



SCOPE

Journal of English Language Teaching

| p-ISSN 2541-0326 | e-ISSN 2541-0334 |

<https://journal.lppmunindra.ac.id/index.php/SCOPE/>



Research Article

Developing an Online Test Battery for Testing EFL Pragmatic Competence: What Can It Tell Us?

Fadhlur Rahman¹, Ella Yuzar², Wanying Zhou³

¹ Institut Agama Islam Negeri (IAIN) Lhokseumawe, Indonesia

² Institut Agama Islam Negeri (IAIN) Langsa, Indonesia

³ The University of Melbourne, Australia

KEYWORDS

Pragmatic Competence;
Online Test Battery;
EFL;
Study Abroad;
Language Proficiency

ABSTRACT

This study aims to develop a practical and reliable test to assess second language pragmatic knowledge as well as to investigate the effect of study abroad experience and proficiency level on EFL learners' pragmatic competence. A total 59 participants, consisting of 34 females and 29 males, were categorized into groups based on their exposure to the target language and proficiency. The results show that the reliability of the test reaches $\alpha=.798$ which means it is acceptable for a language test. Further development of test items is required to achieve more adequate test reliability. The findings also suggest that learners' pragmatic knowledge can be influenced by both study-abroad experience and linguistic competence. These factors may contribute to learners' ability to appropriately respond to pragmatic utterances from the test. Thus, language learners should focus on increasing their linguistic proficiency to enhance their understanding and use of pragmatic language. This study underscores the importance of designing effective language pragmatic tests that incorporate a variety of pragmatics components and comprehensive instruments to achieve the desired test reliability and validity.

CORRESPONDING AUTHOR(S):

E-mail: fadhlur.rahman@iainlhokseumawe.ac.id

INTRODUCTION

Pragmatics is a crucial field of study in modern linguistics because it allows us to understand how meaning is conveyed beyond the literal meaning of words (Yule, 2022). In communication, the meaning of language often goes beyond the literal definitions of the words used. Pragmatics, therefore, explores how context, speaker, and listener contribute to meaning-making in language use (Roever, 2021). It also investigates how language is used to achieve various social functions, such as making

requests, giving directions, expressing emotions, and building relationships (Taguchi, 2022).

The concept of interlanguage was first introduced by Selinker, who suggested that second language learners do not simply replace "the rules of their first language with the rules of the target language" (Ward, 2016). Instead, they create a dynamic system of their own, which he called 'interlanguage.' Interlanguage is characterized by a continuum of development between the first language and the target language, and it is shaped by the learners' prior linguistic experience and their contacted to the intended

language. Interlanguage pragmatics builds on the concept of interlanguage by investigating how second language learners acquire pragmatic competence, which is the ability to use the language appropriately in sociocultural contexts (Kim, 2019). Pragmatic competence includes both socio-pragmatic competence, which involves the knowledge of social conventions and norms in language use, and pragma-linguistic competence, which relates to the capability of conveying meaning through the use of linguistic resources (Roever, 2021; Ishihara & Cohen, 2022). Lately, the concept of pragmatic competence has shifted from an individualistic view that involves grammatical and discourse knowledge to a more interaction based. The later concerns on speaker's understanding of language construct and follows with the ability in using them in culturally distinct social interactions.

The failure to understand pragma-linguistics may result in the use of inappropriate linguistic forms, whereas the failure in comprehending sociopragmatic competence may cause more serious communication issues (Taguchi, Hirschi, & Kang, 2022), such as the violation of specific or particular sociocultural rules or the communication context between interlocutors. The importance of studying interlanguage pragmatics lies in the fact that language learners who lack pragmatic competence may encounter significant difficulties in social interaction (Taguchi & Yamaguchi, 2020). For example, learners who do not understand socio-pragmatic competence may inadvertently use language inappropriately, causing offense or embarrassment (Umale, 2011). They may also misunderstand the intended meaning of utterances and respond inappropriately, leading to further communication breakdowns. In some cases, failure to understand socio-pragmatic competence can result in serious social consequences, such as being excluded from social groups or causing offense to authority figures. Thus, the study of interlanguage pragmatics is important for second language learners to develop their pragmatic knowledge and language use, so as to be able to converse proficiently and competently in various social circumstance (cultural practices and norms).

Pragmatic competence is also essential for effective communication. Learners who lack this competence may struggle to convey the intended meaning of their message or may use language in ways that are grammatically correct but not appropriate for the social context (Kasper & Rose, 2002; Bardovi-Harlig, 2013). For example, they may use overly formal language in informal settings or may use slang or colloquial language in formal settings. This can result in miscommunications or make the speaker appear uncomfortable or impolite.

The current study focuses on two components of pragmatic competence, namely speech act and implicature. According to the seminal work of J.L Austin, the concept of speech act refers to the utterance that assigns a function in communication and affects the real world (Taguchi & Roever, 2017). J.L Austin was the originator of the idea of speech act, who divided it into three components: locutionary act, illocution, and perlocution. The paper analyses three types of speech act, which are apology, request, and refusal.

