbal communicative features within a structured speaking assessment context. In doing so, it contributes to the growing body of literature on multimodal pragmatic competence in EFL professional contexts. It offers pedagogical implications for the design of English-speaking instruction and assessment in Indonesian

higher education, particularly for non-English majors. This study is guided by three research questions: what speech acts do Indonesian EFL learners perform in scripted job interview simulations? How are these speech acts co-expressed through nonverbal behavior? And how do verbal, nonverbal, and prosodic features reflect L1 sociocultural transfer in Indonesian EFL learners' scripted job interview simulations?

Research Method

1. Design

This study employs qualitative multimodal discourse analysis as its methodological framework. Qualitative discourse analysis is appropriate for this study because it foregrounds meaning, context, and communicative intent as manifested in simulated spoken discourse. This framework allows the researcher to account for the complexity of pragmatic performance. The study adopts a descriptive research design to produce a systematic and theoretically grounded description of the pragmatic features, both verbal and nonverbal.

The researchers used three sequential and integrated layers of data analysis combining Austin (Austin, 1962) and Searle's (Searle, 1969) speech act theory, interactional multimodal analysis, and pragmatic transfer theory (Takahashi, 1996). The three layers were connected simultaneously to each sentence in the corpus table. The first research question will be answered using speech act theory to determine kinds of speech acts employed in the scripted expressions. During the categorizing of the speech act on each sentence, the researcher identified nonverbal features employed to match the verbal and nonverbal aspects of each utterance. After both analyses had been conducted, the last step was analyzing the sociocultural transfer from the transcription. The illustration of the data analysis can be seen through the figure of the theoretical framework below:

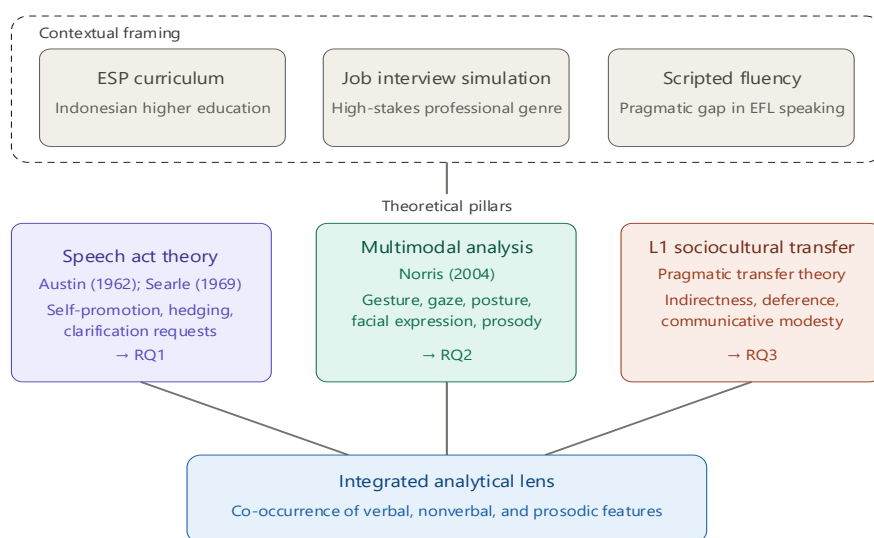


Figure 1. Theoretical Framework

2. Population and Sample

This study was conducted at an Indonesian university within an English for Business course at the undergraduate level. This subject is designed for non-English major students in business and management programs for whom English is a compulsory general education subject. This population is representative of the broader Indonesian undergraduate students who will use English in professional communication contexts and is aligned with the growing demands of the labor market. The job interview simulation was conducted as a speaking task within the course. Students were required to perform a scripted job interview simulation in which both the interviewer and candidate roles were assigned within each group. Students drafted their own dialogue scripts and rehearsed them prior to the recorded performance. Prior to scripting, students received instruction from the lecturer on verbal and nonverbal performance conventions appropriate for job interview contexts. Students had previously been exposed to relevant communicative expressions for self-introduction, motivational statements, strengths and weaknesses, clarification requests, and interview closings.

The original data comprised 15 video recordings of scripted job interview simulations. From this pool, 9 recordings were selected for analysis through purposive sampling. The single inclusion criterion applied in the purposive sampling process was recording duration. Only videos with a minimum duration of three minutes (03:00) were included in the analytical corpus. This threshold was established on the grounds that recordings shorter than three minutes were deemed to lack sufficiently complete interview structure, falling short to provide adequate coverage of key interview phases, such as data such as opening, competency questioning, self-awareness, and closing.

