ATU
PRESS

Document Type: Research Paper

Evaluating EAP Students' Pragmatic Competence Using Authentic Assessment Tasks

Mohammadreza Afshari [iD]

PhD Candidate of TEFL, Islamic Azad University, North Tehran Branch, Tehran, Iran

Seyed Reza Beh-Afarin [iD]

Assistant Professor of TEFL, Islamic Azad University, North Tehran Branch, Tehran, Iran

Jahanbakhsh Nikoopour [iD]

Assistant Professor of TEFL, Islamic Azad University, North Tehran Branch, Tehran, Iran

Abstract

Task-based Language Teaching (TBLT) is a widely adopted pedagogical approach that emphasizes real-world tasks to enhance second language acquisition. In a bid to empirically assess TBLT effectiveness, this study focused on the extent to which TBLT functions in the development of EAP students' pragmatic competence. To this end, 150 adult undergraduate Iranian students from various majors participated in this study. Also, three authentic role play tasks were designed based on a needs analysis, focusing on scenarios relevant to an EAP setting. Given the data nature, parametric statistical approach in the form of MANOVA of both pre-and post-test data was run to measure the students' pragmatic competence prior to the intervention and following it. The pretest data-based MANOVA revealed no significant differences among the four groups not only in their overall pragmatic competence but also in the target sub-competences of it including instrumental, regulatory, interactional, personal, explanation, share knowledge, and imagination competences; thereby indicating groups homogeneity prior to the treatments. However, the post-test data analysis in the light of MANOVA resulted in an exactly opposite direction in that significant differences were reported in all six sub-competences, underscoring the effectiveness of task-based assessment methods in enhancing pragmatic competence in general. Post hoc analysis also confirmed the post-test data-based MANOVA. So, the findings underscore that authentic task-based assessment effectively enhances students' pragmatic competence, fostering their ability to use the language appropriately and confidently in real-life communication situations. This study underscores the significance of methodological rigor in evaluating pragmatic competence in educational contexts. It can be concluded that authentic task-based assessment is beneficial for developing the pragmatic competence of EAP students.

Corresponding author's email: r_behafarin@iau-tnb.ac.ir

INTRODUCTION

Task-based Language Teaching (TBLT) focuses on real-world tasks in language learning, emphasizing authentic language use and specific goals rather than traditional vocabulary and grammar instruction (Shehadeh, 2005). TBLT's effectiveness in facilitating second language acquisition has led to its widespread use in language classrooms worldwide, utilizing task-based activities to provide students with essential learning data (Ellis, 2000). This pedagogical approach integrates educational philosophy, SLA theories, and effective teaching strategies to meet the demands of contemporary language learning environments (Van den Branden et al., 2009). While TBLT has been extensively discussed in educational literature and implemented in language schools, this study focused on exploring its assessment, characteristics, components, development, implementation procedures, and classroom applications (Ellis & Shintani, 2021; Mukhopadhyay & Sudharshana (Eds.), 2021; Sudharshana & Malicka, 2023).

Authentic task-based assessment focuses on assessing students in real-world tasks that are meaningful, and realistic. Authentic task-based assessment aims to assess language learning more engaging, relevant, and effective by providing students with opportunities to apply their language skills in various authentic contexts. This method emphasizes the importance of students actively using language in authentic contexts rather than just being assessed in grammatical rules or vocabulary in isolation. Tasks are designed to require students to use language creatively and purposefully to achieve a specific goal or complete a meaningful activity. By engaging in authentic tasks, students have the opportunity to develop their language skills in a more natural and meaningful way as they are exposed to the types of language they would encounter in real-life situations. This approach can help students improve their communication skills, fluency, and confidence in using the language (Ellis, 2003; Nunan, 2004).

Authentic task-based assessment involves designing activities that mirror real-world tasks and situations. These tasks are meaningful and

purposeful, requiring students to use language in real contexts rather than just focusing on grammar rules or vocabulary lists. In language learning, authentic tasks offer several benefits. Authentic tasks provide students with context, helping them understand how language is used in real-life situations. This contextual understanding not only enhances language comprehension but also fosters practical application, allowing students to navigate real-world scenarios with confidence as they learn language in context. Tasks that are authentic and relevant to students' lives make language learning more engaging and motivating. When tasks resonate with students' interests and daily experiences, they are more likely to invest their energy and focus on the learning process, leading to better retention and use of language skills in authentic situations. By completing authentic tasks, students practice their communication skills in a meaningful way, improving their ability to interact effectively in the target language. Engaging in real-world tasks cultivates not only linguistic proficiency but also promotes fluency, confidence, and cultural sensitivity, essential components of effective communication in diverse linguistic settings (Ahmadian & García Mayo, 2019; Van den Branden, Van Gorp, & Verhelst, 2007).

Task-Based Language Assessment (TBLA) follows the principles of TBLT, focusing on tasks as the central component of assessment to reflect real-world language use (Long & Norris, 2000). TBLA aims to establish a close link between test performance and practical language skills, emphasizing meaningful communication and specific goals (Ellis, 2005). Coombe et al. (2012) identify four critical characteristics of TBLA: formative assessment integrated into teaching, performance-based evaluation, direct assessment through embedded task performance measurement, and emphasis on authenticity in task design to mirror real-world language contexts. This approach requires observing test-takers' performance to make inferences about their underlying language competence, ensuring that the assessment captures authentic language use effectively.

TBLA plays a significant role in language education as it not only facilitates language learning but also promotes meaningful communication,

fosters motivation, and encourages collaboration among students. TBLA is widely recognized as an effective approach that emphasizes learning through the completion of meaningful tasks. Tasks in language learning can vary from real-world activities like ordering food or giving directions to more classroom-based exercises like problem-solving tasks or role plays. One of the key benefits of TBLA is its focus on promoting communication and language use in authentic contexts. By engaging students in tasks that simulate real-life situations, students are encouraged to apply their language skills and knowledge in practical ways which can help them improve their fluency and communicative competence. Furthermore, TBLA can enhance learner motivation and engagement by providing meaningful and relevant learning experiences. Instead of just memorizing grammar rules or vocabulary lists, students are actively using language to accomplish specific goals which can make learning process more enjoyable and rewarding. Another advantage of TBLA is its ability to foster collaboration and interaction among students. Tasks often require students to work together, communicate effectively, and negotiate meaning, which can help develop their interpersonal skills as well as their language proficiency (Ellis, 2005).

The challenge faced by task-based testing strategies lies in designing assessments that go beyond linguistic competence to evaluate actual performance in language use, emphasizing communicative proficiency and specified goals (Skehan et al., 2020). TBLA aligns with TBLT principles. However, it focuses on testing rather than teaching, requiring tasks as fundamental units for analyzing task performance, item selection, and performance rating (Norris, 2009). TBLA aims to connect test performance with real-world language use, extensively emphasizing communicative tasks to assess proficiency and achievement (Ellis et al., 2020). The significance of examining communicative performance is accentuated across proficiency tests (e.g., interactive ability tests) and achievement tests, aiming to predict students' performance realistically within specific language use domains (Skehan et al., 2020). The task difficulty assessment in TBLA involves

considering linguistic, communication demand, and cognitive complexity dimensions to determine overall difficulty ratings.

Research into performance-based testing, such as the Cambridge Main Suite examinations, demonstrates a progression towards task complexity in various levels of tasks in speaking assessments, incorporating interactive-ability models and supporting language use abilities (Urquhart & Weir, 2014; Galaczi & Taylor, 2018). This evolution progresses from controlled to open-ended formats, incorporating various supports and transitioning from facts to evaluations, highlighting the communicative nature and principled importance of tasks in speaking assessment (Ellis et al., 2020).

Studies like Brown-Norris approach and the Belgian group's achievement testing methods explore task-based test development and challenges, contributing to task-based assessment understanding and application in educational settings (Colpin & Gysen, 2006; Norris, 2009). Task-based evaluation serves formative and summative purposes; assisting students in enhancing language competence through self-portfolio, peer assessment, task selection, setting standards, administration, feedback provision, and improving performance on both task-centered and construct-centered language usage levels (Weaver & Gere, 2012). When implementing task-based assessment, the integrated approach of aligning assessments, curricula, and instruction around tasks contributes to meaningful language learning outcomes, emphasizing communicative proficiency and authentic language tasks in L2 exams (Ke, 2006; Willis, 1996). Task-based assessments offer concrete evidence of students' language achievements, opportunities for varied performance stages, instructional feedback, and integrated skills evaluation, enhancing language programs and teaching practices (Norris, 2009).

Pragmatic competence is essential in English for Academic Purposes (EAP) as it enables students to effectively communicate in academic and professional settings by understanding and applying appropriate language use in different social contexts. This skill helps students navigate interactions,

engage in critical thinking, develop cultural awareness, and improve both academic and professional communication. In EAP, pragmatic competence not only enhances language comprehension but also allows students to use language strategically, fostering both personal and academic growth (Bardovi-Harlig & Bastos, 2011). Educators and students alike should prioritize the development of pragmatic competence to succeed in the diverse and dynamic landscape of academic communication. Pragmatic competence plays a crucial role in the realm of EAP as it involves the ability to use language effectively in specific academic contexts. EAP focuses on developing students' English language skills to effectively communicate and succeed in academic settings such as universities or research environments. Pragmatic competence ensures that students can communicate their ideas clearly and appropriately in academic settings. Understanding the nuances of language use, such as politeness strategies, discourse markers, and appropriate tone, is essential for successful communication in academic writing and speaking. In academic environments, students are required to engage in various social interactions, such as group discussions, presentations, and collaborations. Pragmatic competence helps students navigate these interactions by understanding the expected norms, conventions, and communication styles within academic communities (Taguchi, 2019).