Likewise, the conversational implicature, which was initially investigated by H.P. Grice (1913-1988) has been one of the principal subjects of pragmatics. It refers to the implicit meaning conveyed beyond the literal meaning of the speaker's words (Petrey, 2016). In other words, speakers might not use the sentences to carry the exact meaning from how the sentences are sounded and structured. The expressions cannot be translated literally as they have different meanings according to the context of where and when the utterances are being used. Grice was the first to systematically study the cases in which speakers imply meanings other than what are being said during a discourse. He concluded that it is not possible to fully understand what speakers mean without processing what is uttered and what is implicated. For this, it is imperative to investigate how second language learners perform types of speech acts and implicatures in English as a second language and how their performance is affected by their proficiency level and cultural background. By examining these issues, the study aims to contribute to the development of interlanguage pragmatics, which is crucial for second language learners to communicate effectively and appropriately in various social contexts.

The relationship between pragmatics knowledge and English proficiency level, as well as language learners' exposure to a target language obtained from study abroad (SA) experience, has been extensively researched (Taguchi, 2009; Köylü, 2018; Chen & Lin, 2021). Studies have shown that exposure to a target language environment can improve learners' understanding of conversational implicature and pragmatic appropriateness. Additionally, learners become more pragmatically competent with longer stays in an ESL country. Recent research suggests that studying abroad has an impact on how learners' express requests in email communication, especially regarding the imposition level of the written request. The studies showed that the amount of time spent in English speaking countries has benefited English learners in terms of using a more academically and pragmatically appropriate language toward their interlocutors (Soler & Hernández, 2017; Kim & Lantolf, 2018; Vidal & Shively, 2019).

Some recent studies have shown mixed results on the relationship between language proficiency level, exposure to the target language environment, and pragmatic knowledge (Rafieyan, 2018; Ren, 2019; Trebits, 2021). While some studies found that exposure to an ESL environment improves learners' pragmatic competence, others argue that proficiency level has more impact on pragmatic knowledge. Yet, grammatical proficiency does not necessarily translate to high pragmatic competence (Taguchi, 2018; Beltrán-Planques & Querol-Julián, 2018).

Regarding the influence of proficiency level, Tabatabaei and Tabatabaei (2019) state that, in contrast to exposure during a residency in an ESL country; proficiency level does not show a significant difference on pragmatic competence in the time span of utterance. In a similar vein, Bardovi-Harlig (2019) discovered that learners with high proficiency levels, particularly in grammatical competence, also do not have a high degree of pragmatic competence. Yet, Roever (2006) discovered a contradicting conclusion by employing a multi-construct pragmatics. He observed that exposure to an ESL context increased learners' grasp of routine. This was supported by studies suggesting that exposure to a language may improve understanding of implicature (Rose, 2005; Garcia, 2004; Jeon & Kaya, 2006). However, Roever (2006) also mentioned that this is not completely true. He claims that a learner's level of language proficiency has an important role in understanding pragmatic aspects of language, such as implicature and speech acts. Merely living in a country where the target language is spoken is not enough to guarantee improvement in a learner's ability to understand and use these aspects of language. Instead, the learner's level of proficiency and the amount of interaction they have in the language are the key factors that influence individual's ability to recognize and produce pragmatic utterances.

Thus, the main aim of this research, therefore, is to create and validate an assessment instrument that measures the pragmatic knowledge of ESL/EFL learners regarding speech acts and implicatures. It seeks to develop a valid and reliable instrument that can assess EFL learners' pragmatic competence accurately. The instrument will focus on both receptive and productive aspects of speech acts and will cover partial sociolinguistic and pragmalinguistic knowledge. Pragmalinguistic knowledge refers to the capacity to understand the social connotations of utterances and to carry out the speech intentions. Meanwhile, sociopragmatic aspects will be examined through appropriateness judgment items as they are considered an efficient and a common method for assessing students' socio-pragmatic competence (Roever, 2005; Roever, Fraser & Elder, 2014). The study also

purposes to analyze whether exposure of target language during study abroad and language proficiency will impact on learners' socio-pragmatic and pragmatic competence, specifically their ability in recognizing and interpreting conversational implicatures, as well as their ability to produce appropriate speech acts of request, apology, and refusal.

Research Questions:

1. How do the general characteristics of the test impact its reliability in assessing ESL/EFL learners' pragmatic knowledge of English?
2. Do the intensity of interaction with the target language environment during study abroad experiences and proficiency level affect learners' pragmatic competence?
3. To what extent do ESL/EFL learners' receptive and productive knowledge of speech acts align with each other?