3. Data Collection

All data collected in the original pool were recorded by students using their personal devices, either smartphones or laptop built-in cameras, and submitted as part of their course assessment. The recordings were subsequently merged into three videos and were uploaded to Turboscribe.ai for automated transcription. Turboscribe.ai was selected as the primary transcription tool because it enables direct alignment of verbal content, facilitating subsequent sentence-by-sentence corpus construction. After the videos have been merged and the individual timestamp is omitted, the utterances were then identified through sequential sentence-level coding within the corpus. Each of the video was then coded as follows:

Table 1. Corpus Summary: Included and Excluded Recordings

Participant's Code	Original Pool	Duration	Included in Analysis
1	SC01	02.09	No
2	SC02	05.22	Yes
3	SC03	02.56	No
4	SC04	01.48	No
5	SC05	04.23	Yes
6	SC06	02.55	No
7	SC07	03.01	Yes

Participant's Code	Original Pool	Duration	Included in Analysis
8	SC08	07.18	Yes
9	SC09	04.13	Yes
10	SC10	03.50	Yes
11	SC11	02.17	No
12	SC12	03.10	Yes
13	SC13	06.27	Yes
14	SC14	02.23	No
15	SC15	04.12	Yes

To maintain the validity of the transcription data, a triangulation procedure was conducted in three sequential steps. First, the recordings were processed through Turboscribe.ai to generate the primary transcripts. Second, the original recordings were subsequently compiled into a single video and uploaded as an unlisted file on YouTube to preserve participant privacy and enabled caption generation. YouTube's automatic speech recognition (ASR) system generated timestamped captions and served as a secondary transcription source for cross-validation purposes. Third, the transcriptions from both sources were systematically compared to verify consistency. If there were differences among both results of transcription, the researcher reviewed the relevant video segment and determined the accurate transcription through auditory verification. These three steps of the triangulation procedure ensure the final transcripts used in the analysis are as accurate as practically achievable.

4. Data Analysis

The data analysis step was conducted for verbal and nonverbal features. They were identified through analytical procedure and consolidated into an integrated summary table. For each participant utterance, the coded speech act, the corresponding multimodal performance features, and the researcher's interpretive assessment of sociocultural transfer from the learners' L1. All the categories then displayed as in Table 2 below:

Table 2. Corpus Summary: Speech Act and Multimodal Analysis

Participant	Interview Stage	Speech Act Type	Verbal Excerpt	Illocutionary Function	Nonverbal Co-Expression	Alignment
P1	Opening	Self-promotion	"I am a very responsible person and I always finish my work on time"	Assert competence	Intermittent downward gaze; hands clasped on lap; flat intonation with no emphatic stress	⚠ Misaligned
P4	Body	Self-promotion	"I have experience in teamwork and I believe I can	Assert transferable skill	on "responsible." Nodding while speaking; brief eye contact; monotone delivery with	⚠ Partially misaligned

Participant	Interview Stage	Speech Act Type	Verbal Excerpt	Illocutionary Function	Nonverbal Co-Expression	Alignment
P8	Closing	Self-promotion	contribute to this company" "I hope I can be part of this company and I will give my best performance"	Assert commitment; close interaction	reduced stress on contribute Slight smile; sustained eye contact; moderate pace; mild stress on best performance	✓ Aligned

The data were analyzed through an integrated procedure by using speech act theory (Austin, 1962; Searle, 1969) and multimodal interactional analysis (Norris, 2004) to both written transcripts and video recordings. Analytical categories were not predetermined from the literature but emerged from close and repeated engagement following an inductive approach. Each applicant turn was first segmented into utterances based on the shifts in illocutionary intent, then coded according to its dominant communicative function. Three speech act categories emerged from this process were coded as a combined category and analyzed for pragmatic tension between them.

To ensure the validity of both the transcription and the speech act coding, a triangulation procedure was employed involving three undergraduate students of English literature as independent coders. Each coder was provided with the raw data of video recording and transcription. The coders have been explained above how they should label each utterance following the guidelines of the codes. They independently identified and labeled speech acts across the applicant utterances before their coding decisions were compared against the primary researcher's coding. The agreement of the three coders was conducted across discussion and consensual discussion to ensure that the final coding scheme reflected a shared verified reading of the data. This multi-coder verification procedure addresses the reliability concern that is particularly salient in qualitative pragmatic analysis. Finally, nonverbal features observable in the video recordings were coded in relation to their corresponding verbal utterances to assess the degree of alignment or misalignment between the two communicative modes.