Developing pragmatic competence in EAP requires students to analyze language use in context, which in turn enhances their critical thinking skills. By understanding how language functions in academic discourse, students can critically evaluate arguments, identify biases, and construct well-supported claims in their own academic writing. Pragmatic competence also involves an awareness of cultural differences in communication styles and norms. In an academic context that is increasingly globalized, students need to navigate cultural diversity in interactions with peers and professors. Understanding cultural nuances in language use can prevent misunderstandings and foster effective communication across diverse academic communities.

Pragmatic competence is highly valued in academic and professional settings. Students who demonstrate strong pragmatic competence are better equipped to present themselves professionally through academic writing, presentations, and interactions with peers and instructors. This skill is essential for academic success and future career prospects. When students develop pragmatic competence in EAP, they not only improve their language skills but also enhance their overall learning experience. Effective communication fosters better collaboration with classmates, deeper engagement with course materials, and a more enriching academic journey (Clennell, 1999; Taguchi, 2019).

Pragmatic assessment studies in L2 education have drawn interest by operationalizing L2 pragmatics through politeness theory, speech act theory, and pragmalinguistics concepts. Hudson et al. (1995) developed measures like OPDCT, MCDCTs, oral DCT, role plays, and rating criteria to evaluate pragmatic competence, showing reliability and validity in assessing students' pragmatic performance. Despite its impact, critiques by Brown and Ahn (2011), Roever (2006) highlighted flaws in Hudson et al.'s (1995) study, with efforts to address issues like distractor generation in MCDCTs to improve reliability.

Various researchers like Roever (2006) have developed instruments to assess pragmatic competence, exploring constructs like routines, implicatures, and speech acts. Grabowski (2016) and Walters (2009) conducted mixed-methods and discourse analysis studies, respectively, to investigate grammar-pragmatics validity and critique the limitations of DCT-based tests, advocating for conversation analysis-informed test methods (CAIT) to enhance pragmatic assessment through sub-skills evaluation and interactional features detection. Walter's CAIT approach demonstrates improved rating administration and empirically valid results.

Furthermore, the focus on offline pragmatic performance in rating criteria overlooks the significance of online pragmatic production, posing a threat to assessment validity and warranting the reform of pragmatic competence assessment measures (Messick, 1994). The lack of context-

specific pragmatic competence assessment, particularly in fields like English for Specific Purposes (ESP) and English for Academic Purposes (EAP), highlights the necessity for tailored assessments beyond general evaluation of pragmatic competence (Kasper & Ross, 2002).

Pragmatic competence in language education is crucial as it pertains to the ability to use language appropriately in social contexts. It goes beyond just knowing the language rules and grammar; it involves understanding how language is used in different situations to convey meaning effectively and appropriately. Developing pragmatic competence in language education is vital for equipping students with the tools to communicate effectively and confidently in diverse social settings. Having pragmatic competence allows language students to interact with others in a culturally sensitive and context-appropriate manner. It empowers students to navigate various social situations, understand implied meanings, interpret non-verbal cues, and adapt their language use accordingly. This skill is essential for effective communication, building relationships, and avoiding misunderstandings. In language education, fostering pragmatic competence involves exposing students to authentic language use through real-life conversations, role plays, and cultural activities. Teachers play a crucial role in guiding students to develop this competence by providing opportunities for practice, feedback, and reflection (Taguchi & Roever, 2017; Huang, 2022).

TBLA serves as a valuable tool in fostering students' pragmatic competence by allowing them to practice using language in contextually appropriate ways and preparing them to effectively communicate in various real-world situations. TBLA is an approach to language learning that focuses on real-world tasks and activities to enhance students' language skills. When it comes to pragmatic competence which refers to the ability to use language appropriately in various social contexts, TBLA plays a crucial role. Through TBLA, students engage in activities that require them to use language in a meaningful way, such as role plays, problem-solving tasks, and simulations. These activities help students develop not only their linguistic abilities but also their pragmatic competence by providing opportunities to practice

language in authentic situations. By working on tasks that mimic real-life communication scenarios, students learn how to use language appropriately based on the social context, cultural norms, and the relationship between interlocutors. This hands-on approach helps students develop a deeper understanding of how language functions beyond just grammar and vocabulary, ultimately enhancing their pragmatic competence (Timpe-Laughlin, 2018; Taguchi & Kim, 2018).

While discourses around pragmatic competence encompass various testing methods, such as Discourse Completion Tasks (DCTs), scholars like Roever (2006) explored novel approaches to measuring pragmalinguistic knowledge through web-based instruments. The ongoing debate on the importance of grammar versus pragmatics for language students highlights the need for a comprehensive approach that integrates both facets effectively (Tsutagawa, 2013). Despite the utility of DCTs for testing pragmatic abilities, further studies on models like Purpura's (2004) promise enhanced insights into pragmatic competence assessment.

Roever (2006) introduced a web-based ESL pragmatics test evaluating students' knowledge in implicature, routines, and speech acts, recognizing their interdependence shaped by developmental factors like exposure and L1 proficiency. Validity, specially construct validity, and reliability are crucial considerations in assessing pragmatic competence (PC) in L2 education (Tsutagawa, 2013; Bachman & Palmer, 1996). Challenges arise when studies narrowly focus on speech acts, such as requests with small sample sizes, leading to proposals for performance data-driven rating criteria and analytical approaches to address the complexities of assessing pragmatic abilities in interactive contexts (Shohamy et al., 2017; Fulcher et al., 2013). Despite the analytical approach's utilization in various studies, uncertainties persist regarding raters' interpretation of rating category descriptions and their application in ensuring reliable scoring decisions (Youn & Chen, 2021).

This study addresses a notable gap in the current research on task-based language assessment, particularly in the context of English for Academic Purposes (EAP). While previous studies have examined the

effectiveness of TBLT in improving linguistic competence, which often fall short in assessing pragmatic competence. Additionally, as an effort to employ TBL-testing, this study can remedy the pitfalls associating the existing methodologies frequently critiqued for relying on outdated datasets, lacking transparency, or failing to consider essential variables such as cultural norms or the contextual use of language in academic settings. Unlike earlier research, which has primarily focused on grammatical and lexical skills, this study places a significant emphasis on how students use language meaningfully in academic tasks, such as role plays and simulations.

METHOD

Participants

One hundred and fifty Iranian adult undergraduate students from both genders learning English for academic purposes participated in this study. They were from different majors like physical education, management, psychology and electrical engineering. They were selected based availability and convenient sampling, though attempts were made to include participants from multiple majors. The students were enrolled in a Bachelor of Arts (BA) program and came from various ethnic backgrounds and represented the cultural diversity of Iran, with some having attended language schools prior to university, while others learned English through the university curriculum.

Instruments and Materials

A pretest and a posttest adopted from a valid existing pragmatic competence test were taken both at the outset and end of the intervention were used to measure the participants' L2 pragmatic competence. The focus of assessment was six dimensions of pragmatic competence: "instrumental," "regulatory," "personal," "interactional," "wants explanation," and "share knowledge and imagination." These categories reflected distinct components of pragmatic language use in EAP contexts and aligned with established theories of pragmatic competence in SLA. Instrumental competence referred to the

ability to use language effectively to achieve specific goals, such as making requests or giving instructions. Regulatory competence involved managing conversational flow and social control to ensure effective communication in academic interactions. Personal competence was associated with the expression of individual opinions, feelings, and identity through language. Interactional competence emphasized the importance of maintaining social relationships through effective conversation strategies like turn-taking. Additionally, wants explanation competence focused on articulating personal needs and desires in an appropriate manner. Lastly, share knowledge and imagination competence emphasized the collaborative exchange of ideas, information, and creativity, which is often required in academic group discussions or projects. These dimensions were based on pragmatic competence theories that treat language as a social tool, requiring not only linguistic proficiency but also contextual appropriateness. This framework provided a structured approach for task-based assessments and ensured alignment with broader SLA research, focusing on both the linguistic and pragmatic aspects of language use.

Data Collection Procedure

Having administered the pretest, the researcher developed three authentic role play tasks based on the results from the needs analysis. Furthermore, interviews with experts were run enabling the researcher to identify different situations that students might encounter with a range of interlocutors in an EAP setting, such as appropriately requesting a recommendation letter from a professor, politely asking for help in study skills, and using a dictionary. Two female and two male university professors who teach English for academic courses participated as professors for the role play task. Three of these interlocutors were Ph.D. candidates who had experience with content teaching at the university level. The other interlocutors were Ph.D. holding university ESL instructor. The four interlocutors received training by the researcher to standardize the conversation between participants; for example, to accept the students' request or prefer a certain option when two choices are

given. This was needed to minimize the effect of having four different interlocutors. This decision was also justified as the participants being evaluated were examinees rather than the professor interlocutors. The participants were randomly assigned to each interlocutor. These attempts were made to make sure of both the reliability of the role plays in the course of inter-rater reliability estimated through rater training as suggested by Bachman and Palmer (1996) as well as estimating test-retest reliability realized in the form of pretest and posttest possessing same characteristics and involving administering the same task to the same participants at different times to verify the stability of results (Creswell & Creswell, 2017). Test equating; parallels, normally ensures comparability across test versions (Kolen & Brennan, 2014) in large-scale assessments which often requires Item Response Theory (IRT) or other advanced psychometric methods. However, for this small-scale experimental design, involving fewer than 200 participants, implementing full IRT-based test equating was impractical.

