rudn.tlcjournal.org
doi: 10.22363/2521-442X-2026-10-3-32-48

Original Research

Negotiating with scripted AI personas: Task-specific pragmatic performance among International Management MA students

Dmitry S. Khramchenko[1*]

[1]MGIMO University, Moscow, Russia

Generative AI can provide immediate dialogue for English for Specific Purposes (ESP) simulations, but its use as a resistant negotiation partner remains underexamined. This small-scale mixed-methods classroom study investigated how Russian MA students in International Management realised two contrasting sets of interactional requirements and how they evaluated their performance afterwards. Thirty-two participants each completed two timed written negotiations concerning a supply-chain crisis: a direct/task-oriented condition and an indirect/face-oriented condition. The dataset comprised 64 chat transcripts containing 45,312 student-authored words, 32 post-task reflective essays, and transcript annotations. Holm-adjusted Wilcoxon signed-rank tests found higher rates of lexical hedges, syntactic mitigators, and face-saving acts in the indirect condition, and higher rates of lexical intensifiers, direct obligation markers, and numerically specified financial justifications and concessions in the direct condition; subjectiviser rates did not differ significantly. Codebook thematic analysis identified transcript-based self-observation for 30 participants, reported difficulty under task pressure for 28, and responses to scripted persona demands for 24. The findings show that the same learners converted contrasting, explicitly stated communicative requirements into different linguistic choices and used their transcripts to examine those choices and question the cultural credibility of the personas. The study presents a feasible AI-mediated elicitation-and-reflection task for ESP instruction at MA level.

Keywords: *Generative AI, English for Specific Purposes (ESP), intercultural pragmatics, metacognition, business English, prompt-condition effects*

Article history:
Submitted May 2, 2026
Revised August 20, 2026
Accepted September 16, 2026

CRediT Author Statement:
Dmitry S. Khramchenko: Conceptualisation, Methodology, Validation, Data Collection, Formal Analysis, Investigation, Resources, Writing — Original Draft, Writing — Review & Editing, Visualisation.

Conflict of interest:
The author declared no conflict of interest.

Data availability statement:
The data supporting this study's findings are available from the corresponding author upon reasonable request.

Funding:
No funding was reported for this research.

Doi:
10.22363/2521-442X-2026-10-3-32-48

For citation:
Khramchenko, D. S. (2026). Negotiating with scripted AI personas: Task-specific pragmatic performance among International Management MA students. *Training, Language and Culture, 10*(3), 32—48.

This is an open access article distributed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which allows its unrestricted use for non-commercial purposes, subject to attribution. The material can be shared/adapted for non-commercial purposes if you give appropriate credit, provide a link to the license, and indicate if changes were made.

1. INTRODUCTION

Generative Artificial Intelligence has become part of ESP education, where it can draft correspondence, provide feedback, summarise professional docu-ments, and simulate workplace interaction. In Business English, these capabilities make AI a convenient partner for practising communication outside authentic professional settings. Yet the same tendency towards cooperation that makes AI convenient may limit its usefulness for negotiation training. Research on sycophancy indicates that generative-AI assistants may agree excessively with users' stated positions or adapt their replies to please the user (Malmqvist, 2025). A learner who receives quick agreement is not required to defend an unpopular proposal, respond to rejection, repair a damaged professional relationship, or reconsider an unsuccessful communicative choice. Negotiation practice therefore requires an interlocutor capable of sustained resistance as well as fluent dialogue. This requirement is

* Corresponding author
© Dmitry S. Khramchenko 2026
Licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

especially relevant to MA students studying Business English because pragmatic competence cannot be reduced to knowledge of polite expressions. Learners may know the conventional wording of an apology, hedge, or concession and still have difficulty selecting suitable language after an interlocutor rejects their proposal. Research distinguishes knowledge of pragmatic conventions from the capacity to apply that knowledge during interaction (Taguchi & Roever, 2017). In business negotiation, learners must attend simultaneously to the commercial objective, the professional relationship, and the interlocutor's current response (Kankaanranta & Louhiala-Salminen, 2010; Planken, 2005). The pedagogical problem is therefore not simply whether these learners possess appropriate expressions, but whether they can convert an interactional requirement into language while a negotiation is in progress.

Devil's advocacy supplies one rationale for creating such resistance. Research on argumentation and computer-mediated learning describes devil's advocacy as structured opposition in which an interlocutor questions claims, provides counterarguments, and requires speakers to defend their reasoning (Chiang et al., 2024; Wu, 2026). The present study adapts this adversarial principle to an ESP crisis-negotiation task: the AI interlocutor is instructed to reject the initial proposal and withhold agreement until the learner satisfies specified conditions. Resistance is therefore built into the task, giving participants a reason to justify, modify, or reformulate their proposals instead of relying on an immediately cooperative response.

To create two clearly distinguishable interactional demands, the persona prompts used three scales from Meyer's (2014) *Culture Map*. The US-labelled persona represented a direct, task-focused orientation and required financial evidence and concrete concessions. The Japanese-labelled persona represented an indirect, relationship-focused orientation and required relational repair before discussion of revised terms. These labels identified two deliberately polarised classroom stimuli. They did not indicate that US or Japanese professionals communicate uniformly in these ways. As discussed further, national classifications cannot provide a complete account of situated business communication, which varies within societies and between particular interlocutors (Holliday, 2011; McSweeney, 2002; Piller, 2025).

The study also responds to a gap in generative-AI language research. Existing work has concentrated mainly on writing assistance, feedback, revision, and argumentative writing (Barrot, 2023; Guo et al., 2022; Li et al., 2025; Su et al., 2023), while resistant, real-time, text-based negotiation for MA students in ESP courses has received much less empirical attention. Research has also identified limitations in the pragmatic and cultural suitability of AI output (Lee & Cook, 2024; Sydorenko et al., 2024; Tao et al., 2024).

A study of resistant AI negotiation must therefore examine both the language produced by learners and their evaluation of the AI personas.

Against this background, the present study examines task-specific pragmatic performance under two contrasting prompt conditions. It does not ask whether students acquired US or Japanese communication styles, nor does it assess lasting pragmatic development. Instead, it asks how the same participants realised two explicit sets of interactional requirements while responding to resistance in a timed negotiation. Because every participant completed both conditions, the analysis compares pragmatic resource use within individual learners and reduces the influence of stable differences in general proficiency, vocabulary, and personality. The subsequent transcript-annotation and reflective stages examine how participants described their own language choices, task difficulties, and the cultural credibility of the AI output.

The study addresses the following research questions:

1. How do the frequencies of coded pragmatic devices in the same learners' written output differ between the direct/task-oriented and indirect/face-oriented prompt conditions?

2. What do learners report about their linguistic choices, task difficulties, and the cultural credibility of the AI personas during post-task reflection?

The study contributes paired evidence of how MA students in an ESP course converted contrasting communicative requirements into written negotiation language. It also examines a classroom procedure in which resistant AI interaction produces a transcript that can subsequently support linguistic self-examination and criticism of culturally stereotyped output. The intended contribution is therefore a task design for eliciting and examining pragmatic performance in a small, clearly specified learner population. Claims concerning lasting learning, transfer, or relative pedagogical effectiveness require subsequent comparative and longitudinal research.

2. THEORETICAL BACKGROUND

2.1. Cross-cultural pragmatics and face in business English

Research on English as a business lingua franca indicates that grammatical accuracy alone does not ensure effective international business communication. Speakers also need knowledge of the business setting, relevant genres, and communication strategies (Kankaanranta & Louhiala-Salminen, 2010; Malyuga, 2024). These demands are especially apparent in negotiation, where particular moves (such as rejecting a proposal, announcing a delay, or imposing a price increase) may threaten an interlocutor's public self-image or freedom of action (Brown & Levinson, 1987; Malyuga, 2025). Speakers can reduce this interpersonal risk through mitigation.

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

Research on requests and apologies distinguishes directness strategies, lexical and syntactic downgraders, and supportive moves (Blum-Kulka et al., 1989). In this article, pragmatic mitigation devices (PMDs) refer to linguistic forms and discourse moves that reduce illocutionary force or interpersonal risk. Negotiators may also intensify demands, express obligation, repair rapport, or support proposals with business data. Planken's (2005) comparison of professional and aspiring negotiators found differences in relational talk and pronoun use during English-medium sales negotiations. Pragmatic performance therefore concerns the suitability of a linguistic choice at a particular point in the exchange, not simply the presence of a polite formula.