METHOD

Participants

The current study involved 59 EFL learners, consisting of 34 female and 29 male students who voluntarily took the designed online test. The participants were divided into two groups based on their level of exposure to the environment where the target language is spoken. One group had studied abroad (SA) in English speaking countries for at least 6 months, while the other group had not had such an experience (NSA). The NSA group included 31 undergraduate students from a university in Indonesia, all of whom had Indonesian as their native language. Meanwhile, the SA group consisted of 28 graduate students or fresh graduates, with 19 learners having Chinese as their L1 and 9 speaking Indonesian as their native language. The participants self-reported their levels of English proficiency, which ranged from lower intermediate to proficient, based on the results they earned on high-stakes language proficiency tests such as the paper-based TOEFL (Paper-Based Test and Internet-Based Test), and the IELTS. Furthermore, the study involved the participation of nine individuals who were native speakers of Australian English. These individuals took part in the initial two sections of the test alongside the non-native speakers, and their answers were used as a benchmark for evaluating the responses of the non-native speakers. The purpose of this was to establish a standard against which the non-native speakers' performance could be compared and scored in the subsequent sections of the test.

Instrumental Design

Several item types were developed in accordance with both Roever (2005) and Roever, Fraser and Elder's (2014) web-

DOI: <http://dx.doi.org/10.30998/scope.v8i1.17418>

based evaluation battery of ESL pragmatics as well as the personal experience of the researchers. Since the majority of the test takers were located in other countries, the evaluation that was developed here was carried out in the form of an online test battery using Microsoft Form for the sake of convenience.

Adapted from Roever (2005), the test is divided into three sections, namely multiple-choice appropriateness judgement of speech actions, multiple-choice implicature, and speech acts DCT. Those three forms of test sections were chosen due to practicality demanded for a web-based test battery. There are six questions in each area with a total of eighteen questions on the test. In the result sections, the items are labelled with Situation 1 (S1) to Situation 18 (S18). The first two sections are made of multiple-choice questions, and the third section, which is a DCT, asks participants to type in their responses to scenarios that are displayed to them.

To maintain the authenticity of English language both grammatically and pragmatically, a number of English native speaker were invited to evaluate the writing of the questions. Since the native equivalents were recruited in Australia, the goal of the section under "Appropriateness Judgement" was to gain an understanding of the degree to which the test takers' knowledge of social norms is native-like in accordance with the context of Australia. It had six different items pertaining to the spoken acts of requesting, apologizing, refusing, and thanking. The test included a series of questions, with each one providing a scenario that depicts a conversation between two individuals in a specific context.

The second response in each of these conversations was expected to be very straightforward and completely appropriate for the given situation, with no attempt to be overly polite or courteous. This was done to assess the participants' ability to communicate effectively in different social contexts and to determine whether they were able to adjust their language use according to the situation at hand. It succeeded in all three of these objectives. Participants in the survey are to base their decisions on a Likert scale that ranges from very impolite to too polite.

The scenarios presented in the test involved a range of different communication contexts, such as interactions with a professor, conversations with friends, encounters with strangers on the street, and discussions with members of staff in a supermarket. This diverse range of situations was designed to evaluate the participants' ability to use the target language in a variety of settings and to adapt their language use to suit different social contexts.

In the Implications part of the test, there were six questions that tested candidates' knowledge of ironic expressions and

indirect discourse. Roever (2005) served as both an inspiration and a source of material for the items. A typical Pope Q item was modified with another popular expression that is conceptually identical, namely "is the sky blue?" The majority of the scenarios involved interactions between friends and roommates. Participants in the examination are required to select, from among four possible interpretations, the one that most accurately reflects the author's intended meaning of the given passage.

Six different dialogues with a turn-and-talk format made up the final item type of the Discourse Completion Task (DCT) for speech acts. This test was designed to evaluate an individual's capacity for productive knowledge of speech acts. The purpose of this particular section was to enable a comparison and contrast of the participants' ability to both understand and produce different types of speech acts. By assessing both their receptive and productive pragmatic knowledge in relation to speech acts, the study aimed to gain a more comprehensive understanding of the participants' language abilities in this area. This allowed for a more nuanced analysis of their pragmatic competence and provided insight into potential areas of strength and weakness.

A situation prompt was provided, just like in the previous portion, and so was an explanation of the relationship between the interlocutors. Participants in the test are given the instruction to finish the dialogue during the second turn. These items included two requests, two apologies, and two rejections of requests.

Procedure

Prior to the main study, a pilot test of the language pragmatic instrument was conducted with a group of 10 EFL learners who spoke Chinese. The pilot test was aimed to identify any issues or difficulties in the instrument, and to estimate the amount of time required to complete the test. On average, the pilot participants completed the test in approximately 30 minutes, which included the time taken to provide background information. Therefore, the actual study participants were informed that the test should be completed within 30 minutes. Feedback from the pilot participants was taken into consideration to modify or change a few items in the instrument. Additionally, feedback from native speakers was also considered, which highlighted the need for clarification of certain items in the instrument. Some of the feedback concern with the clarity of the context situation written prior to the conversation transcript. Several changes and additional explanation should be made to avoid confusion and unnatural responses. The following is the sample modification resulted from the feedbacks.