In the same vein, following Bachman and Palmer (1996), small-scale studies often rely on content validity to ensure that pretests and posttests assess the same constructs. Then, expert reviews to ensure that the pretest and posttest measured the same pragmatic functions (Ellis & Shintani, 2021). The tasks were aligned to elicit similar types of pragmatic behavior, ensuring that the assessments were comparable in difficulty and content. Given the scope and scale of the study, this approach was deemed sufficient to ensure the integrity of the testing instruments. Furthermore content validity was ensured through expert judgment, which aligns with established methods in language testing (McNamara, 2000). Multiple experts in task-based language assessment reviewed the role-play tasks to determine whether they accurately represented real-world pragmatic functions. Additionally, confirmatory factor analysis (CFA) was conducted to make sure of the construct validity, ensuring that the role plays accurately measured the intended components of pragmatic competence (Field, 2018).

To complete all tasks, each participant was requested to meet with the researcher. The first meeting involved the briefing session in which each

participant was informed about all tasks and the role plays. The second meeting involved the participant engaging in a role play with the professor interlocutor. Each meeting was held with an individual participant. Instructions for the tasks were provided in detail so that participants clearly understood the tasks. All responses made during the tasks were audio-recorded.

The study could be at best practiced through a dynamic assessment approach; however, as outlined by Lantolf & Poehner (2004), dynamic assessment requires repeated measures and interaction over time, which was not feasible given the institutional restrictions and time constraints of our study. To compensate for, pretest-posttest design resembling a variety of dynamic assessment and aligning with traditional task-based assessment (TBA) frameworks used in second language acquisition research, as discussed by Ellis (2003) was implemented which by itself allowed us to capture overall learning gains without the added complexity of dynamic assessment. While repeated measures might provide more detailed insights into the evolving nature of pragmatic competence, our goal was to assess the effectiveness of a single intervention. This approach is commonly used in experimental language learning studies, particularly when resources and participant availability are limited (Mackey & Gass, 2005). The limitations of not using a dynamic assessment approach are acknowledged in the discussion section, but we emphasize that this did not detract from the study's primary goal. To implement the procedure, the following steps were followed:

Sample role play scenario (requesting a recommendation letter from a professor): The student was advised to approach the professor and seek an appointment to discuss the process of drafting a letter of recommendation. The student was encouraged to inquire if the professor would be willing to compose a recommendation letter on their behalf. The professor might have indicated current busyness but expressed availability to attend to this request in a few days. Upon meeting, the student extended greetings to the professor

and expressed gratitude for the opportunity to meet. The student formally requested the professor to write a letter of recommendation, emphasizing the significance of the document in securing a university scholarship. The student conveyed a lack of familiarity with the letter-writing process and expressed uncertainty regarding its contents. The professor reassured the student and promised to provide detailed guidance via email. Subsequently, the professor inquired if the student could advance their presentation slot, originally designated for a busy peer. Despite already committed plans, the student apologized and explained their prior arrangements. The professor reassured the student that the adjustment was permissible given the schedule.

Sample role play scenario (asking for help in study skills): Before the meeting, the student needed to obtain information about study skills and sought advice from her professor. The student approached the professor, requesting a meeting. The professor, being occupied at the time, scheduled to meet the student the following week. The student thanked the professor and confirmed her intention to visit the professor the next week for a discussion. During the meeting, the student greeted the professor enthusiastically, expressing his/her keenness to enhance his/her knowledge of study skills. The professor emphasized the significance of study skills, especially in language learning. The student inquired about references on the topic, to which the professor mentioned having several e-books that he could share. The student agreed to use e-books and also sought permission to share them with his/her peers. The professor consented, under the condition that credit is given for the e-books. The student expressed gratitude and sought permission to consult the professor for future assistance. The professor assured the student he would review his schedule and provide available times for future discussions. Subsequently, the student expressed appreciation once more and departed from the professor's office.

Sample role play scenario (Using a dictionary): On the day the student sent the email, he deliberated on using a dictionary addressed to his/her professor, the professor, with the intention of scheduling an in-person meeting to solicit

guidance on the correct utilization of a dictionary. Having acquainted himself with the conventions of formal email correspondence, he crafted and dispatched the aforementioned email to his/her professor. Subsequently, he awaited a response for a brief period. Upon receiving a reply outlining the specifics of their forthcoming meeting, including the date and time, the student proceeded to compile a list of inquiries and strategized his/her interaction with the professor to ensure a structured and coherent approach. On the appointed day of his/her rendezvous with his/her professor, the student courteously announced his/her presence by knocking on the door, sought permission to enter, and engaged in a courteous discourse with his/her professor. Following a concise dialogue concerning the student's current academic status, the student eloquently articulated his/her queries in a formal manner, as had been previously rehearsed. The professor attentively absorbed the student's questions and offered comprehensive insights on a range of dictionary types, encompassing online resources. The student diligently absorbed the professor's guidance, documenting key points during a discussion that spanned approximately 40 minutes, in which he addressed all of his/her prepared inquiries. Upon the culmination of the meeting, the student conveyed appreciation to his/her professor, reciprocated with warm regards from the professor to the student. Expressing a desire for potential future clarifications, the student expressed interest in possible subsequent visits. The professor acquiesced, suggesting that the student schedule future appointments in advance. The meeting drew to a close with a handshake, cordial farewells exchanged, and the student exiting the office premises.

In line with validity and reliability assurance and enhancement, a scoring rubric used for the role plays was developed based on established frameworks for pragmatic competence (Kasper & Rose, 2002). While it is true that not every role play may elicit all six functions with equal frequency, the tasks were specifically designed to cover a wide range of pragmatic behaviors. Each task was carefully reviewed by experts to ensure that key functions—instrumental, regulatory, personal, interactional, wants explanation, and share knowledge & imagination—were represented

(Taguchi, 2019). The scoring rubric followed best practices in task-based language assessment (TBLA), where each function was weighted according to its relevance to the task. As Roever (2006) points out, pragmatic assessment often involves a degree of variability in task performance, which was accounted for in our analytic scoring system. The rubric allowed for flexibility in scoring while maintaining consistency across different functions, with evaluators receiving specific training on how to assess these functions in each scenario.

This study aimed to address is the limited exploration of pragmatic competence development in English for Academic Purposes (EAP) contexts, particularly how students use language appropriately across various social and academic interactions. While much research has focused on linguistic aspects of second language acquisition, the pragmatic dimension—specifically, instrumental competence, which involves using language to achieve specific goals—has received less attention. This study targeted instrumental competence as a key aspect of pragmatic competence and explored how authentic task-based assessments, such as role play tasks, can enhance students' ability to use language meaningfully and appropriately in academic settings. To validate the analytical rating criteria used to assess the students' performance, a Confirmatory Factor Analysis (CFA) was conducted. The five rating categories—Contents Delivery, Language Use, Sensitivity to Situation, Engaging with Interaction, and Turn Organization—were examined to ensure they effectively measured the relevant components of L2 pragmatic competence. By incorporating three additional monologic tasks (two speaking tasks and one pragmatic task), the study further examined the relationship between performance on role play tasks and broader aspects of L2 proficiency and pragmatics. The detailed turn-by-turn analysis of participant performances allowed for a more refined understanding of task-specific characteristics and contributed to the development of robust rating criteria tailored to the varied pragmatic demands of each task.

Data Analysis

Prior to deciding on the sound statistical paradigm, the data were checked in terms of meeting normality assumptions including normality of data, homogeneity of variances of groups, and homogeneity of covariance matrices. The latter two assumptions will be discussed when reporting the results of MANOVAs Table 1 shows the skewness and kurtosis indices of normality. The results indicated that the present data did not show any significant deviation from normality due to the fact that the skewness and kurtosis indices were within the ranges of ±2. It should be noted that the criteria of ±2 were proposed by Bachman, 2005; Bae and Bachman, 2010; and George and Mallery, 2019. It should also be noted that Zhu et al. (2019) suggested the criteria of ±3. However, Watkins (2021) suggested different criteria for skewness and kurtosis. He states that skewness values should be less than ±2; while kurtosis indices should be evaluated against the criteria of ±7. Additionally, the assumption of linearity was violated which further justified employment of parametric statistical approach. So, the researcher ran two separate MANOVAs for comparing the four groups' means on pretests of pragmatic competence first; and then on posttests, comparing pre- and post-test performances across groups, with a focus on "instrumental," "regulatory," "personal," "interactional," "wants explanation," and "share knowledge and imagination" dimensions of pragmatic competence.