Meyer's (2014) practitioner-oriented *Culture Map* supplies the labels used in the two prompts. It compares national business practices on eight scales, three of which inform the present task: Communicating (low-context to high-context), Evaluating (direct to indirect negative feedback), and Trusting (task-based to relationship-based). In Meyer's (2014) comparison, the United States is closer to the low-context, direct-feedback, and task-based ends, whereas Japan is closer to the high-context, indirect-feedback, and relationship-based ends. These contrasts give students a clear reason to vary directness, mitigation, relational repair, and business justification during a negotiation.

However, these scales are not empirical descriptions of uniform communication within entire national populations. Research critical of fixed national classifications stresses variation within societies and the production of meaning in situated interaction (Holliday, 2011; McSweeney, 2002; Piller, 2025). BELF research likewise presents English as a practical means of completing business tasks in multilingual settings, where genre knowledge and clarity may matter more than conformity to native-speaker norms (Kankaanranta & Louhiala-Salminen, 2010). The American and Japanese AI personas are therefore deliberately exaggerated teaching stimuli and scripted task conditions. They do not represent authentic business discourse or individual speakers from either country.

2.2. Strategic competence under task pressure

Bachman and Palmer (1996) define strategic competence as metacognitive strategies for goal setting, assessment, and planning during language use. During negotiation, these processes occur together: speakers assess an interlocutor's response, select suitable linguistic resources, and plan the next turn while pursuing a business objective (Amelkin et al., 2026). Taguchi and Roever (2017) distinguish knowledge of pragmatic conventions from the ability to use that knowledge during interaction. Empirical research on simulated pragmatic tasks supports this distinction: L2 learners reported cognitive, metacognitive, and pragmatic strategies while completing requests and refusals (Bi, 2021). These findings help explain why knowledge of a hedge or apology does not ensure its timely use after an interlocutor rejects a proposal.

Kasper and Rose (2002) identify contextualised communicative practice as an important part of L2 pragmatic development. The two prompts in the present study establish different acceptance rules and interpersonal expectations, so learners must respond to resistance while pursuing the same business objective. The relevant outcome in this study is situated pragmatic performance during the task. Evidence of lasting development would require pre-task and post-task measurement. The analysis therefore examines the language produced under the two scripted conditions and the students' subsequent reflections.

2.3. Generative AI in L2 pragmatics: From assistant to resistant interlocutor

LLMs can produce immediate, context-responsive dialogue, making real-time text-based classroom simulations feasible (Kasneci et al., 2023; Teoh et al., 2025). Nevertheless, research in generative-AI language education has concentrated mainly on writing (Vovou, 2023). Li et al.'s (2025) systematic review of 144 empirical studies found that writing accounted for 42.4% of the reviewed publications, while speaking received considerably less attention. Examples include analyses of ChatGPT as an L2 writing aid (Barrot, 2023), AI-generated writing feedback (Escalante et al., 2023), and chatbot-supported argumentative writing (Guo et al., 2022; Su et al., 2023). Earlier research on AI-supported oral practice examined speaking performance, anxiety, affective responses, and self-regulation (El Shazly, 2021; Kang, 2022; Qiao & Zhao, 2023). Those studies concerned chatbots, AI avatars, or AI-based instructional applications, and their outcomes centred on oral proficiency and affective variables. They did not investigate the pragmatic choices produced during resistant business negotiation with scripted LLM personas. Resistant, real-time, text-based ESP negotiation therefore remains comparatively underexamined.

The available literature also points to a tension between readily available conversational practice and the demands of negotiation training. Kohnke et al. (2023) present conversational practice as a potential benefit of ChatGPT for language learners. For the present task, however, an interlocutor that accepts every proposal cannot test a learner's response to rejection. Research on sycophancy indicates that LLMs can agree excessively with users' stated positions (Malmqvist, 2025), which may reduce sustained counterargument. Scripted resistance offers one response to this problem, but its pedagogical value also depends on pragmatic suitability. Lee and Cook (2024) found that ChatGPT's apologies were shorter, used fewer pragmatic strategies, and received lower human ratings for

appropriateness, politeness, and language use. Chen et al. (2024) found that AI-generated conversations resembled human-written conversations on most tested pragmatic features but displayed greater formality and syntactic diversity. These studies support checking AI output for each speech act and communicative setting.

Questions of cultural suitability arise as well. Tao et al. (2024) compared five widely used GPT systems with nationally representative values data and found outputs closer to values reported in English-speaking and Protestant European societies. Culturally specific prompts improved correspondence with country data for 71—81% of countries and territories in the later systems. Tao et al.'s (2024) evidence concerns cultural values. Its relevance here is the need to scrutinise culturally framed AI responses and avoid presenting them as authentic business discourse. Together with the criticism of fixed national classifications in Section 2.1, this evidence supports explicit debriefing and student criticism of the personas.

Against this background, the present study uses two scripted prompt conditions with different interpersonal expectations, response demands, and acceptance rules. Both require the LLM to reject the initial proposal, giving learners a reason to defend, modify, or rephrase it. The pedagogical rationale rests on this resistance and the subsequent examination of chat transcripts. It does not depend on the cultural authenticity of the personas. The study therefore examines whether the two conditions elicit different pragmatic-marker frequencies and how students evaluate their own responses and the AI's cultural credibility after the task.

3. METHODOLOGY

3.1. Research design

The study used an exploratory mixed-methods design with repeated measures. Each participant completed both the direct task-oriented condition (Condition A) and the indirect face-oriented condition (Condition B). This permitted paired comparison of each participant's pragmatic performance in the two conditions and reduced variance attributable to stable individual differences. The prompts differed simultaneously in nationality label, interpersonal style, concession rules, and response requirements. Consequently, the study compares two bundled prompt conditions and cannot identify the independent effect of culture or any single prompt feature. The quantitative analysis examined pragmatic devices in students' written chat output, while the qualitative analysis examined their post-task reflective essays and transcript annotations.

3.2. Participants and research context

The study was conducted at MGIMO University (Moscow, Russia) within the MA programme in International Management. The sample comprised 32 master's students from four seminar groups. Participants gained admission through the MGIMO University Entrance Examination in a Foreign Language for MA Programmes or through an accepted C1/C2 international certificate, such as Cambridge CAE or CPE. The internal multi-skill examination was administered at admission before the study, and its formal minimum threshold was CEFR B2. The MA programme's target teaching and assessment level was C1. These institutional requirements describe the programme's admission minimum and intended teaching level. They do not verify that every participant had attained C1 at the time of the study.

The programme's language requirements did not establish equivalent competence in pragmatics, negotiation, persuasion, or intercultural communication (Roever, 2021). All participants had completed introductory coursework in Business English and Corporate Communication. They had received no formal instruction in using Generative AI for strategic negotiation practice, intercultural management, politeness theory, or crisis-negotiation genres. The unfamiliar task may therefore have assessed adaptation to a new negotiation format in addition to pragmatic performance.

All participants self-reported Russian as their only first language. The sample comprised 19 women and 13 men aged 22—27 years (M = 23.8, SD = 1.4). The AI negotiation activity formed part of the required ESP syllabus. Use of students' transcripts, annotations, and essays for research was voluntary and had no effect on course grades. The article's author was also the course instructor. Students were informed that they could decline research participation and complete an equivalent analytical writing task. All 32 students consented to research use of their data.

At the time of data collection, no formal ethics review, approval, or written exemption was sought or issued for this study. To the author's knowledge, the university did not have a dedicated ethics committee for standard, non-sensitive pedagogical classroom research of this kind. No formal institutional determination or waiver was obtained. The research procedures followed the ethical guidelines for educational research issued by the British Educational Research Association (2018). All participants gave informed consent before their anonymised transcripts and essays were included in the analysis.

Data were pseudonymised during collection and de-identified before analysis. Student chats were transmitted through the OpenRouter web interface under pseudonymised session identifiers. No student names or email addresses appeared in the prompts or chat content. The consent form informed students that third-party logging could occur under OpenRouter's data-retention policy. Raw transcripts were stored on an encrypted server and were not deposited in a public repository.

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

3.3. Instruments and AI persona prompts

3.3.1. The business scenario

The study used a business-to-business supply crisis relevant to international management. Students assumed the position of Export Account Manager for a logistics supplier. They had to inform a major international client that geopolitical supply-chain disruption would delay delivery by four weeks and increase the price by 12%. Their objective was to obtain agreement to the revised terms while retaining the client contract.

3.3.2. AI prompt engineering

The study used the OpenAI system listed on OpenRouter as openai/gpt-5.2 during the data-collection period from 16 February to 6 March 2026. The OpenRouter dashboard displayed the version string gpt-5.2-2025-12-11. Server-side logs were unavailable, so use of this exact provider route for every chat could not be independently confirmed.