(Pilot version)

A stranger bump into Sam on the street.

Stranger: *I'm sorry.*

Sam: *I forgive you.*

Look at Sam's response, do you think it is

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

(Modified test item)

Sam bumped into an elderly lady on his way out of a cafe. He couldn't avoid it because the lady was blocking the way. The lady was a bit shocked

Elderly lady: *Oh my goodness!*

Sam: *Excuse me, but you were blocking the way.*

Look at Sam's response, do you think it is

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

The present study employed a two-part survey-based test to evaluate the pragmatic English knowledge of EFL learners. The objective of the first section was to collect participants' background information, including their native language, gender, the length of time they had spent in an English-speaking country, and a reported score on their most recent English language proficiency test. This data was used to provide a clearer picture of the participants' linguistic background and experience/exposure with the English language. By considering these factors, the study aimed to identify any potential differences in language abilities among the participants and to examine how these factors might impact their performance on the test.

The second section consisted of actual items, which were scored on a twelve-point scale with each item receiving two points. Based on the majority of responses from a group of nine Australian native speakers, key answers to questions requiring a judgement of appropriateness were determined.

The final section of the test, the DCT, was graded using a six-factor scale developed by Hudson, Detmer, and Brown's seminal work (Roever & Ellis, 2021). In this particular investigation, expressions and directness were not used as criteria, so the final section's total score was calculated by averaging the scores of four criteria. Due to the distance condition, test supervision was not possible, and participants were given thirty minutes to complete the

test. After the completion of each section, the scores were calculated and converted into percentages. These individual scores were then combined to reach a final total score for the entire test. This approach allowed the researchers to gain an overall understanding of the participants' performance on the test as a whole, as well as to identify any specific areas in which participants excelled or struggled.

Analysis

Scores collected from the test takers would then be inputted and analyzed using the latest version of IBM SPSS Statistics. Descriptive analyses were performed to get the average score and standard deviation for all three sections. Reliability analysis was also run to find out the consistency both within the whole test and among the items. Finally, in order to examine the effect of study abroad experience and second language proficiency on pragmatic competence, an independent t-test was calculated. Test takers were divided into advance and intermediate groups in terms of their language proficiency.

RESULTS AND DISCUSSION

How do the general characteristics of the test impact its reliability in assessing ESL/EFL learners' pragmatic knowledge of English?

From the statistical result of both the whole test and its sub-sections (Table 1), it was determined that the average score for the entire test was just over 60%. This indicates that, on average, the participants performed moderately well on the test. Although the average score for the entire test was slightly above 60%, when the individual sections of the test were examined, it was discovered that the appropriateness judgement section and the implicature section had scores below the overall average.

In contrast, the DCT (Discourse Completion Test) section had a notably higher average score than the rest of the test, and the scores were more tightly clustered, indicating that the participants performed particularly well in this section. This suggests that the participants may have had greater proficiency in their ability to complete discourse tasks compared to their ability to judge appropriateness and recognize implicatures. It is important to note that the different scoring means across the sections might have contributed to this difference.

The DCT section was scored using a rating scale rather than a binary system, meaning that there was no absolute "correct" answer and participants were not penalized with a score of zero. In contrast, the appropriateness judgement section and the implicature section were scored using a system in which points were awarded for correct answers

and no points were awarded for incorrect answers. This distinction is important because it suggests that the DCT section may have been designed to assess a broader range of language skills and abilities, while the other two sections were more focused on testing specific linguistic knowledge. Additionally, the scoring method used in the DCT section may have contributed to the higher scores in this section compared to the other two.

The results also showed that the appropriateness segment (mean=44.49) was significantly below average, given that it tests culture-based sociopragmatic knowledge and that all participants were EFL learners. This section may be more challenging than the other parts due to the cultural differences in sociopragmatic knowledge. In contrast, the DCT section had a higher mean score, which might suggest that it is easier for participants to perform well on tasks that require them to complete a specific discourse.

The present results how well EFL learners performed on two distinct sections of a language pragmatic test: the appropriate judgement section and the implicature section. This sought to investigate how accurately EFL students were able to identify and comprehend the nuances of language use, particularly in relation to sociocultural context in which the language is used.

Table 2 and 3 present the results of the analysis of these two sections. The data reveal a fairly small spread, suggesting that the majority of test takers found the items in these sections moderately challenging. A few items, however, displayed slightly skewed distribution, although these item types did not present any significant problems for the participants.