Table 1. Skewness and Kurtosis Indices of Normality

Group		N	Skewness		Kurtosis	
		Statistic	Statistic	Std. Error	Statistic	Std. Error
	PreInst	38	.45	.38	.39	.75
	PostInst	38	.05	.38	.09	.75
	PreReg	38	-.48	.38	-.60	.75
	PostReg	38	-.97	.38	1.01	.75
	PrePers	38	.58	.38	-.39	.75
Role Play	PostPers	38	-.24	.38	-.92	.75
	PreInter	38	-.26	.38	-1.04	.75
	PostInter	38	.06	.38	-.07	.75
	PreWant	38	-.43	.38	-.15	.75
	PostWant	38	.12	.38	-.19	.75

	PreShare	38	.09	.38	-.81	.75
	PostShare	38	-.74	.38	.52	.75
Dictionary	PreInst	36	-.08	.39	-.95	.77
	PostInst	36	-.01	.39	1.34	.77
	PreReg	36	-.61	.39	-.65	.77
	PostReg	36	.48	.39	.27	.77
	PrePers	36	.54	.39	-.38	.77
	PostPers	36	.24	.39	.47	.77
	PreInter	36	.40	.39	-.33	.77
	PostInter	36	.57	.39	-.06	.77
	PreWant	36	-.02	.39	-40	.77
	PostWant	36	.64	.39	.42	.77
	PreShare	36	.56	.39	-.55	.77
	PostShare	36	.57	.39	.67	.77
Study Skills	PreInst	39	-.05	.38	-.13	.74
	PostInst	39	-.46	.38	-.41	.74
	PreReg	39	.23	.38	-1.00	.74
	PostReg	39	-.23	.88	-.17	.74
	PrePers	39	.11	.38	-.80	.74
	PostPers	39	-.40	.38	-.09	.74
	PreInter	39	.04	.38	-.36	.74
	PostInter	39	-.15	.38	-.57	.74
	PreWant	39	-.22	.38	-.45	.74
	PostWant	39	-.26	.38	-.16	.74
	PreShare	39	-.14	.38	-1.05	.74
	PostShare	39	-.20	.37	-.69	.74
Control	PreInst	38	.14	.38	-.17	.75
	PostInst	38	.68	.38	.09	.75
	PreReg	38	-.75	.38	-.39	.75
	PostReg	38	.11	.38	-.29	.75
	PrePers	38	.85	.38	.52	.75
	PostPers	38	.12	.38	-.67	.75
	PreInter	38	.29	.38	-.20	.75
	PostInter	38	.52	.38	-.60	.75
	PreWant	38	-.22	.38	-.13	.75
	PostWant	38	.50	.38	-.56	.75
	PreShare	38	.28	.38	-.00	.75
	PostShare	38	.08	.38	-.62	.75

Note. Throughout this report, the following abbreviations are employed; Pre = Pretest of, Post = Posttest of, Inst = Instrumental, Reg = Regulatory, Per = Personal, Int = Interactional, Want = Wants explanation, and Share = Share knowledge and imagination.

Comparing Groups' Means on Pretests of Pragmatic Competence

The four groups' means on six components of pragmatic competence were compared using MANOVA. It was mentioned earlier that, besides the assumption of normality, MANOVA requires homogeneity of variances of

groups, and homogeneity of covariance matrices. Table 2 shows the Levene's test of homogeneity of variances for pretests of pragmatic competence. The results indicated that the assumption of homogeneity of variances was retained on instrumental ($F_{(3, 147)}$ = .309, p > .05), regulatory ($F_{(3, 147)}$ = .212, p > .05), personal ($F_{(3, 147)}$ = 1.43, p > .05), interactional ($F_{(3, 147)}$ = 1.39, p > .05), wants explanation ($F_{(3, 147)}$ = 1.10, p > .05), and share knowledge & imagination ($F_{(1, 147)}$ = .658, p > .05).

Table 2. Levene's Test of Homogeneity of Variances for Pretests of Pragmatic Competence by Group

		Levene Statistic	df1	df2	Sig.
PreInst	Based on Mean	.15	3	147	.93
	Based on Median	.31	3	147	.82
	Based on Median and with adjusted df	.31	3	144.58	.82
	Based on trimmed mean	.14	3	147	.93
PreReg	Based on Mean	.86	3	147	.47
	Based on Median	.21	3	147	.89
	Based on Median and with adjusted df	.21	3	139.10	.89
	Based on trimmed mean	.82	3	147	.49
PrePers	Based on Mean	2.07	3	147	.11
	Based on Median	1.44	3	147	.24
	Based on Median and with adjusted df	1.44	3	136.17	.24
	Based on trimmed mean	2.03	3	147	.11
PreInter	Based on Mean	1.86	3	147	.14
	Based on Median	1.40	3	147	.25
	Based on Median and with adjusted df	1.40	3	120.83	.25
	Based on trimmed mean	1.84	3	147	.14
PreWant	Based on Mean	1.77	3	147	.16
	Based on Median	1.11	3	147	.35
	Based on Median and with adjusted df	1.11	3	140.94	.35
	Based on trimmed mean	1.57	3	147	.20
PreShare	Based on Mean	1.02	3	147	.39
	Based on Median	.66	3	147	.58
	Based on Median and with adjusted df	.66	3	140.290	.58
	Based on trimmed mean	1.02	3	147	.39

Table 3 shows the Box's test of homogeneity of covariance matrices. Multivariate ANOVA also requires that the correlations between any two variables be roughly equal across the four groups. The results (Box's M =

97.40, p > .001) indicated that the assumption of homogeneity of covariance matrices was retained. It should be noted that the results of the reported at .001 levels; Pallant, 2016l; Field, 2018, and Tabachnick & Fidell, 2019.

Table 3. Box's Test of Equality of Covariance Matrices for Pretests of Pragmatic Competence by Group

Box's M	97.41
F	1.43
df1	63
df2	50322.25
Sig.	.01

The main results of MANOVA will be discussed next. Table 4 shows the results of MANOVA. The results (F (18, 432) = .905, p > .05, partial η^2 = .036 representing a weak effect size[1]) indicated that there were not any significant differences between six components' overall means on pretests of pragmatic competence. The groups' means on each of the components of pretests of pragmatic competence will be discussed in Table 4.5 and Table 4.6.

Table 4. Multivariate Tests for Pretests of Pragmatic Competence by Group

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.98	2226.9	6	142	.00	.98
	Wilks' Lambda	.01	2226.9	6	142	.00	.98
	Hotelling's Trace	94.09	2226.9	6	142	.00	.98
	Roy's Largest Root	94.09	2226.9	6	142	.00	.98
Group	Pillai's Trace	.10	.90	18	432	.57	.03
	Wilks' Lambda	.89	.89	18	402.1	.58	.03
	Hotelling's Trace	.11	.89	18	422	.58	.03
	Roy's Largest Root	.05	1.36	6	144	.23	.05

Table 5 shows the four groups' means on pretests of pragmatic competence. The results indicated that the four groups had roughly equal means on pretests

---

[1] Partial Eta Squared should be interpreted using the following criteria; .01 = Weak, .06 = Moderate, and .14 = Large (Gray and Kinnear 2012, p 323).

of pragmatic competence. Based on these results and the results of the Between-Subject Effects (Table 6) it can be concluded that;

     A: There were no significant differences between the four groups' means on the pretest of instrumental competence, $F(3, 147) = 0.822$, $p > .05$, $p\eta^2 = .017$, indicating a weak effect size. Thus, the groups were homogeneous in terms of their instrumental competence prior to the implementation of the treatments.

Table 5. Descriptive Statistics for Pretests of Pragmatic Competence by Group

Dependent Variable	Group	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
PreInst	Role Play	6.26	.12	6.02	6.51
	Dictionary	6.06	.13	5.81	6.30
	Study Skills	6.03	.12	5.79	6.27
	Control	6.18	.12	5.94	6.43
PreReg	Role Play	3.42	.10	3.22	3.62
	Dictionary	3.36	.10	3.16	3.57
	Study Skills	3.56	.10	3.37	3.76
	Control	3.45	.10	3.25	3.65
PrePers	Role Play	8.29	.15	7.99	8.59
	Dictionary	7.89	.16	7.58	8.20
	Study Skills	7.95	.15	7.66	8.24
	Control	8.11	.15	7.81	8.40
PreInter	Role Play	21.158	.392	20.383	21.933
	Dictionary	21.000	.403	20.204	21.796
	Study Skills	20.692	.387	19.927	21.457
	Control	21.263	.392	20.488	22.038
PreWant	Role Play	5.474	.111	5.254	5.693
	Dictionary	5.278	.114	5.052	5.503
	Study Skills	5.333	.110	5.117	5.550
	Control	5.500	.111	5.281	5.719
PreShare	Role Play	10.211	.194	9.828	10.593
	Dictionary	9.944	.199	9.551	10.338
	Study Skills	9.744	.191	9.366	10.121
	Control	10.053	.194	9.670	10.435

B: There were no significant differences between the four groups' means on the pretest of regulatory competence, $F(3, 147) = 0.714$, $p > .05$, $p\eta^2 = .014$,

indicating a weak effect size. Therefore, the groups were homogeneous in their regulatory competence before the treatments.