Results and Discussions

This chapter presents the findings of the study in three sequential sections. The data were drawn from coded verbal and nonverbal transcripts of nine Indonesian EFL learners, coded as P02–P15, performing scripted job interview simulations. Through all the utterances, a total of 122 coded utterances from applicants were analyzed. From all the coded utterances, 98 were assigned a speech act type, including self-promotion, hedging, or clarification request, and the rest is interactional responses such as scripted greetings and closings. Nonverbal co-expression and alignment were coded across all 122 utterances.

1. Speech Acts Performed in Scripted Job Interview Simulations

The first research question asked: "What speech acts do Indonesian EFL learners perform in scripted job interview simulations?" To answer this question, the analysis employed three dominant speech act types and displayed in the coded data as presented in the table below:

Table 3. Distribution of Speech Act Types per Participant

Participant	Self-Promotion	Hedging	Clarification Request	Total Utterances
P02	5	2	1	8
P05	13	1	4	18
P07	3	2	0	5
P08	14	4	4	22
P09	4	2	0	6
P10	8	2	1	11
P12	7	0	1	8
P13	11	2	2	15
P15	2	3	0	5
Total (n)	67	18	13	98
Percentage (%)	68.4%	18.4%	13.3%	100%

As shown in Table 3, self-promotion was the most frequent speech act, accounting for 67 out of 98 coded utterances (68.4%). Hedging constituted 18.4% (n=18), while Clarification Request appeared in 13.3% of utterances (n=13). These findings reveal that no participant relied exclusively on a single act type. However, the proportion of each type varied considerably across individuals. P08 and P05 produced the highest number of self-promotion utterances (14 and 13, respectively), while P15 exhibited a notably higher proportion of hedging relative to total utterances (3 out of 5). P07, P09, and P15 produced no clarification requests. It suggests that the scripted format constrained some participants' ability or willingness to enact inquiry-based acts.

Table 4. Speech Act Distribution Across Interview Stages

Interview Stage	Self-Promotion	Hedging	Clarification Request	Total Utterances
Opening	12 (17.9%)	0 (0%)	0 (0%)	12
Body	38 (56.7%)	17 (94.4%)	4 (30.8%)	59
Closing	17 (25.4%)	1 (5.6%)	9 (69.2%)	27
Total	67	18	13	98

Table 4 reveals a systematic staging of speech act deployment. During the opening stage, all 12 coded utterances were self-promotion. It reflects the conventional function of the opening as a site of professional identity construction. The Body stage contained the greatest overall density of speech acts (n=59), with Self-Promotion (n=38) dominant but Hedging appearing almost exclusively here (n=17, 94.4%). The concentration of hedging in the body is particularly notable. It can be assumed when the interviewees responded to questions about strengths, weaknesses, and challenges, contexts that pragmatically invite mitigating language.

The clarification request acts were distributed differently. Most of this speech act appeared in the closing stage ($n=9$, 69.2%). It existed typically after participants had concluded their self-presentation and used the closing turn to perform what the data reveal as proactive engagement. The function of this speech act is to request information about job scope, training opportunities, or recruitment timelines. This distribution reflects a pragmatic strategy aligned with scripted interview conventions, wherein the closing is recognized as the appropriate slot for candidate-initiated questions.

2. Nonverbal Co-Expression of Speech Acts

The second research question asked: "How are these speech acts co-expressed through nonverbal behavior?" This question can be answered from the coding which identified four principal nonverbal channels. They were gaze behavior ($n=66$ of 121 nonverbal entries), facial expression and smile ($n=70$), prosodic and intonation features ($n=46$), and bodily gesture ($n=42$). These channels were not mutually exclusive; most utterances were coded with combinations of nonverbal features, reflecting the multimodal nature of face-to-face interaction.