Table 1. Sub-section Test and Score

| Section | N | Mean | Std. Dev |
|----------------------------------|----|-------|----------|
| <i>DCT</i> | 59 | 88.04 | 7.27 |
| <i>Appropriateness Judgement</i> | 59 | 44.49 | 22.14 |
| <i>Implicature</i> | 59 | 55.38 | 27.24 |
| Total | 59 | 62.63 | 15.39 |

Table 2. Appropriate Judgement Section

| Item | N | Mean | Std. Dev |
|-----------|----|-------|----------|
| S2 | 59 | .695 | .8760 |
| S1 | 59 | .712 | .9658 |
| S5 | 59 | .864 | .7978 |
| S6 | 59 | .915 | 1.0050 |
| S3 | 59 | 1.000 | .8906 |
| S4 | 59 | 1.153 | .9968 |

In the appropriateness judgement section, the data suggest that test takers struggled more with certain items than with

others. Specifically, a request to a grocery staff (S4) was found to be the easiest item in this section, with the majority of participants scoring close to that of native speakers. The following is the complete item illustration of S4.

(S4) Kelly is doing grocery shopping at a supermarket. She found her favorite brand of shampoo and she wanted to know the price. A shop assistant comes by, and she asks,

Kelly: *Excuse me, I'm sorry to bother you. I would be really grateful if you could tell me how much this costs?*

Look at Kelly's utterance, do you think it is...

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

In contrast, a refusal (S2) was identified as the most difficult item, with relatively lower scores compared to other items (Mean .695).

Table 3. Implicature Section

| Item | N | Mean | Std. Dev |
|------------|----|-------|----------|
| S12 | 59 | .847 | .9968 |
| S9 | 59 | .915 | 1.0050 |
| S10 | 59 | .949 | 1.0073 |
| S7 | 59 | 1.051 | 1.0073 |
| S8 | 59 | 1.085 | 1.0050 |
| S11 | 59 | 1.797 | .6097 |

Regarding the implicature section, the data revealed that participants had more difficulty in interpreting sarcasm (S12) than indirect speech (S11). This result implies that EFL learners may have a greater challenge in recognizing and interpreting sarcastic utterances in comparison to indirect speech. These findings indicate that the appropriateness judgement and implicature sections of the pragmatic language test present unique challenges for EFL learners. The results suggest that EFL learners may benefit from targeted instruction and practice in these areas to improve their pragmatic competence in English.

Table 4. DCT Section

| Item | N | Mean | Std. Dev |
|------------|----|--------|----------|
| S14 | 59 | 3.8686 | .7633 |
| S17 | 59 | 4.3475 | .7865 |
| S18 | 59 | 4.3861 | .5696 |
| S15 | 59 | 4.5996 | .4299 |
| S13 | 59 | 4.6907 | .7167 |
| S16 | 59 | 4.7118 | .3011 |

In comparison to the other two sections of the test, the DCT section displayed a greater degree of variability. The

participants in the test generally demonstrated an ability to perform acceptable speech acts; however, the level of difficulty varied greatly across the various types of speech acts that were tested. Making a request to an authority (S16), on the other hand, was found to be relatively simple for those who took the test. The data suggests that performing a proper apology act (S14) to an authority might be the most difficult task for them.

The Reliability of the Test and Its Sections

The reliability of a test is a key measure of its quality and is important for ensuring that the test results are accurate and reliable. In this study, the section reliabilities were not found to be satisfactory, which suggests that the test may have some issues with internal consistency. However, despite this, the overall reliability of the test was acceptable, given the small size of the items and the participants.

Table 5. The Test Reliability in Each Section

| Section | N | Alpha |
|-----------------------|----|-------|
| DCT | 59 | .58 |
| Implicature | 59 | .59 |
| Appropriate Judgement | 59 | .33 |
| Total | 59 | .798 |

In order to further understand the issues with the reliability of the test, an item analysis was conducted. The results of the analysis are presented in Table 6.1, 6.2 and 7 which shows the Item Facility (IF) and Discrimination Index (DI) for each item. The IF is a measure of how easy or difficult the item is, while the DI measures how well the item discriminates between high and low scorers.

Table 6.1 Item Difficulty

| Item | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| IF | .36 | .42 | .61 | .57 | .60 | .45 | .52 | .53 | .46 |

Table 6.2 Item Difficulty

| S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| .36 | .42 | .61 | .57 | .60 | .45 | .52 | .53 | .46 |

Table 7. Item Facility (IF) and Discrimination Index (DI)

| Item | IF | DI |
|------|------|------|
| S1 | .720 | .085 |
| S2 | .698 | .285 |
| S3 | .713 | .136 |
| S4 | .684 | .401 |
| S5 | .699 | .267 |
| S6 | .703 | .246 |
| S7 | .673 | .490 |
| S8 | .662 | .576 |
| S9 | .691 | .343 |

| | | |
|-----|------|------|
| S10 | .700 | .272 |
| S11 | .706 | .178 |
| S12 | .698 | .290 |
| S13 | .713 | .103 |
| S14 | .695 | .314 |
| S15 | .700 | .304 |
| S16 | .700 | .379 |
| S17 | .694 | .321 |
| S18 | .691 | .412 |

The item analysis results of this study suggest that most of the items in the test have an appropriate level of difficulty falling between the desired range of .33 to .67. However, some items appeared to be too easy or too difficult for the test takers. Among these items, indirect speech implicature (S11) was found to be the easiest item, while apology acts (S14, S18) proved to be the most challenging for the participants.