Table 6. Tests of Between-Subjects Effects for Pretests of Pragmatic Competence by Group

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	PreInst	1.41	3	.47	.82	.48	.02
	PreReg	.82	3	.28	.71	.55	.01
	PrePers	3.62	3	1.21	1.40	.25	.03
	PreInter	7.17	3	2.39	.41	.75	.01
	PreWant	1.30	3	.43	.92	.43	.02
	PreShare	4.44	3	1.48	1.04	.38	.02
Error	PreInst	83.94	147	.57			
	PreReg	56.55	147	.39			
	PrePers	126.85	147	.86			
	PreInter	858.73	147	5.84			
	PreWant	68.86	147	.47			
	PreShare	209.54	147	1.43			
Total	PreInst	5764.00	151				
	PreReg	1855.00	151				
	PrePers	9939.00	151				
	PreInter	67625.00	151				
	PreWant	4469.00	151				
	PreShare	15274.00	151				

C: There were no significant differences between the four groups' means on the pretest of personal competence, $F(3, 147) = 1.39$, $p > .05$, $p\eta^2 = .014$, indicating a weak effect size. Thus, the groups were homogeneous in personal competence before the treatments.

   D: No significant differences were found between the four groups' means on the pretest of interactional competence, $F(3, 147) = 0.409$, $p > .05$, $p\eta^2 = .008$, indicating a weak effect size. Therefore, the groups were homogeneous in their interactional competence prior to the treatments.

   E: There were no significant differences between the four groups' means on the pretest of wants explanation competence, $F(3, 147) = 0.922$, $p > .05$, $p\eta^2 = .018$, indicating a weak effect size. This confirmed homogeneity in terms of wants explanation competence prior to the treatments.

F: Finally, there were no significant differences between the four groups' means on the pretest of share knowledge and imagination competence, $F(3, 147) = 1.03$, $p > .05$, $p\eta^2 = .021$, indicating a weak effect size. Thus, the groups were homogeneous in terms of share knowledge and imagination competence prior to the treatments.
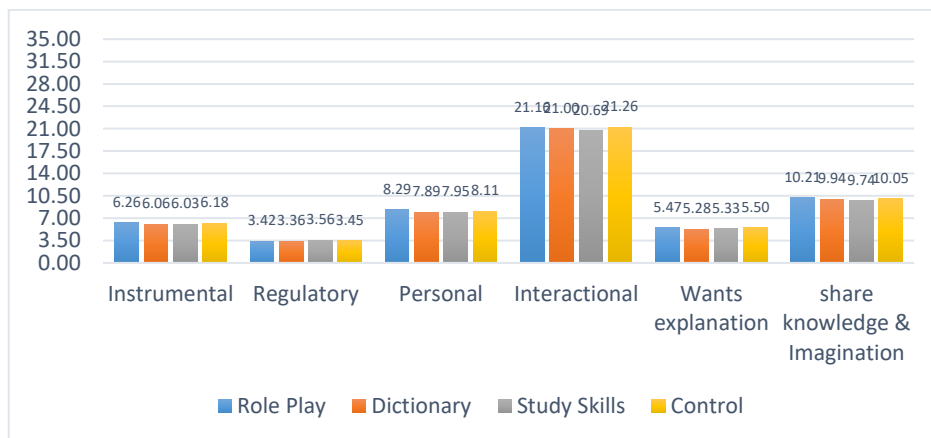

Figure 1. Means on Pretests of Pragmatic Competence by Group

Testing Null-Hypothesis

The only null-hypothesis raised in this study stated that EAP students' pragmatic competence cannot be significantly developed in the light of authentic task-based assessment. Multivariate ANOVA was run to compare the four groups' means on posttests of pragmatic competence in order to probe the null-hypothesis. It was mentioned earlier that, besides the assumption of normality, MANOVA requires homogeneity of variances of groups, and homogeneity of covariance matrices.

Table 7 shows the Levene's test of homogeneity of variances for posttests of pragmatic competence. The results indicated that the assumption of homogeneity of variances was retained on regulatory (F (3, 147) = .559, p > .05), personal (F (3, 147) = 1.83, p > .05), interactional (F (3, 147) = 2.42, p > .05), wants explanation (F (3, 147) = 1.85, p > .05), and share knowledge & imagination (F (1, 147) = 138, p > .05). However, it was violated on

instrumental (F (3, 147) = 3.29, p < .05). As noted by Tabachnick, Fidell & Ullman, 2013; in case the assumption of homogeneity of variances is violated, the alpha level should be reduced to .01 instead of .05. That was why the results related to instrumental competence in Table 4.12 and Table 4.13 were reported at .01 levels.

Table 4.7: Levene's Test of Homogeneity of Variances for Posttests of Pragmatic Competence by Group

		Levene Statistic	df1	df2	Sig.
PostInst	Based on Mean	2.64	3	147	.05
	Based on Median	3.30	3	147	.02
	Based on Median and with adjusted df	3.30	3	136.8	.02
	Based on trimmed mean	2.59	3	147	.06
PostReg	Based on Mean	1.57	3	147	.20
	Based on Median	.56	3	147	.64
	Based on Median and with adjusted df	.56	3	131.5	.64
	Based on trimmed mean	1.56	3	147	.20
PostPers	Based on Mean	1.72	3	147	.17
	Based on Median	1.84	3	147	.14
	Based on Median and with adjusted df	1.84	3	144.5	.14
	Based on trimmed mean	1.68	3	147	.17
PostInter	Based on Mean	2.83	3	147	.04
	Based on Median	2.43	3	147	.07
	Based on Median and with adjusted df	2.43	3	128.2	.06
	Based on trimmed mean	2.85	3	147	.04
PostWant	Based on Mean	3.09	3	147	.029
	Based on Median	1.86	3	147	.139
	Based on Median and with adjusted df	1.86	3	124.372	.140
	Based on trimmed mean	3.06	3	147	.030
PostShare	Based on Mean	2.32	3	147	.077
	Based on Median	1.39	3	147	.249
	Based on Median and with adjusted df	1.39	3	123.166	.249
	Based on trimmed mean	2.22	3	147	.088

Table 8 shows the Box's test of homogeneity of covariance matrices. Multivariate ANOVA also requires that the correlations between any two variables be roughly equal across the four groups. The results (Box's M = 106.72, p > .001) indicated that the assumption of homogeneity of covariance matrices was retained. It should be noted that the results of the reported at .001 levels; Pallant, 2016 et al. and Tabachnick & Fidell, 2019.

Table 8. Box's Test of Equality of Covariance Matrices for Posttests of Pragmatic Competence by Group

Box's M	106.72
F	1.57
df1	63
df2	50322.25
Sig.	.003

The main results of MANOVA will be discussed next. Table 4.10 shows the results of MANOVA. The results (F (18, 432) = 17.58, p < .05, partial η^2 = .423 representing a large effect size) indicated that there were significant differences between six components' overall means on posttests of pragmatic competence. Thus; the null-hypothesis was rejected. The groups' means on each of the components of posttests of pragmatic competence will be discussed in Tables 10, 11 and 12.

Table 9. Multivariate Tests for Posttests of Pragmatic Competence by Group

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.99	2661.63	6	142	.00	.99
	Wilks' Lambda	.01	2661.36	6	142	.00	.99
	Hotelling's Trace	112.46	2661.63	6	142	.00	.99
	Roy's Largest Root	112.46	2661.63	6	142	.00	.99
Group	Pillai's Trace	1.27	17.59	18	432	.00	.42
	Wilks' Lambda	.08	32.85	18	402.12	.00	.57
	Hotelling's Trace	7.65	59.74	18	422	.00	.72
	Roy's Largest Root	7.08	169.80	6	144	.00	.88

Table 10 shows the four groups' means on posttests of pragmatic competence. The results indicated that role play group had the highest mean on all six components of posttests of pragmatic competence. These were followed by the dictionary and study skills groups. The control group had the lowest mean on all six posttests. Based on these results and the results of the Between-Subject Effects (Table 11), and the results of the Scheffe's post-hoc tests (Table 12) it can be concluded that;

A: There were significant differences between the four groups' means on posttest of instrumental competence (F (3, 147) = 196.13, p < .01[2], pη² = .800 representing a large effect size). The results of post-hoc Scheffe's tests (Table 4.13) indicated that a) The role play group (M[3] = 13.97) had a significantly higher mean than the dictionary (M = 12.11) group (MD = 1.86, p < .01), b) The role play group (M = 13.97) had a significantly higher mean than the study skills (M = 9.89) group (MD = 4.08, p < .01), c) The role play group (M = 13.97) had a significantly higher mean than the control (M = 7.42) group (MD = 6.55, p < .01), d) The dictionary group (M = 12.11) had a significantly higher mean than the study skills (M = 9.89) group (MD = 2.21, p < .01), e) The dictionary group (M = 12.11) had a significantly higher mean than the control (M = 7.42) group (MD = 4.69, p < .01), f) And finally, the study skills group (M = 9.89) had a significantly higher mean than the control (M = 7.42) group (MD = 2.48, p < .01).