Table 5. Nonverbal Co-Expression and Alignment by Speech Act Type

Speech Act Type	Dominant Nonverbal Feature	Prosodic Feature	Alignment n (%)	Misalignment n (%)
Self-Promotion (n=67)	Intermittent downward gaze; smiling; eye contact with gestures	Flat/monotone; occasional emphatic rise	51 (76.1%)	16 (23.9%)
Hedging (n=18)	Downward gaze; subdued expression; smile at end of turn	Tone shift: confident on strengths, subdued on weaknesses	11 (61.1%)	7 (38.9%)
Clarification Request (n=13)	Flat/steady eye contact; minimal facial expression	Flat intonation; minimal rising intonation for questions	13 (100%)	0 (0%)

There were four non-verbal co-expression performed mostly by the applicants: intermittent downward gaze, smile, prosodic, and gesture.

First, Intermittent downward gaze was the most frequently observed nonverbal feature. It occurred with self-promotion utterances. This created a paradoxical disjunction: while verbal content was assertive and face-enhancing, the accompanying gaze enacted deference rather than confidence. Contrastively, clarification request acts were consistently paired with steady and maintained eye contact with the total 13 utterances were coded as aligned. It suggests that when candidates shifted to inquiry-based acts, their gaze behavior more closely approximated English-medium professional norms.

Another non-verbal co-expression perform that was mostly found is smile. Smiling was pervasive throughout the dataset, yet its function varied systematically by speech act type. In aligned self-promotion utterances, smiling co-occurred with eye contact and gesture. It enacted a stance of confident warmth, whereas in hedging utterances, smiling appeared predominantly at turn-end as a face-saving device. A flat or neutral expression,

by contrast, was the defining feature of misaligned utterances. It was found from P10's data who produced a consistent disjunction between competent verbal content and disengaged nonverbal presentation.

Third, non-verbal co-expression was prosodic. Prosodic co-expression was documented in 46 of 121 nonverbal entries. This co-expression appeared with monotone or flat intonation co-occurring with self-promotion as the most frequent pattern. Both were deployed in the performance of the data P09 and P10. Despite well-organized verbal content, these participants' delivery lacked the emphatic stress and intonational variation. It is associated with confident English-medium self-presentation. A contrastive pattern emerged in hedging utterances, where participants such as P08 and P09. They produced within-turn tone shifts. P08 and P09 are more expressive when asserting strengths than subdued when disclosing weaknesses. It reflects multimodal coordination with the verbal concessive-compensatory structure, even against an overall flat prosodic baseline.

Last, gesture that was co-occurred in 42 entries. It appeared most prominently in aligned utterances and frequently paired with eye contact and smiling. Although gesture types were not systematically differentiated in the data, the co-occurrence pattern indicates that gesture functioned as an amplifier of verbal assertiveness. The participants who used gesture consistently producing higher alignment scores. Conversely, misaligned utterances were almost entirely gesture-free. It suggested that nonverbal disengagement was not channel-specific but extended simultaneously across gaze, expression, and bodily movement.

Table 6. Summary of Verbal-Nonverbal Alignment

Alignment Category	n (%)	Associated Speech Act	Representative Nonverbal Pattern
Aligned	95 (78.5%)	All types: highest in Clarification Request (100%)	Smiling + sustained eye contact + gestural emphasis
Partially Misaligned	12 (9.9%)	Hedging (6); Self-Promotion (6)	Flat expression + steady gaze; tone shift at weakness disclosure
Misaligned	13 (10.7%)	Self-Promotion (10–13)	Monotone delivery + gaze aversion throughout; flat expression

Table 6 summarizes the three alignment categories. Most of the utterances (78.5%) were coded as "aligned". It indicates that most participants broadly co-expressed their verbal content with appropriate nonverbal behavior. However, the 20.6% of utterances coded as "partially misaligned" or "misaligned" reveal systematic patterning: misalignment was concentrated in self-promotion utterances (particularly those of P10) and hedging utterances, while clarification request acts achieved 100% alignment. The high alignment of clarification request acts may reflect the relative familiarity of questioning behavior as a pragmatic act, which requires less performative self-regulation than assertive self-presentation.

3. Verbal, Nonverbal, and Prosodic Features Reflecting L1 Sociocultural Transfer

The third research question asked, "How do verbal, nonverbal, and prosodic features reflect L1 sociocultural transfer in Indonesian EFL learners' scripted job interview simulations?" This question requires interpretation of the patterned mismatches between verbal content and nonverbal/prosodic delivery observed in RQ1 and RQ2, viewed through the lens of Indonesian, particularly Javanese-influenced sociocultural norms.