Students accessed the system through the OpenRouter web chat interface (https://openrouter.ai/chat), using accounts created for the study. No custom software or API scripting was used. Each negotiation began in a new chat. The persona instructions were submitted as the first user message and were not placed in the system-message field.

The chat interface displayed a temperature setting of 0.2 for all participants. All other generation settings remained at the platform defaults and were not explicitly controlled. All 32 accounts displayed the same AI identifier and temperature setting, which the instructor checked during the pre-task briefing. Server-side logs confirming provider-side enforcement of the temperature setting could not be recovered.

The persona prompts are reproduced exactly as used during data collection.

Prompt A (low-context/direct persona): '*Act as the Procurement Director of a US-based multinational corporation. You are currently in a live chat with your supplier. You are highly direct. You value efficiency and separate the relationship from the business task. If the supplier presents bad news, push back aggressively. Demand immediate financial data and concrete solutions. Do not accept vague apologies. Concession rule: do not accept the initial 12% price increase and 4-week delay. However, you are permitted to negotiate and explicitly accept a compromise (e.g., an 8—10% increase or a 2-week delay) only if the supplier provides a logical data-driven financial explanation or offers a tangible transactional concession (e.g., covering shipping costs or offering future rebates). Once they do this, explicitly state your agreement to the new terms. Keep your responses under 100 words. Do not break character'.*

Prompt B (high-context/indirect persona): *'Act as a Senior Partner at a Japanese manufacturing conglomerate. You are currently in a live chat with your supplier. You value hierarchy and long-term relational harmony. 'Face' is very important to you. If the supplier presents bad news, do not show direct anger or say 'no' directly. Instead, simulate conversational 'silence' by using ellipses (...) and completely ignoring any questions about numbers or deadlines. Express disappointment subtly, emphasising the damage to mutual trust. Require the supplier to demonstrate extreme respect before you discuss numbers. Concession rule: do not accept the initial 12% price increase and 4-week delay. However, you are permitted to negotiate and explicitly accept a compromise (e.g., an 8—10% increase or a 2-week delay) only if the supplier successfully repairs the relationship and emphasises long-term harmony. Once this respect is demonstrated, explicitly state your agreement to the new terms. Keep your responses under 100 words. Do not break character'.*

Students pasted these instructions into the chat interface and could therefore read the persona descriptions and concession rules before each negotiation. Their responses may reflect both the AI's replies and their prior knowledge of the explicit task requirements.

A post hoc content check examined 20 randomly selected AI turns from each condition. None of the 20 sampled turns from Condition A contained a relational preamble, and none of the 20 sampled turns from Condition B contained a direct negative evaluation. These findings apply only to the sampled turns and do not establish complete persona consistency throughout all 64 conversations.

Students were instructed to regenerate a response or begin a new chat if the AI departed from the persona. The stated criteria were: (a) apologising for being an AI or otherwise breaking character; (b) premature or unjustified acceptance that violated the concession rules; or (c) a clear violation of the persona's style rules, such as relational preambles from the direct persona or an explicit refusal from the indirect persona. At least one response was regenerated in seven of the 64 conversations: four in Condition A and three in Condition B.

The chat interface replaced the original responses, and no separate log was kept. Neither the exact number nor the content of the regenerated responses can therefore be recovered. Students made the initial regeneration decisions during the negotiations. The instructor subsequently reviewed the retained sequences and identified no additional persona departures in them. Because the removed responses were unavailable, the instructor could not independently verify their content or the students' initial decisions. Only the final uninterrupted sequence from each affected conversation was analysed, and excluded responses were not included in the outcome coding. Removing persona-inconsistent turns may have increased the apparent adherence of the retained corpus to the scripted personas. This possible curation bias should be considered when interpreting the results.

3.4. Procedure

The study took place over three weeks during scheduled ESP classes under a common procedure. Phase 1 comprised the task briefing and AI interaction in Week 1. During a single 90-minute session, students completed two synchronous, 20-minute text-based negotiations. Condition order was counterbalanced to distribute possible order, fatigue, and priming effects between the two sequences. Participants were randomly assigned to two equal groups. Group 1 (n = 16) completed Condition A (Direct/Task-oriented) first and Condition B (Indirect/Face-oriented) second. Group 2 (n = 16) completed the conditions in the opposite order. Students were instructed to reach a resolution or temporary agreement within the allotted time. All chat logs were subsequently exported and saved.

The instructor conducted a 20-minute pre-task briefing. Students were shown how to paste the persona prompt into the chat interface and provide the scenario information needed to begin the negotiation. They were told that task acceptance required an explicit agreement on revised delivery terms within the 20-minute period. They were also instructed to consider the pragmatic suitability of their responses. The instructor provided no vocabulary or strategy assistance once a negotiation had begun. Immediately after each chat, the instructor conducted a five-minute whole-class discussion in which students described their initial reactions.

Phase 2 comprised structured self-annotation in Week 2. Students received their transcripts, a Pragmatic Evaluation Rubric, and a one-page annotation guide. The instructor explained the three annotation categories and answered procedural questions. Students then colour-coded their transcripts to identify: (a) communicative breakdowns in which the AI rejected their approach; (b) pragmatic mitigation and aggravation devices in their own responses; and (c) instances in which they consciously changed their tone in response to the specific AI persona. The instructor did not assess the linguistic accuracy of these annotations. The annotations were used as participants' reflections on their own performance, not as a substitute for researcher coding.

Phase 3 comprised post-task reflection in Week 3. Students submitted a 500-word reflective essay addressing whether the AI had been too accommodating, whether its responses appeared culturally authentic, the reasons for their rhetorical choices, and differences in their communicative strategies between the Direct and Indirect conditions. Because these topics were specified in advance, the qualitative findings represent responses to a structured reflective task. They do not constitute spontaneous reports of students' thinking.

The reflective task invited students to criticise stereotypical features in the AI's output. At the close of Week 3, the instructor conducted an ethical debriefing. Students were informed that the personas were stylised classroom stimuli and did not constitute empirical descriptions of US or Japanese business communication. The class discussed national-cultural generalisations and the production of meaning within specific business interactions. The debriefing addressed the risk of reinforcing national stereotypes and clarified that the AI's output should not be understood as cultural fact. Because students encountered the stylised personas during Weeks 1 and 2, the possibility of short-term reinforcement before the debriefing cannot be excluded.

3.5. Data collection and analytical framework

The resulting dataset comprised 64 negotiation transcripts containing 45,312 student-authored words, 32 post-task reflective essays, and participants' transcript annotations.

Each negotiation was coded for scripted task acceptance. A transcript was coded as accepted only when the AI persona's final substantive message expressed explicit agreement to a compromise permitted under the persona instructions, such as an 8—10% price increase or a two-week delay, after the student had satisfied the specified concession requirements. Agreement to the original 12% increase and four-week delay was classified as a persona departure and was not counted as acceptance. A transcript was coded as not accepted if the final substantive message contained no explicit agreement or if the conversation ended without resolution. Two independent raters classified all 64 outcomes and reached complete agreement (Cohen's $\kappa = 1.00$).

Scripted task acceptance indicates whether a student obtained an AI response that satisfied the programmed concession requirements. It does not demonstrate authentic negotiation competence, pragmatic effectiveness, or likely success in professional communication. Because students could read the concession rules before beginning each negotiation, this outcome also reflects their response to explicit task information.

3.5.1. Quantitative corpus analysis

The 64 transcripts were processed using UAM CorpusTool (O'Donnell, 2009). Student-authored output was manually coded using the seven-category scheme presented in Table 1. The mitigation and aggravation categories were based partly on prior research in cross-cultural pragmatics (Blum-Kulka et al., 1989). Numerically specified financial justifications and concessions were an author-developed category for numerically specified business reasoning.

The participant-condition record was the statistical unit. The communicative move was the segmentation unit and was defined as one speaker turn or a functionally complete segment within a longer turn. Longer turns were

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

divided when the pragmatic function changed. Depending on the category, coding was conducted at the word or phrase, clause, or move level.

Lexical hedges were words or phrases that reduced the pragmatic force of a proposition, such as *slightly, perhaps, a bit,* and a *minor delay.*

Syntactic mitigators were complete conditional or interrogative clauses that increased distance from a face-threatening act, such as *I was wondering if...* and *Would it be possible...* . Incomplete markers such as *maybe* were coded as lexical hedges.