Based on the results of the Discrimination Index (DI), four items (S1, S3, S11, and S13) did not effectively differentiate between strong and weak learners. The low DI scores suggest that these items might not be useful in measuring the participants' sociopragmatic knowledge. In addition, the reliability of the entire test would increase if these items were discarded.

The judgement of items S1 and S3 might be affected by the test takers' sociopragmatic knowledge in their first language, as their ratings showed a high level of diversity. The problem with S11 could be that it was too easy, with nearly 90% of test takers receiving a full score. This indicates that the item might not have been challenging enough to differentiate between strong and weak learners. Lastly, for S13, it appears that some weak learners were rated higher than some stronger learners.

These findings are consistent with the overall results of the test, which demonstrated that the test takers performed well in general but struggled with certain aspects of sociopragmatic knowledge, particularly those related to making apologies. The insights gained from the item analysis can be useful in further refining the test items and improving the reliability of the test. Equally important, these results highlight the importance of conducting item analyses in order to identify problematic items and improve the quality of the test. By addressing these issues, future versions of the test could potentially have higher reliability and validity, increasing confidence in the test results and providing more accurate assessments of the participants' language abilities.

Do the intensity of interaction with the target language environment during study abroad experiences and proficiency level affect learners' pragmatic competence?

An independent t-test was conducted to examine if there were any significant differences between the two groups in terms of their overall test scores and scores on each sub-section. The results indicated that the SA (Study Abroad) group significantly outperformed the NSA (No Study Abroad) group on all measures, with statistically significant differences between the two groups.

The current results imply that exposure to an L2 environment through study abroad could have a positive impact on EFL learners' pragmatic knowledge. The SA group might have been exposed to more authentic language input and opportunities to practice real-life communication in the target language, which could have contributed to their better performance on the test. These findings highlight the importance of incorporating study abroad opportunities in language learning programs to enhance learners' pragmatic competence. Further research should explore the long-term effects of study abroad on pragmatic development and investigate the specific features of study abroad programs that contribute to pragmatic development.

Table 8. A Comparison of the two Group (SA and NSA)

| | Exposure to L2 | N | Mean | Std. Dev |
|---------------------------|----------------|----|---------|----------|
| DCT | SA | 28 | 91.4770 | 6.0031 |
| | NSA | 31 | 84.9352 | 6.9823 |
| Implicature | SA | 28 | 65.4762 | 20.7516 |
| | NSA | 31 | 46.2366 | 29.4108 |
| Appropriateness Judgement | SA | 28 | 58.0357 | 18.4929 |
| | NSA | 31 | 32.2581 | 17.7093 |
| Total | SA | 28 | 71.6630 | 11.1173 |
| | NSA | 31 | 54.4767 | 14.2043 |

Table 9. A Comparison of Two English Proficiency Groups

| | Proficiency | N | Mean | Std. Dev |
|---------------------------|--------------|----|---------|----------|
| DCT | intermediate | 28 | 84.0131 | 6.6940 |
| | advanced | 31 | 91.6768 | 5.7482 |
| Implicature | intermediate | 28 | 41.6667 | 28.5089 |
| | advanced | 31 | 67.7419 | 19.2139 |
| Appropriateness Judgement | intermediate | 28 | 33.9286 | 19.5011 |
| | advanced | 31 | 54.0322 | 20.1695 |
| Total | intermediate | 28 | 53.2028 | 14.6959 |
| | advanced | 31 | 71.1503 | 10.2718 |

This study appeared to investigate the impact of learners' proficiency level on their pragmatic competence as measured by a newly developed test. To accomplish this, participants were divided into two groups based on their IELTS scores: intermediate and advanced.

An independent t-test was used to investigate the differences in test performance between the two proficiency groups. The advanced group outperformed the intermediate group in all three sections of the test, as well

as in total scores. Furthermore, the effect size was large, indicating a significant difference in pragmatic competence between the two groups.

The current study's findings are consistent with previous research indicating that higher proficiency levels are associated with better pragmatic performance. It is also consistent with the assumption that increased exposure to language input and opportunities to practise in a target language environment can lead to increased pragmatic competence.

The current findings support the importance of taking learners' proficiency levels into account when designing and implementing pragmatic assessments. Proficiency level also becomes an essential element to consider when developing appropriate pedagogical interventions to improve pragmatic competence.

To what extent do ESL/EFL learners' receptive and productive knowledge of speech acts align with each other?

Table 10. Correlations Among Sub-sections

| | Appr.J | Implicature | DCT |
|--------------------|--------|-------------|-----|
| Appr.J | - | - | - |
| Implicature | .491** | - | - |
| DCT | .318* | .386** | - |

The present study examined the correlation between the Appropriateness Judgement and Discourse Completion Test (DCT) sections in a pragmatic test. The analysis revealed a statistically significant correlation between the two sections, but with a modest effect size. The correlation coefficients indicate that only 10% of the DCT scores can be explained by the scores of the Appropriateness Judgement section. Although this result may seem disappointing, it is important to consider the possible reasons behind it.