Table 10. Descriptive Statistics for Posttests of Pragmatic Competence by Group

Dependent Variable	Group	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
PostInst	Role Play	13.97	.20	13.58	14.37
	Dictionary	12.11	.21	11.70	12.52
	Study Skills	9.90	.20	9.50	10.29
	Control	7.42	.20	7.02	7.82
PostReg	Role Play	9.42	.13	9.17	9.67
	Dictionary	8.17	.13	7.91	8.42
	Study Skills	6.62	.13	6.37	6.86
	Control	4.47	.13	4.22	4.72
PostPers	Role Play	19.97	.26	19.46	20.48
	Dictionary	16.83	.27	16.31	17.36
	Study Skills	14.21	.26	13.70	14.71
	Control	11.37	.26	10.86	11.88
PostInter	Role Play	43.87	.60	42.69	45.05
	Dictionary	37.22	.62	36.01	38.44
	Study Skills	31.10	.59	29.94	32.27
	Control	26.05	.60	24.87	27.24
PostWant	Role Play	14.63	.20	14.23	15.03

[2] Since the assumption of homogeneity of variances was violated in posttest of instrumental competence (Table 4.8), the results were reported at .01 levels.
[3] M and MD stand for Mean and Mean Difference.

	Dictionary	12.47	.21	12.06	12.89
	Study Skills	10.28	.20	9.89	10.68
	Control	8.76	.20	8.36	9.17
	Role Play	28.00	.35	27.30	28.70
PostShare	Dictionary	24.33	.36	23.62	25.05
	Study Skills	19.77	.35	19.08	20.46
	Control	13.61	.35	12.91	14.30

B: There were significant differences between the four groups' means on posttest of regulatory competence (F (3, 147) = 284.48, p < .05, pη² = .853 representing a large effect size). The results of post-hoc Scheffe's tests (Table 4.12) indicated that a) The role play group (M = 9.42) had a significantly higher mean than the dictionary (M = 8.16) group (MD = 1.25, p < .05), b) The role play group (M = 9.42) had a significantly higher mean than the study skills (M = 8.16) group (MD = 2.81, p < .05), c) The role play group (M = 9.42) had a significantly higher mean than the control (M = 4.47) group (MD = 4.95, p < .05), d) The dictionary group (M = 8.16) had a significantly higher mean than the study skills (M = 6.61) group (MD = 1.55, p < .05), e) The dictionary group (M = 8.16) had a significantly higher mean than the control (M = 4.47) group (MD = 3.69, p < .05), f) And finally, the study skills group (M = 6.61) had a significantly higher mean than the control (M = 4.47) group (MD = 2.14, p < .05).

Table 11. Tests of Between-Subjects Effects for Posttests of Pragmatic Competence by Group

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	PostInst	910.14	3	303.38	196.13	.00	.80
	PostReg	516.53	3	172.18	284.49	.00	.85
	PostPers	1537.85	3	512.62	201.93	.00	.81
	PostInter	6763.70	3	2254.57	165.71	.00	.77
	PostWant	749.13	3	249.71	159.20	.00	.77
	PostShare	4377.48	3	1459.16	309.97	.00	.86
Error	PostInst	227.38	147	1.55			
	PostReg	88.97	147	.61			
	PostPers	373.18	147	2.54			
	PostInter	2000.05	147	13.61			
	PostWant	230.58	147	1.57			
	PostShare	692.00	147	4.71			

	PostInst	18841	151
	PostReg	8330	151
Total	PostPers	38515	151
	PostInter	188526	151
	PostWant	21007	151
	PostShare	74076	151

C: There were significant differences between the four groups' means on posttest of personal competence ($F(3, 147) = 201.92$, $p < .05$, $p\eta^2 = .805$ representing a large effect size). The results of post-hoc Scheffe's tests (Table 4.12) indicated that a) The role play group (M = 19.97) had a significantly higher mean than the dictionary (M = 16.83) group (MD = 3.14, p < .05), b) The role play group (M = 19.97) had a significantly higher mean than the study skills (M = 14.20) group (MD = 5.77, p < .05), c) The role play group (M = 19.97) had a significantly higher mean than the control (M = 11.36) group (MD = 8.61, p < .05), d) The dictionary group (M = 16.83) had a significantly higher mean than the study skills (M = 14.20) group (MD = 2.63, p < .05), e) The dictionary group (M = 16.83) had a significantly higher mean than the control (M = 11.36) group (MD = 5.46, p < .05), f) And finally, the study skills group (M = 14.20) had a significantly higher mean than the control (M = 11.36) group (MD = 2.84, p < .05).

Table 12. Post-Hoc Scheffe's Tests for Posttests of Pragmatic Competence by Group

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PostInst	Role Play	Dictionary	1.86*	.29	.00	1.04	2.68
		Study Skills	4.08*	.28	.00	3.27	4.88
		Control	6.55*	.29	.00	5.75	7.36
	Dictionary	Study Skills	2.21*	.29	.00	1.40	3.03
		Control	4.69*	.29	.00	3.87	5.51
	Study Skills	Control	2.48*	.28	.00	1.67	3.28
PostReg	Role Play	Dictionary	1.25*	.18	.00	.74	1.77
		Study Skills	2.81*	.18	.00	2.30	3.31
		Control	4.95*	.18	.00	4.44	5.45
	Dictionary	Study Skills	1.55*	.18	.00	1.04	2.06
		Control	3.69*	.18	.00	3.18	4.20
	Study Skills	Control	2.14*	.18	.00	1.64	2.64
PostPers	Role Play	Dictionary	3.14*	.37	.00	2.09	4.19
		Study Skills	5.77*	.36	.00	4.74	6.80

	Control	8.61*	.37	.00	7.57	9.64
Dictionary	Study Skills	2.63*	.37	.00	1.59	3.67
	Control	5.46*	.37	.00	4.42	6.51
Study Skills	Control	2.84*	.36	.00	1.81	3.86
	Dictionary	6.65*	.86	.00	4.22	9.07
Role Play	Study Skills	12.77*	.84	.00	10.39	15.14
	Control	17.82*	.85	.00	15.42	20.21
Dictionary	Study Skills	6.12*	.85	.00	3.71	8.53
	Control	11.17*	.86	.00	8.74	13.60
Study Skills	Control	5.05*	.84	.00	2.67	7.43
	Dictionary	2.16*	.29	.00	1.34	2.98
Role Play	Study Skills	4.35*	.29	.00	3.54	5.16
	Control	5.87*	.29	.00	5.06	6.68
Dictionary	Study Skills	2.19*	.29	.00	1.37	3.01
	Control	3.71*	.29	.00	2.89	4.53
Study Skills	Control	1.52*	.29	.00	.71	2.33
	Dictionary	3.67*	.51	.00	2.24	5.09
Role Play	Study Skills	8.23*	.50	.00	6.83	9.63
	Control	14.39*	.50	.00	12.99	15.80
Dictionary	Study Skills	4.56*	.50	.00	3.15	5.98
	Control	10.73*	.51	.00	9.30	12.16
Study Skills	Control	6.16*	.50	.00	4.77	7.56

(PostInter group label spans the Role Play / Dictionary / Study Skills block; PostWant and PostShare label the subsequent blocks.)

*. The mean difference is significant at the .05 level.

D: There were significant differences between the four groups' means on posttest of interactional competence (F (3, 147) = 165.70, p < .05, pη² = .772 representing a large effect size). The results of post-hoc Scheffe's tests (Table 12) indicated that a) The role play group (M = 43.86) had a significantly higher mean than the dictionary (M = 37.22) group (MD = 6.65, p < .05), b) The role play group (M = 43.86) had a significantly higher mean than the study skills (M = 31.10) group (MD = 12.77, p < .05), c) The role play group (M = 43.86) had a significantly higher mean than the control (M = 26.05) group (MD = 17.82, p < .05), d) The dictionary group (M = 37.22) had a significantly higher mean than the study skills (M = 31.10) group (MD = 6.12, p < .05), e) The dictionary group (M = 37.22) had a significantly higher mean than the control (M = 26.05) group (MD = 11.17, p < .05), f) And finally, the study skills group (M = 31.10) had a significantly higher mean than the control (M = 26.05) group (MD = 5.05, p < .05).

E: There were significant differences between the four groups' means on posttest of wants explanation competence (F (3, 147) = 159.19, p < .05, pη^2 = .765 representing a large effect size). The results of post-hoc Scheffe's tests (Table 12) indicated that a) The role play group (M = 14.63) had a significantly higher mean than the dictionary (M = 12.47) group (MD = 2.16, p < .05), b) The role play group (M = 14.63) had a significantly higher mean than the study skills (M = 10.28) group (MD = 4.35, p < .05), c) The role play group (M = 14.63) had a significantly higher mean than the control (M = 8.76) group (MD = 5.87, p < .05), d) The dictionary group (M = 12.47) had a significantly higher mean than the study skills (M = 10.28) group (MD = 2.19, p < .05), e) The dictionary group (M = 12.47) had a significantly higher mean than the control (M = 8.76) group (MD = 3.71, p < .05), f) And finally, the study skills group (M = 1028) had a significantly higher mean than the control (M = 8.76) group (MD = 1.52, p < .05).