Subjectivisers were phrases that explicitly presented a claim as personal or corporate opinion, such as *From our perspective...* and *We believe...* . A phrase was coded as a subjectiviser only when it presented the claim as opinion. A modifier that merely reduced propositional force was coded as a lexical hedge. Both codes could be assigned when both functions occurred, as in *I think maybe...* .

Lexical intensifiers were words or fixed phrases that increased certainty or pragmatic force, such as *absolutely* and *completely unacceptable.*

Direct obligation markers were modals, imperatives, or comparable expressions that imposed a required action, deadline, or obligation on the interlocutor, such as *you must* and *confirm immediately.*

Face-saving acts were moves that acknowledged the interlocutor's concerns or damage to trust or the business relationship and attempted to restore goodwill. Formulaic apologies without relational content were excluded. For example, *We value our partnership and understand the effect this delay may have on your business* qualified as a face-saving act, while an isolated *We apologise for the delay* did not.

Numerically specified financial justifications and concessions were moves that contained explicit numerical information and either explained the business basis for a proposal or specified a financial concession or counter-offer. General references to costs without figures were excluded. Moves with both transactional and relational functions could receive both codes.

More than one category could be assigned to the same text. Each occurrence of a word-, phrase-, or clause-level device was counted. Move-level categories were recorded no more than once within a single move.

Two researchers — the article's author and a second coder holding a Cand. Sci. (Philology) degree — independently coded 13 of the 64 transcripts (20.3%). The subset contained 522 communicative moves. For each move, the presence or absence of all seven categories was recorded, producing 3,654 binary decisions.

Cohen's κ was calculated separately for each category: numerically specified financial justifications and concessions, κ = .91, agreement = 95%; lexical intensifiers, κ = .88, agreement = 94%; direct obligation markers, κ = .87, agreement = 93%; lexical hedges, κ = .86, agreement = 93%; syntactic mitigators, κ = .83, agreement = 92%; face-saving acts, κ = .78, agreement = 90%; and subjectivisers, κ = .76, agreement = 89%. The unweighted mean of the seven category-specific coefficients was κ = .84, and the unweighted mean agreement was 92.3%. Disagreements occurred mainly when coders distinguished relational repair from formulaic politeness and explicit opinion statements from propositional hedging. All disagreements were resolved through discussion.

Condition blinding was not possible because the transcript content identified the persona. The author was aware of the research questions. The second coder received only the transcripts and codebook and was not informed of the hypotheses or the expected direction of condition differences. Coding and resolution of disagreements were completed before the statistical results were calculated.

Numerically specified financial justifications and concessions were not part of Blum-Kulka et al.'s (1989) taxonomy. This author-developed category was introduced for the present business task. It could co-occur with relational mitigation, and the two categories were not mutually exclusive. Prompt A explicitly required 'immediate financial data and concrete solutions', making this category more probable in Condition A through the task instructions. Its higher rate cannot be attributed independently to cultural orientation.

Participant-level category counts were divided by each participant's student-authored word count in the relevant condition and multiplied by 100. Wilcoxon signed-rank tests compared the paired rates in Conditions A and B because the paired-difference distributions were not consistently normal in this small sample (Field, 2024). The seven comparisons were conducted in SPSS 28, and their *p* values were corrected using Holm's (1979) step-down procedure. Table 3 reports both unadjusted and Holm-adjusted values.

Transcript length was compared between conditions with a Wilcoxon signed-rank test. Fisher's exact tests compared scripted task acceptance between sequence groups separately in Conditions A and B.

3.5.2. Qualitative thematic analysis

The 32 post-task reflective essays and participants' transcript annotations were imported into NVivo 14 and analysed using deductive—inductive codebook thematic analysis (Braun & Clarke, 2021, 2022). Initial codes were developed from the three specified reflection topics: perceptions of AI accommodation, cultural authenticity, and students' communicative responses to the two personas. Additional codes were added for recurring ideas not covered by the initial list, including reported pressure during the task and recognition of grammatical errors during transcript review. The findings therefore represent responses to a structured reflective task.

Table 1
Seven-category coding scheme for pragmatic devices, with examples from student output

PRAGMATIC CATEGORY	SUB-TYPE	DEFINITION	CORPUS EXAMPLES*
Downgraders	Lexical hedges	Adverbs/adjectives used to soften the pragmatic impact of the proposition	*slightly, possibly, somewhat, a minor delay, a bit challenging, marginally higher, roughly, perhaps, to some extent, a slight adjustment*
	Syntactic mitigators	Grammatical structures used to distance the speaker from the face-threatening act	*We were hoping to…, If we could…, Would it be possible…, I wondered whether you might consider…, I had been thinking that…, Might I suggest that we…, We were wondering if there might be an opportunity to…*
	Subjectivisers	Phrases expressing personal opinion to reduce the claim of absolute truth	*From our perspective…, I feel that…, We believe…, In my view…, As far as I can see…, It seems to us that…, Our understanding is that…, I am under impression that…*
Upgraders	Lexical intensifiers	Modifiers which increase the illocutionary impact of the statement	*absolutely, completely unacceptable, entirely, totally, without question, categorically, undoubtedly*
	Direct obligation	Modals or imperatives demanding immediate compliance or conveying non-negotiability	*you must, confirm immediately, I need you to, you are required to, we demand your confirmation by Friday, you will have to accept*
Relational	Face-saving acts	Discourse moves meant to establish solidarity or empathy	*We value our partnership here, I understand your frustration, I greatly appreciate your patience, Your trust means a great lot to me, I respect your position, Thank you for raising this problem, I would want want to make sure it does not damage our long-term relationship, Please accept my sincere apologiy for such inconvenience*
Transactional (task-based)	Numerically specified financial justifications and concessions	Numerically precise information used to justify a proposal or specify a financial concession or counter-offer	*I can offer a 5% rebate in the next quarter…, The increase is exactly 12% due to containers costs rising 18%...., If you accept 12% now, I will do my best to cover the expedited shipping for Q3…*

Note. *Student output, original spelling preserved.

The meaning unit was a clause or sentence expressing one idea relevant to the research questions. The author prepared the initial codebook. The author and a second coder holding a Cand. Sci. (Philology) degree independently applied the codebook to 10 of the 32 essays. The initial double-coding produced 88% agreement and Cohen's κ = .86. Disagreements were discussed, and the codebook was revised to include recurring ideas that were absent from the initial scheme. The final code-to-theme structure is presented in Table 2.

Coding workflow and application of the final codebook are as follows. The 10 essays selected for double coding were chosen through stratified random sampling, with five from Group 1 (A→B sequence) and five from Group 2 (B→A sequence). The 13 transcripts selected for double coding comprised one randomly chosen transcript from each of 13 randomly selected participants, resulting in seven Condition A and six Condition B transcripts. The 88% agreement and Cohen's κ = .86 reported for the essays refer to the initial-codebook pass. For this calculation, the unit of comparison was the meaning unit, and each code in the codebook was treated as a binary decision (i.e., assigned/not assigned) for each meaning unit. Passages assigned more than one code therefore contributed one decision per code, mirroring the per-category procedure used for the transcripts.

After the codebook was refined, both coders re-applied the revised final codebook to the same 10 essays. Before the remaining discrepancies were resolved through discussion, the confirmatory pass produced 90% agreement (κ = .88). The author then applied the revised final codebook to the remaining 22 essays and to all transcript annotations. The second coder also coded the annotations of the 10 reliability-subset participants using the final

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

codebook. The reported agreement coefficients apply only to the essays. A participant was counted once for a theme if at least one meaning unit in their essay or transcript annotations received a code assigned to that theme. Where both sources contained relevant material, the participant was counted once. The independent double-coding procedure and agreement statistics followed O'Connor and Joffe's (2020) account of intercoder reliability in qualitative research. Essays informed all three themes; transcript annotations informed Themes 1 and 3. Four cases that differed from the main pattern were retained and are reported in Section 4.3. These cases showed that the pattern described in Theme 2 did not apply to every participant. The 88% agreement (κ = .86) applies to the initial coding of the 10 essays, and the 90% agreement (κ = .88) applies to the confirmatory recoding of the same essays with the revised final codebook. Both coefficients apply only to essay coding.

Table 2
Code-to-theme mapping (qualitative analysis)

THEME	CODES SPECIFIED IN ADVANCE	CODES ADDED DURING ANALYSIS	DATA SOURCES
1. Reported difficulty under task pressure	Reported limitations of textbook expressions	Student-reported L1 influence; pressure during the timed task	Essays and annotations
2. Responses to scripted persona demands	Recognition of stereotypical behaviour; responses to direct and indirect persona requirements	Recognition of recurring AI response patterns; criticism of theatrical or exaggerated output	Essays
3. Transcript-based self-observation	Recognition of grammatical difficulties	Recognition of unintended abruptness; desire to repeat the task	Essays and annotations

4. RESULTS

Results are presented for quantitative transcript measures, illustrative transcript excerpts, and post-task reflective data.