Table 7. Identified L1 Sociocultural Transfer Features

Alignment Category	n (%)	Associated Speech Act	Representative Nonverbal Pattern
Transfer Feature	Verbal Manifestation	Nonverbal/Prosodic Manifestation	Sociocultural Explanation (Javanese/Indonesian Norms)
Downward gaze during self-assertion	Self-Promotion utterances asserting competence and identity	Intermittent downward gaze; avoidance of sustained eye contact	<i>Sungkan</i> (deference) and avoiding display of direct confidence to authority figures; sustained eye contact perceived as rude or challenging
Flat/monotone prosody on assertive content	Formulaic tricolon assertions (e.g., 'responsible, disciplined, hard-working')	Flat intonation; no emphatic stress; monotone delivery	Indonesian oral tradition values even-keel delivery; exaggerated prosody may index arrogance or performativity
Hedging strength disclosure	Epistemic hedges: 'I think,' 'I believe,' concessive structures	Subdued tone, downward gaze when disclosing strengths; smile at turn-end	<i>Rukun</i> (social harmony) and face-saving norms; direct self-praise is face-threatening in Javanese-influenced communication
Clarification Request as a Closing Ritual	Questions about job scope, next steps at interview close	Flat, professional tone; steady gaze; no gestural elaboration	Turn-taking and politeness norms: questions reserved for structured closings; forward-facing inquiry signals commitment without aggression

Discussion

This chapter discusses the findings. It includes the existing theoretical and empirical literature on speech act theory, multimodal communication, and L1 pragmatic transfer in EFL contexts. The discussion is organized around three thematic concerns that cut across the three research questions: (1) the pragmatic competence of the job interview as a genre and how Indonesian EFL learners navigate it; (2) the multimodal constitution of professional identity and its sociocultural dimensions; and (3) the pedagogical implications of the identified transfer features.

1. Pragmatic Competence of the Job Interview and Indonesian EFL Learner Navigation

The distribution of speech act types reveals that participants in this study demonstrated a broadly genre-appropriate understanding of the job interview. This architecture can be understood as a structured communicative event. The concentration of

self-promotion in the opening stage, hedging in the body, and clarification request in the closing stage aligns with the three-part structure described by Bhatia as a negotiated genre (Bhatia, 1993). Each stage carries distinct illocutionary expectations. The ability of Indonesian EFL learners in producing this structure suggests that genre awareness is in minimal performance of speech acts of three stages. The formation of the professional speech acts can be developed through explicit preparation and simulation practice.

However, the high proportion of formulaic utterances such as the utterance "I am responsible, disciplined, and detail-oriented" raises questions about the depth of pragmatic competence enacted. The scripted simulation format naturally privileges formulaic performance, which may explain why speech act selection was appropriate while prosodic and nonverbal enactment lagged. The findings, therefore, support the view of Kasper and Rose that distinguish formulaic (relying on rehearsed) and rich (adapting to the interlocutor in real time) pragmatic performance that pragmatic competence. This is a multi-layered construction and may provide only partial preparation for spontaneous, embodied professional interaction.

2. Multimodal Constitution of Professional Identity and Sociocultural Norms

The most theoretically significant finding of this study lies in the systematic disjunction between assertive verbal content and deferential nonverbal/prosodic delivery. This pattern cannot be adequately explained by reference to L2 proficiency or anxiety alone. It is because the verbal content of most participants was grammatically and pragmatically competent, while the nonverbal features identified are patterned and consistent rather than random. The interpretation advanced here is that these nonverbal features constitute pragmatic transfer from L1 sociocultural norms. The concept of sociocultural transfer in this context aligns with Odlin's (1989) influential framework. It posits that transfer operates not only at the phonological and syntactic levels but also at the pragmatic and discourse levels, including at the level of nonverbal and paralinguistic behavior (Odlin, 1989).

More recently, Ishihara and Cohen (2010) have argued that pragmatic transfer is particularly likely in high-stakes, emotionally charged communicative events, precisely the conditions that a job interview, even in simulation form, creates (Ishihara & Cohen, 2021). The present findings support this claim as employed in the Javanese-Indonesian's sociocultural values of *sungkan* (deference), *rukun* (harmony), and *kalem* (composed restraint) appear to exert significant influence on the nonverbal and prosodic dimensions of participants' performance, even when verbal content has been carefully prepared to approximate English-medium professional norms.