F: There were significant differences between the four groups' means on posttest of share knowledge and imagination competence (F (3, 147) = 309.96, p < .05, pη^2 = .863 representing a large effect size). The results of post-hoc Scheffe's tests (Table 12) indicated that a) The role play group (M = 28.00) had a significantly higher mean than the dictionary (M = 24.33) group (MD = 3.67, p < .05), b) The role play group (M = 28.00) had a significantly higher mean than the study skills (M = 19.76) group (MD = 8.23, p < .05), c) The role play group (M = 28.00) had a significantly higher mean than the control (M = 13.60) group (MD = 14.29, p < .05), d) The dictionary group (M = 24.33) had a significantly higher mean than the study skills (M = 19.76) group (MD = 4.56, p < .05), e) The dictionary group (M = 24.33) had a significantly higher mean than the control (M = 13.60) group (MD = 10.73, p < .05), f) And finally, the study skills group (M = 19.76) had a significantly higher mean than the control (M = 13.60) group (MD = 6.16, p < .05). Figure 4.2 shows the four groups' means on posttests of pragmatic competence.
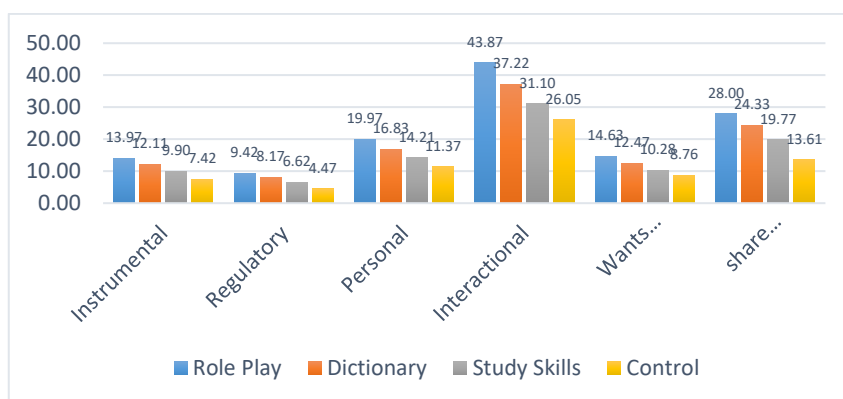
Figure 2. Means on Posttests of Pragmatic Competence by Group

The researcher attempted to apply Multivariate Analysis of Covariances (MANCOVA) to compare the means of four groups on posttests of various pragmatic aspects while controlling for pretest effects. However, it was found that the assumption of linearity, crucial for MANCOVA, was violated. The results of the linearity test suggested that for instrumental, regulatory, personal, interactional, wants explanation, and share knowledge and imagination aspects, there was no significant linear relationship between pretests and posttests. Thus, due to the violation of the linearity assumption, the researcher conducted separate MANOVAs to compare group means on pretests and posttests of pragmatic competence. It's important to note that MANOVA assumes normality of data, homogeneity of variances among groups, and homogeneity of covariance matrices. The normality of data was confirmed based on the skewness and kurtosis indices presented in Table 4.2. The skewness and kurtosis indices fell within the proposed ranges of ±2 by Bachman (2005), Bae & Bachman (2010), and George & Mallery (2019). Although Zhu et al. (2019) suggested a criterion of ±3, Watkins (2021) recommended a stricter criterion of less than ±2 for skewness and ±7 for kurtosis indices. The adherence to statistical assumptions and the methodology followed in this study provide a robust foundation for understanding the impact of authentic task-based assessment on the pragmatic

competence development of EAP students. Further research could explore alternative statistical approaches to address violations of assumptions and enhance the validity of the findings.

The study emphasized the importance of meeting assumptions such as normality, homogeneity of variances, and homogeneity of covariance matrices for performing MANOVA accurately. The results from Levene's test and Box's test indicated that these assumptions were met in the study, ensuring the validity of the analysis.

DISCUSSION

The present study aimed to investigate the development of EAP students' pragmatic competence through authentic task-based assessment. The subsequent MANOVA results revealed that there were no significant differences in the overall means of the six components of pragmatic competence among the four groups. This suggests that the groups had similar levels of competence across instrumental, regulatory, personal, interactional, wants explanation, and share knowledge & imagination aspects prior to treatment administration. The weak effect sizes observed further supported the conclusion of homogeneity among the groups regarding their pretest competencies. The findings presented in the text underscore the importance of ensuring the homogeneity of variances and covariance matrices when conducting statistical analyses like MANOVA. Despite no significant differences found in the groups' means on pretests of pragmatic competence, the detailed analysis of each component revealed that the groups exhibited homogeneity in instrumental, regulatory, personal, interactional, and wants explanation competencies. These results are crucial in establishing a baseline understanding of the groups' initial competence levels, which will be essential for evaluating the impact of any treatments or interventions in subsequent analyses. Overall, the meticulous examination of homogeneity across various dimensions of pragmatic competence provides valuable insights into the

readiness of the groups for the forthcoming interventions and highlights the robustness of the statistical procedures employed in the study.

The study focused on assessing the pragmatic competence development of EAP students through authentic task-based assessment by exploring the null-hypothesis that such development is not significant. Utilizing Multivariate ANOVA, the examination involved comparing posttest means across four groups to scrutinize this hypothesis. While the assumption of homogeneity of variances was largely upheld across various aspects of pragmatic competence, it was noted to be violated specifically in instrumental competence.

The discussion further extended to examining the covariance matrices and revealed the maintenance of homogeneity, as indicated by the results from Box's test, reinforcing the validity of the analysis. Subsequent analyses unveiled substantial differences among the means of the groups regarding various components of pragmatic competence posttests, ultimately leading to the rejection of the null-hypothesis. Notably, the role play group emerged with the highest mean scores across all components, closely followed by the dictionary and study skills groups, while the control group recorded the lowest mean scores. Detailed comparisons through post-hoc tests highlighted significant disparities in instrumental competence among the groups, with the role play group exhibiting notably superior performance compared to the dictionary, study skills, and control groups. These findings underscored the benefits of task-based assessment in enhancing pragmatic competence and shed light on the effectiveness of different instructional approaches in academic settings.

The study questioned whether authentic task-based assessment could significantly improve pragmatic competence. The research utilized multivariate ANOVA to compare different groups' means on posttests of pragmatic competence in order to investigate this null-hypothesis. It evaluated aspects like homogeneity of variances and covariance matrices, which are crucial in statistical analysis for drawing valid conclusions. The results indicated that while the assumption of homogeneity of variances was

generally met for most components, it was violated when it came to instrumental competence. Hence, adjustments were made in the statistical analysis to maintain accuracy. The text also delves into the importance of conforming to specific alpha levels when assumptions are not fully met, as this ensures the reliability of the findings.

Further analysis through MANOVA revealed significant differences between the overall means of the six components of pragmatic competence across the studied groups. Notably, the null-hypothesis was rejected, indicating that pragmatic competence could indeed be developed through the task-based assessment. The findings showed that the role play group had the highest mean across all components, followed by the dictionary and study skills groups, while the control group scored the lowest. In-depth comparisons of the groups' mean on posttests of pragmatic competence were provided, highlighting the effectiveness of role play in enhancing instrumental competence compared to other instructional methods.

CONCLUSION

Despite encountering a violation of the linearity assumption necessary for Multivariate Analysis of Covariances (MANCOVA), the researcher adjusted their approach by conducting separate MANOVAs to compare group means on pretests and posttests of pragmatic competence. The confirmation of data normality, homogeneity of variances among groups, and homogeneity of covariance matrices further solidified the study's foundation. While the findings contribute to understanding the impact of authentic task-based assessment on EAP students' pragmatic competence development, future research could explore alternative statistical approaches to address assumption violations and enhance the validity of results. This study underscores the significance of methodological rigor in evaluating pragmatic competence in educational contexts. It can be concluded that authentic task-based assessment is beneficial for developing the pragmatic competence of EAP students. The adherence to statistical assumptions and methodology in

the study provides a solid foundation for understanding the impact of such assessments. Further research could explore alternative statistical approaches to address violations of assumptions and improve the validity of the findings.

The study emphasizes the importance of meeting statistical assumptions for accurate analysis. Despite no significant differences found in the groups' initial competence levels, the detailed component analysis revealed homogeneity, laying a crucial baseline for evaluating future interventions. The meticulous examination of homogeneity provides valuable insights for assessing group readiness and underscores the robustness of the statistical procedures employed. The study on authentic task-based assessment of EAP students' pragmatic competence highlights the importance of meeting statistical assumptions for accurate analysis. While the groups showed similar levels of competence across various aspects before treatment, the meticulous examination of homogeneity in variances and covariance matrices underscored the study's robustness. Despite no significant differences in the overall means, the detailed analysis revealed insights into the groups' initial competency levels. These findings provide a crucial baseline for evaluating the impact of future interventions. Ensuring homogeneity in statistical analyses like MANOVA is vital for drawing accurate conclusions. The weak effect sizes observed further supported the conclusion of homogeneity among the groups. Therefore, the study emphasizes the significance of rigorous statistical procedures in assessing students' pragmatic competence effectively and setting the stage for future evaluations and interventions.

The study yields significant results, ultimately rejecting the null hypothesis that such development is not significant. Utilizing Multivariate ANOVA, the study found that the role play group excelled, followed closely by the dictionary and study skills groups, while the control group performed less well. Notably, the analysis showed that differences among the groups were particularly pronounced in instrumental competence. The findings emphasized the effectiveness of task-based assessment in enhancing pragmatic competence in academic settings and highlighted the varying

impacts of instructional approaches. The study's approach offers valuable insights into evaluating and improving EAP students' pragmatic skills, indicating the importance of incorporating authentic tasks into assessments to better understand and enhance students' language proficiency. The study on assessing EAP students' pragmatic competence showed that task-based assessment significantly improves students' skills, with the role play group outperforming other groups. This highlights the effectiveness of interactive instructional methods in developing pragmatic competence in academic settings.