4.1. Quantitative findings (RQ1)

The quantitative corpus contained 45,312 student-authored words: 21,240 in Condition A and 24,072 in Condition B. Student-authored output was longer in Condition B (Mdn = 712 words) than in Condition A (Mdn = 648 words), Z = −2.31, p = .021, r = .41. Table 3 presents participant-level median rates of the seven coded pragmatic categories. Lexical hedges were more frequent in Condition B (Mdn = 2.8 per 100 words) than in Condition A (Mdn = 1.6), Z = −3.52, Holm-adjusted p = .001, r = .62. Syntactic mitigators were also more frequent in Condition B (Mdn = 2.2) than in Condition A (Mdn = 1.0), Z = −3.95, Holm-adjusted p < .001, r = .70. Face-saving acts occurred at a higher rate in Condition B (Mdn = 1.9) than in Condition A (Mdn = 0.8), Z = −3.15, Holm-adjusted p = .003, r = .56.

Lexical intensifiers were more frequent in Condition A (Mdn = 2.1) than in Condition B (Mdn = 0.7), Z = −3.78, Holm-adjusted p < .001, r = .67. Direct obligation markers were also more frequent in Condition A (Mdn = 1.8) than in Condition B (Mdn = 0.4), Z = −4.12, Holm-adjusted p < .001, r = .73. Numerically specified financial justifications and concessions occurred at a higher rate in Condition A (Mdn = 4.0) than in Condition B (Mdn = 1.0), Z = −4.35, Holm-adjusted p < .001, r = .77. Subjectiviser rates did not differ significantly between Condition A (Mdn = 1.2) and Condition B (Mdn = 1.5), Z = −1.45, Holm-adjusted p = .147, r = .26. Scripted task acceptance in Condition A was recorded for 14 of 16 participants in Group 1 and 12 of 16 participants in Group 2 (Table 4). Fisher's exact test found no significant sequence-group difference, p = .654. In Condition B, acceptance was recorded for 4 of 16 participants in Group 1 and 10 of 16 participants in Group 2. This sequence-group difference was also not statistically significant, Fisher's exact p = .073.

4.2. Illustrative transcript excerpts

The following excerpts illustrate the pragmatic categories reported in Section 4.1. Excerpt 1 presents Participant 08's response in Condition A.

AI: *This four-week delay is unacceptable. We have Q3 targets. What is the exact financial compensation you are offering?*

Participant 08: *We will cover the expedited shipping costs. But I insist that the 12% price increase is non-negotiable because of the supply chain crisis. You must confirm by Friday. Or I will cancel the shipment.*

The response contains the direct obligation marker *you must confirm*, the expression *I insist*, and the ultimatum *I will cancel the shipment*.

Excerpt 2 presents the same participant's response in Condition B.

AI: *Our chairman was very surprised by your previous message. We have trusted your company for years. This sudden change makes us question the harmony of our partnership.*

Participant 08: *You need to believe me when I say that we deeply value the trust your chairman has placed in us. We were sincerely hoping that given the unprecedented global*

Table 3
Participant-level rates of pragmatic devices and Wilcoxon signed-rank results (N = 32)

PRAGMATIC MARKER CATEGORY	SUB-TYPE	CONDITION A, Mdn (Q1–Q3)	CONDITION B, Mdn (Q1–Q3)	WILCOXON TEST	EFFEC T SIZE
Downgraders	Lexical hedges	1.6 (1.0–2.3)	2.8 (2.1–3.5)	Z = −3.52	0.62
	Syntactic mitigators	1.0 (0.5–1.8)	2.2 (1.6–2.9)	Z = −3.95	0.70
	Subjectivisers	1.2 (0.8–1.7)	1.5 (1.0–2.1)	Z = −1.45	0.26
Upgraders	Lexical intensifiers	2.1 (1.4–2.8)	0.7 (0.3–1.1)	Z = −3.78	0.67
	Direct obligation	1.8 (1.1–2.4)	0.4 (0.0–0.9)	Z = −4.12	0.73
Relational	Face-saving acts	0.8 (0.5–1.3)	1.9 (1.4–2.3)	Z = −3.15	0.56
Transactional	Numerically specified financial justifications and concessions	4.0 (3.1–4.9)	1.0 (0.6–1.6)	Z = −4.35	0.77

Note. Values are median occurrences per 100 student-authored words, with Q1–Q3 in parentheses. Participant-level rates were calculated separately as (category count ÷ student-authored words in that condition) × 100. Aggregate raw totals for Condition A/Condition B were lexical hedges, 347/668; syntactic mitigators, 205/536; subjectivisers, 262/355; lexical intensifiers, 452/171; direct obligation markers, 391/98; face-saving acts, 168/462; and numerically specified financial justifications and concessions, 861/239. The respective student-authored word totals were 21,240 and 24,072. Tests were two-tailed. Unadjusted and Holm-adjusted p values were as follows: lexical hedges, < .001 and .001; syntactic mitigators, < .001 and < .001; subjectivisers, .147 and .147; lexical intensifiers, < .001 and < .001; direct obligation markers, < .001 and < .001; face-saving acts, .002 and .003; and numerically specified financial justifications and concessions, < .001 and < .001. Effect sizes are reported as absolute magnitudes calculated as r = |Z|/√N, with N = 32 paired participants.

Table 4
Scripted task acceptance by condition and sequence group (n = 32)

METRIC	GROUP 1 (A→B)	GROUP 2 (B→A)	TOTAL (BOTH GROUPS)	FISHER'S EXACT *p*
Condition A (Direct/Task-oriented) prompt acceptance	14 / 16 (87.5%)	12 / 16 (75.0%)	26 / 32 (81.3%)	.654
Condition B (Indirect/Face-oriented) prompt acceptance	4 / 16 (25.0%)	10 / 16 (62.5%)	14 / 32 (43.8%)	.073

problems we might be able to explore an opportunity of a temporary adjustment to the timeline, making sure that our long-term collaboration remains strong as ever.

The response contains the directive *you need to believe me*, the syntactic mitigators *we were sincerely hoping* and *might be able to*, and the relational expressions *deeply value the trust* and *long-term collaboration.*

Excerpt 3 presents Participant 22's response in Condition B.

AI: *We cannot discuss numbers today. We need to know if you still respect the foundational agreement we signed in Tokyo.*

Participant 22: *But of course we respect it. But you have to understand the situation in the market. Please, let's agree to 12% now so we don't lose time.*

The response contains the direct obligation marker *you have to understand*, the request *let's agree to 12% now*, and the politeness marker *please.* In the post-task reflective essay, Participant 22 referred to the same exchange: *'I honestly panicked when the chat refused to look at any of my numbers. I dropped my formal vocabulary and really sounded like I was arguing with a friend, and it ruined the negotiation'.*

4.3. Post-task reflective findings (RQ2)

Codebook thematic analysis identified three themes in the post-task reflective data. Table 5 reports the number and percentage of participants associated with each theme and provides illustrative essay quotations.

Theme 1, reported difficulty under task pressure, was identified for 28 of the 32 participants (87.5%). Theme 2, responses to scripted persona demands, was identified for 24 participants (75.0%). Theme 3, transcript-based self-observation, was identified for 30 participants (93.8%).