This finding has implications for how professional identity is theorized in L2 contexts. The Indonesian EFL learners in this study appear to be simultaneously enacting two communicative selves: the scripted professional identity (verbally) and the culturally embedded relational self (nonverbally). The tension between these two selves may constitute a productive site for pedagogical intervention.

Notably, clarification request acts achieved 100% alignment, suggesting that when participants shifted from self-presentation to inquiry, the sociocultural pressure to perform

deference was reduced or at least that the act of asking questions did not activate the same cultural scripts of *sungkan* that assertive self-promotion did. This finding is consistent with Dörnyei and Thurrell's observation that interactional acts involving information-seeking tend to be less culturally marked than acts of self-promotion because they are constituted as other-directed rather than self-directed and therefore less vulnerable to face threat from the perspective of the Indonesian sociocultural value system (Dörnyei & Thurrell, 1994).

3. Hedging: Pragmatic Strategy or Sociocultural Obligation?

The finding that hedging appeared almost exclusively in the body stage and primarily in response to questions about strengths, weaknesses, and challenges is consistent with the established pragmatic literature on interview discourse. What is less expected is the systematic use of epistemic hedging within strength disclosure, not only weakness disclosure. In the interview genre, strength disclosure is typically a legitimate self-promotional context where the candidate is explicitly invited to assert competence. Yet, Indonesian EFL learners in this dataset consistently softened their strength assertions with modal hedges such as "I think" and "I believe". In line with this expression, the frequency of hedges such as "usually" and "sometimes" was formulaically assertive.

The patterns performed by Indonesian EFL learner's in this dataset suggests that hedging functions not only as pragmatic uncertainty marking but also as a sociocultural obligation. It also shows a performative enactment of Javanese norms of social modesty that constrains even scripted speech acts. This interpretation aligns with Wierzbicka's cross-cultural analysis of hedging as a culturally variable phenomenon in which speakers of different language backgrounds deploy hedges according to different social logics rather than universal epistemic conventions (Wierzbicka, 1991).

The partially misaligned hedging utterances in which flat expression and monotone delivery co-occurred with verbal concessive-compensatory structures. It suggests that the pragmatic work of hedging extends beyond the verbal channel. The nonverbal and prosodic dimensions of hedging reinforce the interactional message of the verbal hedge. It can be seen through utterance such as: "I am not overclaiming", "I acknowledge my limitations", "I am oriented toward growth rather than superiority". This multimodal coherence in hedging acts demonstrates a sophisticated.

4. Misalignment and Individual Variation

The concentration of misaligned utterances in P10's data warrants specific discussion. P10 produced 13 of the 13 fully misaligned utterances. This pattern consistent across all interview stages and speech act types. This systematic misalignment which is characterized by flat facial expression and monotone delivery throughout, is unlikely to reflect L1 sociocultural transfer alone. It may also reflect higher levels of anxiety by performing lower L2 oral proficiency or reduced familiarity with the simulation format. The co-occurrence of verbal competence with nonverbal expression across all of P10's utterances suggests a dissociation between prepared verbal content and spontaneous nonverbal regulation. It reflects a dissociation that may be heightened under conditions of evaluative stress.

At the other end of the spectrum, P05 and P08 produced the highest volumes of self-promotion utterances. They demonstrated the highest rates of aligned delivery, with frequent combinations of smiling, eye contact, and gesture. These participants appeared to have internalized or successfully performed a multimodal repertoire closer to English-medium professional norms. Whether this reflects greater L2 pragmatic awareness or individual communicative style cannot be determined from the available data. Whereas, the variation underscores the importance of treating Indonesian EFL learners as a heterogeneous population rather than a monolithic cultural category.

5. Pedagogical Implications

The finding of this study carries several concrete implications for EFL instruction oriented toward professional communication. First, they suggest that explicit instruction in multimodal professional norms is needed alongside the teaching of speech act content. Learners who can select and structure appropriate speech acts may still underperform in authentic professional contexts if the nonverbal and prosodic dimensions of their delivery do not align with target-culture norms.

Second, the findings suggest that awareness-raising activities addressing L1 sociocultural transfer in professional contexts may be valuable. The instructors might engage learners in cross-cultural pragmatic reflection rather than treat cultural norms as obstacles to L2 performance. It is an activity that foregrounds the cultural logics underlying their own communicative behavior and invites critical comparison with the norms of the target professional discourse community. Such activities can support learner agency without requiring wholesale abandonment of L1 cultural identity.