The study also reveals intriguing insights into the effectiveness of such assessments. The text underlines the importance of adhering to specific alpha levels when assumptions are not fully met to guarantee the credibility of the findings. This study showcases the value of authentic task-based assessment in enhancing pragmatic competence among EAP students, while also emphasizing the need for meticulous analysis to yield reliable results.

The study underscores the effectiveness of task-based assessment, particularly role play, in enhancing EAP students' pragmatic competence. The rejection of the null-hypothesis suggests that pragmatic competence is indeed fostered through such assessments. The role play group demonstrated the highest mean proficiency, outperforming the dictionary and study skills groups, while the control group lagged behind. Comparative analysis of the groups' posttest results highlighted role play as a potent tool in developing instrumental competence. The significant differences identified between the groups emphasize the influence of varied teaching methods on pragmatic competence growth among EAP students. These findings advocate for the integration of authentic task-based assessments, especially role play, into EAP curricula to bolster students' pragmatic skills effectively.

The findings from this study clearly indicate that authentic task-based assessments, particularly role play, significantly enhance the development of pragmatic competence among EAP students. The role play group, in particular, outperformed other groups, demonstrating superior instrumental competence, a key aspect of pragmatic ability. This suggests that authentic,

interactive tasks provide students with valuable opportunities to practice language in real-world contexts, reinforcing their ability to use language appropriately in academic settings. By simulating real-life communication, task-based assessment effectively helps students build confidence and proficiency in navigating various social and academic interactions, making it a highly beneficial tool for fostering pragmatic competence.

The implications of these findings are particularly relevant for educational settings that aim to equip students with the practical language skills needed for academic and professional success. Incorporating authentic task-based assessments, such as role plays and simulations, into EAP curricula can significantly enhance students' ability to apply language meaningfully in context. This method of assessment not only improves linguistic competence but also addresses the nuances of social interaction, such as cultural sensitivity and contextual appropriateness, which are critical for effective communication. Educators should consider adopting such approaches to ensure a holistic development of language skills that go beyond traditional grammar and vocabulary assessments.

However, the study also faced some limitations, particularly related to the violation of the linearity assumption required for MANCOVA, which led the researchers to use MANOVA instead. While this adjustment allowed for robust analysis, future research should explore alternative statistical methods to overcome such limitations. Additionally, expanding the participant sample to include a more diverse range of learners or incorporating longitudinal studies could provide further insights into the long-term impact of task-based assessment on pragmatic competence development. Exploring the role of different instructional methods and task types across various educational contexts could also yield more comprehensive results.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Mohammadreza Afshari   ⓘD   http://orcid.org/0009-0004-4550-7929

Seyed Reza Beh-Afarin   ⓘD   http://orcid.org/0000-0001-8666-9458

Jahanbakhsh Nikoopour   ⓘD   http://orcid.org/0000-0002-9532-0108

References

Ahmadian, M. J., & García Mayo, M. P. (2019). Current trends and new developments in task-based language teaching. ELT Journal, Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bae, J., & Bachman, L. F. (2010). An investigation of four writing traits and two tasks across two languages. *Language Testing*, *27*(2), 213-234. https://doi.org/10.1177/0265532209349470

Bardovi-Harlig, K., & Bastos, M.-T. (2011). Proficiency, length of stay, and intensity of interaction in L2 pragmatics. Intercultural Pragmatics, 8, 347–384.

Clennell, C. (1999). Promoting pragmatic awareness and spoken discourse skills with EAP classes. ELT Journal, 53(2), 83–91.

Colpin, M., & Gysen, S. (2006). Developing and introducing task-based language tests. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 151–174). Cambridge University Press.

Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge University Press.

Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, *4*(3), 193-220. https://doi.org/10.1177/136216880000400302

Ellis, R. (2003). Task-based instruction. Language Teaching, 36(1), 1-14.

Ellis, R., & Shintani, N. (2021). Task-based language teaching (2nd ed.). Cambridge University Press.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Field, A. (2018). *Discovering statistics using IBM SPSS, Statistics for statistics*. (5th ed.). SAGE Publications.

Fulcher, G. (2013). *Practical language testing*. Routledge.

Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, *28*(1), 5-29. https://doi.org/10.1177/0265532209359514

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualizations, operationalization, and outstanding questions. *Language Assessment Quarterly*, *15*(3), 219-236. https://doi.org/10.1080/15434303.2018.1453816

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

Grabowski, K. C. (2016). *Assessing pragmatic competence.* Chicago Press.

Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. Psychology Press.

Huang, N. (2022). Revisiting L2 pragmatic competence through implicit vs. explicit instructional framework. Frontiers in Psychology, 13, 987729.

Kasper, G., & Ross, S. J. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics*, *39*(11), 2045-2070. https://doi.org/10.1016/j.pragma.2007.07.011

Ke, C. (2006). A model of formative task-based language assessment for Chinese as a foreign language. *Language Assessment Quarterly: An International Journal*, *3*(2), 207-227. https://doi.org/10.1207/s15434311laq0302_6

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23. https://doi.org/10.3102/0013189X023002013

Mukhopadhyay, L., & Sudharshana, N. P. (Eds.). (2021). Task-based language teaching and assessment: Contemporary reflections from across the world. Springer. https://doi.org/10.1007/978-3-319-73423-7

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. Doughty (Eds.), *The handbook of language teaching*, 578-594. https://doi.org/10.1002/9781444315783.ch30

Nunan, D. (2004). Task-based language teaching. Cambridge University Press.

Pallant, J. F., Haines, H. M., Green, P., Toohill, J., Gamble, J., Creedy, D. K., & Fenwick, J. (2016). Assessment of the dimensionality of the Wijma delivery

expectancy/experience questionnaire using factor analysis and Rasch analysis. *BMC Pregnancy and Childbirth*, *16*, 1-11. https://doi.org/10.1186/s12884-016-1157-8

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, *23*(2), 229-256. https://doi.org/10.1191/0265532206lt329oa

Shehadeh, A. (2005). Task-based language learning and teaching: Theories and applications. In C. Edwards & J. R. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 13-30). Palgrave Macmillan.

Shohamy, E., Or, I. G., & May, S. (Eds.). (2017). *Language testing and assessment* (pp. 441-454). Springer.

Skehan, P., & Luo, S. (2020). Developing a task-based approach to assessment in an Asian context. *System*, *90*, 102223. https://doi.org/10.1016/j.system.2020.102223

Sudharshana, N. P., & Malicka, A. (2023). Task-based language assessment: Components, development, and implementation. In R. Ellis & N. P. Sudharshana (Eds.), Task-based language teaching: Early days, now and into the future (pp. 207-215). Springer.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Taguchi, N. (2019). Second language acquisition and pragmatics: An overview. Routledge.

Taguchi, N., & Kim, Y. (Eds.). (2018). Task-based approaches to teaching and assessing pragmatics. John Benjamins Publishing.

Taguchi, N., & Roever, C. (2017). Second Language Pragmatics. Oxford University Press.

Timpe-Laughlin, V. (2018). Pragmatics in task-based language assessment: Opportunities and challenges. In N. Taguchi & Y. Kim (Eds.), Task-based approaches to teaching and assessing pragmatics (pp. 287-304). John Benjamins Publishing. https://doi.org/10.1075/tblt.10.12tim

Tsutagawa, F. S. (2013). Pragmatic knowledge and ability in the applied linguistics and second language assessment literature: A review. *Working Papers in TESOL & Applied Linguistics*, *13*(2), 1-20. https://doi.org/10.7916/salt.v13i2.1324

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.

Van den Branden, K., Van Gorp, K., & Verhelst, M. (2007). Task-based language teaching in practice: Research and pedagogical implications. Language Teaching, Cambridge University Press.

Van den Branden, K., Van Gorp, K., & Verhelst, M. (Eds.). (2009). *Tasks in action: Task-based language education from a classroom-based perspective*. Cambridge Scholars Publishing.

Walters, F. S. (2009). A conversation analysis-informed test of L2 aural pragmatic comprehension. *TESOL Quarterly, 43*(1), 29-54. https://doi.org/10.1002/j.1545-7249.2009.tb00226.x

Watkins, M. W. (2021). *A step-by-step guide to exploratory factor analysis with SPSS*. Routledge.

Weaver, W., & Gere, J. M. (2012). *Matrix analysis framed structures*. Springer Science & Business Media.

Willis, J. (1996). *A flexible framework for task-based learning.* Longman.

Youn, S. J., & Chen, S. (2021). Creating valid and reliable scoring rubrics for performance-based classroom assessment. *European Journal of Applied Linguistics & TEFL*, *10*(1).

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, *382*(8), 727-733. https://doi:10.1056/NEJMoa2001017

Zhu, X., Raquel, M., & Aryadoust, V. (2019). Structural equation modeling to predict performance in English proficiency tests. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment volume II* (pp. 101-126). Routledge.