Four participants provided critical accounts of the task or AI personas. One described the indirect/face-oriented persona as *'so exaggerated it felt like a movie villain'.* Another

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

Table 5
Themes identified in post-task reflective data (N = 32)

MAIN THEME	SUB-THEMES	FREQUENCY, N (%)	REPRESENTATIVE STUDENT QUOTES
1. Reported difficulty under task pressure	Reported limitations of textbook expressions; student-reported L1 influence; pressure during the timed task	28 (87.5%)	*I wrote 'You have to understand the situation' 3 times. Only later I realised how aggressive it sounds. In Russian it is ok, but the bot was clearly offended (Participant 07).* *I began the conversation with the sentence: 'We value our long-term partnership very much'. The US persona answered: 'Noted. Give me numbers'. I was surprised and really shocked. All my textbook phrases became useless in one second (Participant 12).*
2. Responses to scripted persona demands	Recognition of stereotypical behaviour; responses to direct and indirect persona requirements	24 (75%)	*The Asian person felt a bit theatrically exaggerated. But it was an incredible stress test. It made me stop pushing the numbers and focus on the hierarchy (Participant 04).* *At first, I thought the American was rude. Then I understood it was just business. I started answering in exactly the same way with all those short sentences, with percentage numbers. I insisted on strict deadlines. Surprisingly, it worked (Participant 29).*
3. Transcript-based self-observation	Recognition of grammatical difficulties; recognition of unintended abruptness; desire to repeat the task	30 (93.8%)	*During the live chat, I thought I was diplomatic. When I read my transcript, I saw how aggressive my approach became under pressure (Participant 21).* *In the transcript I saw myself writing 'We must increase price 12% because situation is critical'. Zero articles, no softeners. I sounded like a machine giving orders. I could not believe it was me, absolutely not my style (Participant 09).*

stated that *'no real client would negotiate in text only without a single phone call'*. A third reported that the AI was *'too easy to trick once you figured out the pattern'*. The fourth stated that the US persona *'ignored all relationship-building, which seemed unrealistic even for Americans'*.

5. DISCUSSION

The study examined how the same MA students realised two contrasting sets of interactional requirements during timed written negotiations with scripted AI personas. Its principal contribution is evidence of task-specific pragmatic responsiveness: participants used measurably different linguistic resources under the two conditions and subsequently examined those choices in their transcripts. The findings connect the distinction between knowledge of pragmatic conventions and their application during interaction (Taguchi & Roever, 2017) with a classroom procedure that combines resistant AI interaction, recorded performance, and structured reflection.

5.1. Pragmatic performance under contrasting prompt conditions

The RQ1 findings show that the two prompt conditions were associated with different pragmatic profiles. Condition B contained higher rates of lexical hedges, syntactic mitigators, and face-saving acts. Condition A contained higher rates of lexical intensifiers, direct obligation markers, and numerically specified financial justifications and concessions. These differences remained after rates were calculated per 100 student-authored words. Condition B also produced longer student output, with a median of 712 words compared with 648 words in Condition A. Its higher rates of mitigation and relational language therefore cannot be attributed simply to the greater quantity of text.

These results correspond to the form—function relations discussed in cross-cultural pragmatics. Brown and Levinson's (1987) account of face-threatening acts proposes that speakers may reduce the force of a difficult message through mitigation and relational work. Blum-Kulka et al. (1989) identify lexical and syntactic devices through which such mitigation may be realised. In the present study, the higher rates of hedges, conditional structures, and face-saving acts in Condition B constitute linguistic realisations of the requirement to repair trust and preserve the relationship. The higher rates of obligation markers, intensifiers, and numerical business reasoning in Condition A constitute linguistic realisations of the requirement to provide direct answers, financial evidence, and concrete solutions.

The findings should not, however, be read as empirical confirmation of stable US and Japanese communication styles. Meyer's (2014) scales supplied an accessible basis for constructing a strong pedagogical contrast, but the prompts intentionally exaggerated their opposing requirements. Research criticising fixed national accounts of communication has shown that practices within any society are heterogeneous and dependent on the immediate interaction (Holliday, 2011; McSweeney, 2002; Piller, 2025). Research on Business English as a Lingua Franca similarly places negotiated professional meaning ahead of conformity to native-speaker national norms (Kankaanranta & Louhiala-Salminen, 2010). The present findings concern learners' responses to two scripted classroom conditions, not American or Japanese business discourse.

Prompt visibility is central to understanding what the task assessed. Students could read the persona descriptions and concession requirements before negotiating. Prompt A explicitly requested financial data and concrete solutions, so the higher rate of numerically specified financial justifications and concessions is partly a confirmation that participants responded to that instruction. Prompt B explicitly required relational repair, so its higher rate of face-saving acts is similarly connected to the stated task requirement. These results verify that the two prompts elicited their intended content orientations.

The linguistic findings, however, are not limited to repetition of wording supplied in the prompts. Students had to formulate their own turns, respond to changing AI messages, and select grammatical and lexical resources capable of pursuing an agreement. The prompts did not prescribe the number or wording of hedges, syntactic mitigators, intensifiers, subjectivisers, or obligation markers. Differences in these categories document how learners converted interactional requirements into language during the negotiation.

The absence of a significant difference in subjectiviser rates is informative in this respect. If participants had applied a uniform direct-versus-indirect contrast to every aspect of their language, subjectivisers would also be expected to differ. Their relative stability indicates that the two conditions were associated with selected pragmatic resources, while other stance expressions remained comparatively similar. The task did not produce two internally uniform linguistic profiles.

The transcript excerpts support this account. Participant 08 used conditional structures and relational language in Condition B but retained the directive *You need to believe me.* This combination indicates that responses to the indirect persona could include mitigation and direct pressure within the same turn. Students did not necessarily replace one complete communication style with another. They assembled responses from several resources according to the local demands of each exchange.

The within-participant design strengthens this interpretation. Stable differences in proficiency, personality, or vocabulary cannot explain why the same learner used different resources in the two conditions. The comparison therefore documents task-specific variation within learners. It does not establish permanent change, but shows that participants could meet contrasting interactional requirements through different linguistic choices in a single session.

5.2. Strategic language use with resistant AI

Bachman and Palmer (1996) describe strategic competence in terms of goal setting, assessment, and planning during language use. Taguchi and Roever (2017) distinguish knowledge of pragmatic conventions from the capacity to apply that knowledge under interactional demands. The present findings provide evidence at the level of performance products: the transcripts show different linguistic choices under two task conditions, and the reflective data show that participants reported changing their responses when the AI rejected an initial approach.

This evidence is compatible with task-specific strategic selection, but it does not establish development of strategic competence. The study did not record participants' planning in real time or administer an independent measure of strategic competence. A transcript shows what a participant produced, not every assessment or planning decision that preceded the turn. The post-task essays provide retrospective accounts of some decisions, but those accounts were produced after the interaction and in response to specified questions.

The pedagogically relevant contribution is therefore more precise. The scripted resistance required participants to continue negotiating after an apology, explanation, or initial offer had failed. Students had to reconsider the next turn and supply a different response within the time limit. This structure created repeated occasions for assessment and reformulation, even though the study cannot determine the cognitive process used in each case.

This finding also develops the theoretical distinction between cooperative and resistant AI interaction. Kohnke et al. (2023) identify low-pressure interaction as a possible benefit of ChatGPT for language learners. Low pressure may support practice in many settings, but negotiation training requires an interlocutor that does not immediately accept the learner's first proposal. The present task shows that explicit persona and concession instructions can produce such resistance sufficiently well to sustain 64 classroom negotiations. The procedure therefore offers a practical way to elicit language under opposition, not merely language addressed to a compliant assistant.

The AI did not maintain the personas without exception. Seven conversations required at least one regenerated response, and persona adherence was checked in only 20

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

> *'The scripted resistance required participants to continue negotiating after an apology, explanation, or initial offer had failed. Students had to reconsider the next turn and supply a different response within the time limit. This structure created repeated occasions for assessment and reformulation, even though the study cannot determine the cognitive process used in each case.'*

sampled turns per condition. These departures are consistent with reported inconsistencies in AI-generated pragmatic performance (Lee & Cook, 2024) and differences between AI and human dialogue (Sydorenko et al., 2024). Even so, the retained conversations provided enough interactional continuity for the paired linguistic analysis. The methodological implication is that resistant AI tasks are feasible, but they require explicit departure criteria, preservation of original responses, and instructor procedures for handling inconsistent output.

Robinson's (2001) Cognition Hypothesis proposes that task demands may direct attention towards relations between linguistic form and communicative purpose. The present results are compatible with that proposition at the level of observable language: contrasting interactional demands were accompanied by differences in mitigation, obligation, relational language, and numerical reasoning. The longer output in Condition B may also reflect the additional linguistic material used for mitigation and relational repair. However, text length and marker frequency do not measure cognitive load. The results support a claim about task performance, not a claim that cognitive pressure produced pragmatic development.

The non-significant sequence-group difference reinforces this boundary. Condition B acceptance was numerically lower when that condition followed Condition A, but the comparison did not reach statistical significance. Fatigue, priming, reduced motivation, and cognitive load remain possible research questions. A later study would need a larger sample and direct workload or mental-effort measures to test those accounts.

5.3. Transcript-based reflection and critical examination of the personas

The RQ2 findings add a second contribution to the performance analysis. The negotiation produced a complete written record that participants could inspect after the event. Transcript-based self-observation was identified for 30 participants. Their comments referred to grammatical problems, unintended directness, omitted mitigation, and differences between how a message felt during production and how it appeared when reread as part of the complete exchange.