Third, the high alignment of clarification request acts suggests that inquiry-based communicative acts, where the learner adopts an other-directed, information-seeking stance, may offer a productive entry point for developing confidence in L2 professional interaction. The asymmetry between alignment in Clarification Request (100%) and Self-Promotion (76.1%) suggests that learners may benefit from extended practice specifically in the multimodal performance of assertive, self-referential acts, which appear to be the communicative contexts most susceptible to L1 transfer effects.

Finally, the scenario for scripted job interview simulation as a pedagogical format warrants critical reassessment. The format may have provided participants with sufficient preparation time to produce verbally appropriate content. However, it may have reduced the ecological validity of the performance as a measure of pragmatic competence. It privileges rehearsed verbal production but leaves spontaneous nonverbal and prosodic regulation underprepared. To design comprehensive design, future instruction and assessment might benefit from combining scripted preparation phases with semi-spontaneous performance opportunities that foreground multimodal rather than purely verbal dimensions of professional communication.

Conclusions

This discussion has argued that the findings of this study cannot be adequately explained by reference to L2 proficiency deficits alone. Rather, the patterned disjunction between verbal and nonverbal-prosodic dimensions of Indonesian EFL learners' job interview simulations reflect the operation of deep-seated sociocultural values. This value exists in the data set are the acts of smiling, downward gaze, and nodding, that shape the embodied performance of professional identity. It shows systematic, culturally coherent, and largely independent of verbal competence. The pedagogical implication we can draw from the results is the more holistic conception of communicative competence that encompasses multimodal, prosodic, and sociocultural dimensions alongside the verbal speech act content as the primary focus of EFL instruction.

Referensi

- Ageeva, J. (2016). COMMUNICATION SKILLS AT JOB INTERVIEW: PEDAGOGICAL INSIGHT INTO THE PROBLEM. *The Education and Science Journal*, (1), 109–117. <https://doi.org/10.17853/1994-5639-2016-1-109-117>
- Al-Issa, A. (2003). Sociocultural transfer in L2 speech behaviors: Evidence and motivating factors. *International Journal of Intercultural Relations*, 27(5), 581–601. [https://doi.org/10.1016/S0147-1767\(03\)00055-5](https://doi.org/10.1016/S0147-1767(03)00055-5)
- Alviar, C., Kello, C. T., & Dale, R. (2023). Multimodal coordination and pragmatic modes in conversation. *Language Sciences*, 97, 101524. <https://doi.org/10.1016/j.langsci.2022.101524>
- Austin, L. (1962). *HOW TO DO THINGS WITH WORDS*.
- Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman.
- Blake, H. L., Hewat, S., & Spencer, E. (2011). Native Speaker and Non-Native Speaker Communication During Job Interviews. *Asia Pacific Journal of Speech, Language and Hearing*, 14(3), 171–178. <https://doi.org/10.1179/jslh.2011.14.3.171>
- Burns, A. (2019). CONCEPTS FOR TEACHING SPEAKING IN THE ENGLISH LANGUAGE CLASSROOM. *LEARN Journal: Language Education and Acquisition Research Network*, 12(1), 1–11. <https://doi.org/10.61587/mmit.uz.vi.31>
- Chen, R. T.-H. (2022). Effects of Global Englishes-oriented pedagogy in the EFL classroom. *System*, 111, 102946. <https://doi.org/10.1016/j.system.2022.102946>
- Darong, H. C. (2024). Teaching speech acts in EFL context. *Frontiers in Education*, 9, 1423498. <https://doi.org/10.3389/feduc.2024.1423498>
- Deda, N. (2013). The role of Pragmatics in English Language Teaching. Pragmatic Competence. *Academic Journal of Interdisciplinary Studies*. <https://doi.org/10.5901/ajis.2012.v2n4p63>
- Dörnyei, Z., & Thurrell, S. (1994). Teaching conversational skills intensively: Course content and rationale. *ELT Journal*, 48(1), 40–49. <https://doi.org/10.1093/elt/48.1.40>
- 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>