One possible explanation for the modest effect size could be the small size of the population and the total item count. The study might have lacked the necessary statistical power to detect significant correlations between the two sections. Additionally, it is worth noting that the skills assessed in each section are distinct, which could have contributed to the low correlation coefficients. While the items on the DCT assessed productive skills, the items on the Appropriateness Judgement section assessed receptive skills. Hence, the two sections assessed different aspects of pragmatic competence, which might have affected the correlation results.

It is important to keep in mind that the present study compared only two sections of the pragmatic test. Future research could explore the correlations between different

sections of the same test or between different tests assessing pragmatic competence. Furthermore, it could be informative to investigate the relationship between the scores of different sub-sections within each section of the test. Such analyses could provide more insights into the factors affecting the overall scores of the pragmatic tests.

To begin with, the findings related to the internal structure of the test are presented. Subsequently, we will examine how the results differ based on the participants' study abroad experience, level of proficiency, and the relationship between their receptive and productive language skills. Later, we will discuss the potential causes of the problematic question items identified in the results.

Ensuring the reliability of a test is crucial in its construction, as it determines the degree of consistency in the test's internal structure (Roever, Higuchi, Sasaki, Yashima, & Nakamuro, 2023). Examining the reliability is a practical way to gauge the test homogeneity. It is unfavourable for a test if all sorts of test takers would result in identical scores. While the reliability of each section of the test was not satisfactory, the overall reliability of the entire test was .789, which is deemed acceptable for a low-stakes test. However, given the small number of items and participants, the reliability results might not be ideal. The test consists of only eighteen items that measure three test components, namely appropriateness judgment speech act, multiple-choice implicature, and DCT speech act. The test could potentially increase its reliability and be suitable for a high-stakes test by adding more items to the construct and administering it to a larger sample size.

In terms of the external criteria of the test, the results of the study indicate that there is a significant relationship between learners' exposure during their residence in the target language country and their performance on the pragmatic test. Specifically, participants who had studied abroad demonstrated higher scores on the test than those who had only learned English in their home country. This suggests that being immersed in a native speaker environment can help EFL learners develop their pragmatic knowledge, particularly in relation to speech acts and implicature. The findings suggest that exposure to the target language country may be a key factor in developing proficiency in these areas of language use. The results are in line with Taguchi (2018) and Vidal and Shively (2019), in which living in an English speaking will contribute to learners' ability in understanding second language pragmatics because it provides them with linguistic and sociocultural knowledge. The opportunities also enable learners to participate in target language communities where they can practice this knowledge input, increasing their second language pragmatic competence. These insights could have implications for language

teaching and learning, as they highlight the importance of providing opportunities for learners to engage with the target language in authentic contexts.

During the detailed analysis of the test results, it was found that a few participants who had spent more than 24 months in an English-speaking country provided non-native-like answers in the appropriateness judgment section of the test. This can be explained from a sociopragmatic perspective, where not all language learners have intensive interaction with native speakers or individuals from other cultures during their stay. Some L2 learners prefer to interact mainly with people from the same cultural background, which may limit their exposure to diverse pragmatic contexts, and as a result, their interlanguage pragmatics may not improve much.

Equally significant, the study revealed that there was a significant difference in the performance of intermediate and advanced learners on the pragmatic test. Specifically, advanced learners demonstrated greater pragmatic competence than their intermediate-level counterparts. It was also found that participants with higher levels of linguistic proficiency tended to score higher on the test than those with lower proficiency levels. These results suggest that proficiency in the target language is a key factor in developing pragmatic competence, as learners with a more advanced level of language knowledge were better able to recognize and respond appropriately to pragmatic language use in the test (Roever, 2006; Su, 2021).

The test design includes four question items that pose difficulties in distinguishing between learners of high and low proficiency. Specifically, items S1 and S3 in the appropriateness section may be attributed to sociopragmatic factors.

Item S1: You and your friends booked a study room at the library for two hours today, from 2 pm – 4 pm. When you arrived at the library, it was already 2.10 pm but you found a professor sitting by himself in the room. You knocked on the door and opened it.

You said: "*Excuse me, Sir. I am really sorry to bother you, but could you please kindly get out of this room now? We have made the booking from 2 pm to 4 pm.*"

Upon further analysis, it was found that the Chinese participants, regardless of their proficiency level, generally perceived the utterance in item 1 as rude. This perception can be explained by the sociopragmatic perspective (Taguchi, Chen, & Qin, 2023), as Chinese learners tend to use conventionally indirect request strategies when addressing someone of higher status. As a result, when the

word "get out" is used in the utterance, it is perceived as an impolite request. On the other hand, the Indonesian context has a more diverse range of cultural values, especially regarding interactions with individuals of higher status. This diversity of values has resulted in Indonesian EFL learners providing more varied responses to this question compared to Chinese-Indonesian ESL learners. This same reasoning can also be applied to problematic item no. 3.