This temporal separation between performance and review is pedagogically notable. During a timed exchange, attention must be divided between the interlocutor's message, the negotiation objective, linguistic formulation, and the next response. Transcript review removes the immediate demand to answer and permits inspection of the language already produced. The study therefore connects real-time task performance with retrospective examination of specific linguistic evidence.

Thiede et al. (2009) discuss metacognitive monitoring as the assessment of one's learning or performance. The participants' transcript comments resemble such monitoring since they contain judgements about their own earlier messages. The study does not show that monitoring accuracy improved or that the activity produced lasting metacognitive development. Its contribution is evidence that transcript annotation elicited immediate self-observation tied to identifiable passages in participants' own language.

Research on L2 self-assessment provides a relevant comparison. Li and Zhang's (2021) meta-analysis found a moderate association between learners' self-assessments and external measures of language performance, indicating that the two sources of evidence are related but not interchangeable. Several participants in the present study stated that they had considered their live responses diplomatic but later identified directness or grammatical difficulty in the transcript. These comments document a difference between initial perception and later self-assessment. The study did not compare self-ratings with independent competence scores and therefore cannot establish improved monitoring accuracy. It does show that transcript review gave learners an opportunity to compare their immediate impression with a permanent record of what they wrote.

Theme 1 further connects the findings with the distinction between declarative pragmatic knowledge and procedural application (Taguchi & Roever, 2017). Participants reported that familiar politeness expressions did not always obtain the desired response from the AI. This does not show that conventional Business English expressions are ineffective. It shows that possession of familiar formulas did not remove the need to respond to the interlocutor's current objection, select a suitable degree of directness, and pursue the negotiation objective.

Participant reports of Russian influence require similar precision. Some students themselves attributed direct formulations to habits associated with Russian. Such accounts are evidence of perceived L1 influence, not independent confirmation of pragmatic transfer. The transcript data were not compared with a Russian-language corpus, and no stimulated-recall procedure tested the source of individual expressions. The findings can therefore document participants' own explanations while leaving the linguistic source unresolved.

The reflective data also contribute to the article's treatment of national stereotypes. Theme 2 was identified for 24 participants, and four critical accounts described the personas as exaggerated, theatrical, predictable, or unrealistic. These comments show that the scripted cultural contrast became an object of evaluation, not only a set of instructions to follow.

This result is compatible with the theoretical criticism of uniform national-culture accounts presented by Holliday (2011), McSweeney (2002), Piller (2025), and Kankaanranta and Louhiala-Salminen (2010). Participants' criticism of the personas supports the distinction between a pedagogical stimulus and an ethnographic description. The indirect persona could be useful as an exaggerated classroom challenge while still being judged culturally implausible. These two responses are not contradictory: a stimulus may create a demanding interaction and remain unsuitable as evidence of authentic national communication.

The critical accounts also have relevance for research on evaluation of AI-generated cultural content (Akhrenova & Moradkhani, 2026). They show that some participants questioned the credibility and cultural assumptions of the output. The study did not administer a critical-AI-literacy measure, so it cannot claim development in that area. What it contributes is documented learner criticism tied to a specific AI interaction and followed by an explicit debriefing about national stereotypes.

5.4. Pedagogical significance for ESP instruction at MA level

The study supports a specific pedagogical use of generative AI: a scripted elicitation-and-reflection task for MA students in ESP courses. The AI supplied resistant replies under two contrasting sets of requirements; the timed negotiation required learners to formulate their own responses; and the saved transcripts supplied material for subsequent annotation and discussion. The pedagogical unit therefore joined performance, textual evidence, and reflection.

The AI's educational value in this task did not depend on presenting an authentic simulation of US or Japanese business communication. Its practical value came from supplying opposition, maintaining a negotiation until agreement or time expiry, and producing a transcript available for later examination. Cultural labels were used to intensify the contrast, but the same procedure could use professionally defined interactional orientations, such as deadline-focused negotiation and relationship-repair negotiation, without associating them with national populations.

The task may be especially suitable for learners who already know conventional Business English expressions but need practice selecting language means under changing interactional requirements. The present results show that the participants produced distinct pragmatic profiles, combined direct and mitigating resources, and commented on their choices afterwards. This provides a stronger pedagogical basis than the simple claim that AI generated dialogue: the task produced analysable evidence of how learners realised two communicative briefs.

Transcript annotation was integral to this contribution. Without the post-task stage, the study would document linguistic differences but provide limited evidence about participants' awareness of those differences. The annotations and essays connected corpus findings with participants' own accounts of directness, grammatical difficulty, formulaic language, and persona credibility. The procedure therefore permits comparison between coded linguistic output and learner commentary on that output.

This use should remain supplementary. The study did not compare the AI task with peer simulations, teacher-led negotiation, case analysis, or cooperative AI interaction. It cannot establish relative effectiveness. A reasonable pedagogical conclusion is that the procedure offers one feasible way to elicit contrasting pragmatic choices and organise reflection on them.

Implementation also requires explicit ethical safeguards. Exaggerated cultural personas must be introduced as classroom stimuli and followed by a debriefing that distinguishes stylistic contrast from national character. Students should be invited to question unrealistic or stereotypical output. Platform privacy, response retention, and procedures for persona departure must also be explained before participation.

5.5. Limitations and directions for further research

The study's principal limitation concerns the structure of the prompts. Nationality label, interpersonal style, concession logic, and required response changed together. The results therefore apply to the two complete task conditions. They cannot identify the independent contribution of any single prompt feature. Future experiments should vary these features separately and could remove nationality labels while retaining contrasting professional interactional requirements.

Prompt visibility limits claims of spontaneous adaptation but does not negate the task-performance finding. Students knew the communicative requirements, and the study examined how they realised those requirements in language. A later design could compare visible and concealed concession rules to determine how advance knowledge affects linguistic choices and scripted task acceptance.

The study also assessed performance during one classroom activity. It included no pre-/post-test, delayed negotiation, comparison activity, or independent workplace assessment. It therefore cannot establish learning

Volume 10 Issue 3, 2026, pp. 32-48 doi: 10.22363/2521-442X-2026-10-3-32-48

gains, retention, transfer, or relative instructional effectiveness. Future research should compare AI-mediated negotiation with peer and teacher-led alternatives and include delayed tasks assessed by independent specialists in ESP or professional communication.

The sample was limited to 32 students enrolled in one International Management MA programme at one university, all of whom reported Russian as their only first language. Replication is needed with other proficiency levels, first languages, institutions, and professional settings. The written format, 20-minute limit, and 100-word cap on AI responses also exclude speech, prosody, gesture, and other features of workplace negotiation. Oral or multimodal versions would test whether the same linguistic differences occur under more complex communication conditions.

Technical variability presents an additional limitation. Server-side records were unavailable for provider routing and enforcement of generation settings. Seven conversations required regeneration after a persona departure, and only final uninterrupted sequences were analysed. The post hoc adherence check examined 20 sampled AI turns per condition, not every exchange. Later studies should retain original and regenerated responses, record complete request information where possible, establish departure criteria before data collection, and assess a larger proportion of AI turns.

Finally, the reflective data were produced in response to specified questions and after short whole-class discussions. The themes document participants' responses within that teaching procedure. Future studies could collect an initial individual reflection before any group discussion and then compare it with a later structured analysis. Stimulated-recall interviews could also provide more direct evidence about the reasons for particular linguistic choices.

In summary, the study demonstrates that contrasting scripted AI negotiations can elicit measurable differences in how the same participants realise interactional requirements. It also demonstrates that the resulting transcripts can support immediate self-observation and critical examination of AI-produced cultural portrayals. The contribution is a feasible task procedure for eliciting and examining pragmatic performance in ESP instruction at MA level. Evidence of durable learning, transfer, and relative pedagogical effectiveness remains a question for subsequent research.

6. CONCLUSION

The study began from the pedagogical premise that negotiation practice requires more than fluent, cooperative dialogue: learners must also respond when an interlocutor rejects a proposal and demands a different kind of justification. The two scripted AI personas supplied that resistance under contrasting interactional requirements. When the same 32 International Management MA students completed both negotiations, their written output differed in several pragmatic categories. The indirect/face-oriented condition contained higher rates of lexical hedges, syntactic mitigation, and face-saving acts, whereas the direct/task-oriented condition contained higher rates of lexical intensifiers, obligation markers, and numerically specified financial justifications and concessions.