- Ho, W. Y. J. (2019). Developing Professional Communication: The Construction of a Multimodal Understanding of Job Interviews. *Languages*, 4(1), 5. <https://doi.org/10.3390/languages4010005>
- Ishihara, N., & Cohen, A. D. (2021). *Teaching and Learning Pragmatics: Where Language and Culture Meet* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003168188>
- Kamaliah, N., Kasim, U., & Azis, Z. A. (2018). *USE OF THE THREE-STEP INTERVIEW TECHNIQUE IN TEACHING ESL SPEAKING*.
- Kasper, G., & Rose, K. R. (2002). Introduction to Second Language Pragmatic Development. *Language Learning*, 52(s1), 1–11. <https://doi.org/10.1111/j.1467-1770.2002.tb00022.x>
- Krishnan, I. A., Ramalingam, S., Kaliappen, N., Uthamaputhran, S., Suppiah, P. C., Mello, G. D., & Paramasivam, S. (2021). Graduate employability skills: Words and phrases used in job interviews. *Australian Journal of Career Development*, 30(1), 24–32. <https://doi.org/10.1177/1038416220980425>
- Leech, G. N. (1983). *Principles of pragmatics*. Longman.
- Marsakawati, N. P. E., Mujiyanto, J., & Agustien, H. I. R. (2019). An Indonesian EFL Learner's Use of Evaluative Language and Gestures in a Spoken Persuasive Presentation: A Case Study. *The Journal of AsiaTEFL*, 16(1), 401–410. <https://doi.org/10.18823/asiatefl.2019.16.1.30.401>
- Morell, T. (2018). Multimodal competence and effective interactive lecturing. *System*, 77, 70–79. <https://doi.org/10.1016/j.system.2017.12.006>
- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework* (0 ed.). Routledge. <https://doi.org/10.4324/9780203379493>
- Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524537>
- Otłowski, M. (2008). *Preparing University EFL Students for Job Interviews in English: A Task-Based Approach*.
- Peng, J.-E., Zhang, L., & Chen, Y. (2017). The Mediation of Multimodal Affordances on Willingness to Communicate in the English as a Foreign Language Classroom. *TESOL Quarterly*, 51(2), 302–331. <https://doi.org/10.1002/tesq.298>
- Reeves, J. (2010). Teacher learning by script. *Language Teaching Research*, 14(3), 241–258. <https://doi.org/10.1177/1362168810365252>
- Sbisa, M. (2002). Speech acts in context. *Language and Communication*, 11, 421–436.
- Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139173438>
- Stivers, T., & Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica*, 2005(156), 1–20. <https://doi.org/10.1515/semi.2005.2005.156.1>
- Takahashi, S. (1996). Pragmatic Transferability. *Studies in Second Language Acquisition*, 18(2), 189–223. <https://doi.org/10.1017/S0272263100014881>
- Takaya, K. (2024). Taiwanese EFL Learners' Use of Learning Strategies in Developing Job Interview Skills. *Journal of International and Comparative Education*, 13(2), 93–109. <https://doi.org/10.14425/jice.2024.13.2.0913>

- Wierzbicka, A. (1991). *Cross-Cultural Pragmatics: The Semantics of Human Interaction*. Mouton de Gruyter.
- Wilmes, S. E. D., & Siry, C. (2021). Multimodal Interaction Analysis: A Powerful Tool for Examining Plurilingual Students' Engagement in Science Practices: Proposed Contribution to RISE Special Issue: Analyzing Science Classroom Discourse. *Research in Science Education*, 51(1), 71–91. <https://doi.org/10.1007/s11165-020-09977-z>
- Yan, Y. (2022). *The Effect of Pragmatic Competence on the Communicative Competence of Second Language Learners*: 2022 3rd International Conference on Language, Art and Cultural Exchange (ICLACE 2022). <https://doi.org/10.2991/assehr.k.220706.074>
- Yu, C. (2013). Two interactional functions of self-mockery in everyday English conversations: A multimodal analysis. *Journal of Pragmatics*, 50(1), 1–22. <https://doi.org/10.1016/j.pragma.2013.01.006>
- Yu, M. (2011). Learning how to read situations and know what is the right thing to say or do in an L2: A study of socio-cultural competence and language transfer. *Journal of Pragmatics*, 43(4), 1127–1147. <https://doi.org/10.1016/j.pragma.2009.09.025>
- Zarrinabadi, N., & Khodarahmi, E. (2017). L2 Willingness to Communicate and Perceived Accent Strength: A Qualitative Inquiry. *Journal of Intercultural Communication Research*, 46(2), 173–187. <https://doi.org/10.1080/17475759.2017.1301981>