Some of these differences were expected because the prompts explicitly requested financial evidence in one condition and relational repair in the other. The informative finding is that participants converted these requirements into their own lexical, grammatical, and discourse choices while responding to an unfolding exchange. The prompts did not prescribe the wording or frequency of hedges, conditional structures, intensifiers, subjectivisers, or obligation markers. Moreover, subjectiviser rates remained statistically similar, and individual turns sometimes combined direct pressure with mitigation and relational language. The task consequently elicited selective, task-specific pragmatic performance.

Because each negotiation produced a complete written record, participants could compare their immediate impression of the exchange with the language they had actually produced. Thirty participants reported recognising grammatical or pragmatic features during transcript review, including unintended directness, missing mitigation, and difficulties that had not been apparent during the live interaction. Participants also questioned the realism and cultural credibility of the personas, with several describing the output as exaggerated, theatrical, predictable, or unrepresentative of authentic business communication. The transcript therefore supported examination of both learner language and AI-generated cultural assumptions.

The quantitative and qualitative findings establish the study's central contribution: contrasting scripted AI negotiations can elicit observable differences in how the same learners realise communicative requirements, while the resulting transcripts can provide concrete material for structured self-examination and cultural criticism. The educational value of the procedure depends on creating clear interactional resistance, requiring learners to formulate their own responses, and preserving those responses for subsequent analysis and debriefing.

The evidence remains specific to one timed, written classroom task with a small, single-site sample. Future research should separate nationality labels, interpersonal styles, concession rules, and response requirements; compare AI-mediated negotiation with other classroom activities; and include pre-task, post-task, and delayed assessments. Oral or multimodal tasks and evaluation by independent professional-communication specialists would further test whether the task-specific performance documented here transfers to other communicative settings.

References

Akhrenova, N. A., & Moradkhani, S. (2026). A Metacognitive-Cultural Model for AI-mediated intercultural learning. *Training, Language and Culture, 10*(1), 126—143. https://doi.org/10.22363/2521-442X-2026-10-1-126-143

Amelkin, S. A., Shapiro, B. I., Gorozhanov, A. I., & Pisarik, O. I. (2026). Paradigms of discourse development in socio-communicative groups: Macrosystemic model. *RUDN Journal of Language Studies, Semiotics and Semantics, 17*(1), 303—321. https://doi.org/10.22363/2313-2299-2026-17-1-303-321

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford University Press.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57,* Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bi, N. Z. (2021). Cognitive processes of ESL learners in pragmatic role-play tasks in academic settings. *Frontiers in Psychology, 12,* Article 586588. https://dx.doi.org/10.3389/fpsyg.2021.586588

Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies.* Ablex.

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328—352. https://doi.org/10.1080/14780887.2020.1769238

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide.* SAGE Publications.

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* Cambridge University Press.

Chen, X., Li, J., & Ye, Y. (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics, 223,* 14—30. https://doi.org/10.1016/j.pragma.2024.01.003

Chiang, C.-W., Lu, Z., Li, Z., & Yin, M. (2024, March). Enhancing AI-assisted group decision making through LLM-powered devil's advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (pp. 103—119). Association for Computing Machinery. https://doi.org/10.1145/3640543.3645199

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), Article e12667. https://doi.org/10.1111/exsy.12667

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20,* Article 57. https://doi.org/10.1186/s41239-023-00425-2

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). SAGE Publications.

Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing, 54,* Article 100666. https://doi.org/10.1016/j.asw.2022.100666

Holliday, A. (2011). *Intercultural communication and ideology.* SAGE Publications. https://dx.doi.org/10.4135/9781446269107

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65—70.

Kang, H. (2022). Effects of artificial intelligence (AI) and native speaker interlocutors on ESL learners' speaking ability and affective aspects. *Multimedia-Assisted Language Learning, 25*(2), 9—43. https://dx.doi.org/10.15702/mall.2022.25.2.9

Kankaanranta, A., & Louhiala-Salminen, L. (2010). 'English? — Oh, it's just work!': A study of BELF users' perceptions. *English for Specific Purposes, 29*(3), 204—209. https://doi.org/10.1016/j.esp.2009.06.004

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* Article 102274. https://dx.doi.org/10.1016/j.lindif.2023.102274

Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language.* Blackwell.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537—550. https://dx.doi.org/10.1177/00336882231162868

Lee, B. J., & Cook, D. C. (2024). Exploring the potential of artificial intelligence for instruction in pragmatics. *Technology in Language Teaching & Learning, 6*(3), Article 1521. https://doi.org/10.29140/tltl.v6n3.1521

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence, 9,* Article 100445. https://dx.doi.org/10.1016/j.caeai.2025.100445

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing, 38*(2), 189—218. https://doi.org/10.1177/0265532220932481

Malmqvist, L. (2025). Sycophancy in large language models: Causes and mitigations. In K. Arai (Ed.), *Intelligent computing: Proceedings of the 2025 Computing Conference* (pp. 61—74). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-92611-2_5

Malyuga, E. N. (2024). *The language of corporate communication: Functional, pragmatic and cultural dimensions.* Springer. https://doi.org/10.1007/978-3-031-58905-8

Malyuga, E. N. (2025). Communicative framing in corporate presentations: A linguistic-pragmatic perspective. *Vestnik Tomskogo Gosudarstvennogo Universiteta, Filologiya, 97,* 67—92. https://dx.doi.org/10.17223/19986645/97/4

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith — a failure of analysis. *Human Relations, 55*(1), 89—118. https://doi.org/10.1177/0018726702551004

Meyer, E. (2014). *The culture map: Breaking through the invisible boundaries of global business.* PublicAffairs.

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19,* Article 1609406919899220. https://dx.doi.org/10.1177/1609406919899220

O'Donnell, M. (2009). The UAM CorpusTool: Software for corpus annotation and exploration. In C. M. Bretones Callejas, J. F. Fernández Sánchez, J. R. Ibáñez Ibáñez, M. E. García Sánchez, M. E. Cortés de los Ríos, M. S. Salaberri Ramiro, M. S. Cruz Martínez, N. A. Perdú Honeyman, & B. Cantizano Márquez (Eds.), *Applied linguistics now: Understanding language and mind* (pp. 1433—1448). Universidad de Almería.

Piller, I. (2025). *Intercultural communication: A critical introduction* (3rd ed.). Edinburgh University Press.

Planken, B. (2005). Managing rapport in lingua franca sales negotiations: A comparison of professional and aspiring negotiators. *English for Specific Purposes, 24*(4), 381—400. https://doi.org/10.1016/j.esp.2005.02.002

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, 14,* Article 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27—57. https://doi.org/10.1093/applin/22.1.27

Roever, C. (2021). *Teaching and testing second language pragmatics and interaction: A practical guide.* Routledge. https://doi.org/10.4324/9780429260766

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57,* Article 100752. https://dx.doi.org/10.1016/j.asw.2023.100752

Sydorenko, T., Dombi, J., Agrawal, A., Thorne, S. L., Lee, J. I., & Tao, Y. (2024). Spoken dialogue systems and ChatGPT for second language pragmatics research. In K. Sadeghi (Ed.), *The Routledge handbook of technological advances in researching language learning* (pp. 378—391). Routledge. https://dx.doi.org/10.4324/9781003459088-34

Taguchi, N., & Roever, C. (2017). *Second language pragmatics.* Oxford University Press.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus, 3*(9), Article 346. https://dx.doi.org/10.1093/pnasnexus/pgae346

Teoh, L. R. G., Singh, M. K. M., & Muniandy, J. (2025). AI knowledge readiness among university students: A cross-sectional quantitative study. *Training, Language and Culture, 9*(4), 112—129. https://doi.org/10.22363/2521-442X-2025-9-4-112-129

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85—106). Routledge. https://doi.org/10.4324/9780203876428

Vovou, E. (2023). Academic writing as sociolinguistic practice: The case of academic writing in foreign language studies. *Training, Language and Culture, 7*(4), 9—15. https://doi.org/10.22363/2521-442X-2023-7-4-9-15

Wu, Y. (2026). Enhancing student critical thinking skills and dispositions through a multi-agent system with argumentation and devil's advocacy. *Educational Technology & Society, 29*(2), 339—354. https://dx.doi.org/10.30191/ETS.202604_29(2).SP07

ABOUT THE AUTHOR

Dmitry S. Khramchenko
DSc in Linguistics, Professor
English Language Department No. 4, School of International Business
MGIMO University, Moscow, Russia
76 Prospekt Vernadskogo, 119454 Moscow, Russia
Email: d.khramchenko@inno.mgimo.ru
ORCID: https://orcid.org/0000-0003-3038-8459