Item S3: Lisa and John are friends, and they've known each other for a few years. Lisa is much older than John. She's just read a book and she's recommending it to John.

Lisa: "Have you read the book *Never Let Me Go*? It's a great book. I can lend it to you if you haven't".

John: "Really? Thank you, I haven't read that".

Lisa: "I'll have the book for you when I see you next time".

John: "Thanks, I am so honoured".

Look at the final utterance, do you think it is...

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

The test items that measure the learners' ability to provide a suitable response to a particular context are inadequate in discriminating their proficiency level. This is because learners with lower linguistic competence can provide brief yet appropriate responses, while high proficiency learners tend to give longer responses that include more linguistic devices, but may end up being irrelevant or excessive.

This study might have some limitations. First, it only covers a small part of pragmatics, specifically speech act and implicature, so it may not accurately represent the participants' overall pragmatic competence. Thus, caution should be exercised when interpreting the results (Longworth, 2019). Second, the written format of the test may have led to some participants' unclear understanding of the question as they were not able to hear the speakers' tone of voice. For example, the phrase "excuse me, you're blocking the way" could be interpreted as either rude or polite depending on the speaker's tone. This might also be the reason why native speakers did not agree on the same answers. Finally, the small number of native speakers involved in the study may not be sufficient to establish reliable scoring criteria.

CONCLUSION

The purpose of this study was to design a form of language pragmatic test and investigate whether the external criteria of the test, such as level of proficiency and previous study abroad experience in the country where the target language is spoken, have an effect on the ability of learners to understand and produce pragmatic knowledge. The findings indicate that the test has an adequate level of reliability; however, this level of reliability is not sufficient for high-stakes examinations.

To achieve the level of reliability that is desired, additional components need to be developed. A further important finding is that learners' pragmatic knowledge may improve as a result of both their linguistic competence and the language exposure they receive while studying abroad. This is due to the fact that students learning a second language require the linguistic tools in order to recognise pragmatic utterances, which allows them to comprehend what is being said and respond appropriately.

When taken together, these results imply that the design of a language pragmatic test ought to involve more pragmatics components using instruments that are more comprehensive. The next factor is linguistic competence, which, along with exposure to the target language via interaction, could support the acquisition of pragmatic knowledge. In conclusion, it is suggested that additional research be carried out, this time involving a greater number of both test items and participants.

REFERENCE

- Bardovi-Harlig, K. (2013). Developing L2 Pragmatics. *Language learning*, 63, 68-86.
- Bardovi-Harlig, K. (2019). Routines in L2 pragmatics research. In *The Routledge handbook of second language acquisition and pragmatics* (pp. 47-62). Routledge.
- Beltrán-Planques, V., & Querol-Julián, M. (2018). English language learners' spoken interaction: What a multimodal perspective reveals about pragmatic competence. *System*, 77, 80-90.
- Chen, Y. S., & Lin, M. F. (2021). Effects of peer collaboration on EFL learners' comprehension of conversational implicatures. *System*, 97, 102441.
- Garcia, P. (2004). Pragmatic comprehension of high and low level language learners. *Test-Ej*, 8(2), n2.
- Ishihara, N., & Cohen, A. D. (2022). *Teaching and Learning Pragmatics: Where Language and Culture Meet (Second)*. Routledge.
- Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development. *Synthesizing research on language learning and teaching*, 165-211.

- Kasper, G., & Rose, K. R. (2002). Pragmatic development in a second language. *Language learning*.
- Kim, J. (2019). Being polite or appropriate? The sociolinguistic aspect of politeness in speech acts and a pedagogical approach for pragmatic awareness and strategies focused on refusal. *MA TESOL Journal*, 5(1), 77-89.
- Kim, J., & Lantolf, J. P. (2018). Developing conceptual understanding of sarcasm in L2 English through explicit instruction. *Language Teaching Research*, 22(2), 208-229.
- Köylü, Y. (2018). Comprehension of conversational implicatures in L2 English. *Intercultural Pragmatics*, 15(3), 373-408.
- Longworth, G. (2019). Illocution and understanding. *Inquiry*, 18(1), 40-57. <https://doi.org/10.1080/0020174X.2019.1667869>
- Petrey, S. (2016). *Speech Acts and Literary Theory*. Routledge.
- Rafieyan, V. (2018). Knowledge of formulaic sequences as a predictor of language proficiency. *International Journal of Applied Linguistics and English Literature*, 7(2), 64-69.
- Ren, W. (2019). Pragmatic development of Chinese during study abroad: A cross-sectional study of learner requests. *Journal of Pragmatics*, 146, 137-149.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229-256.
- Roever, C. (2021). *Teaching and Testing Second Language Pragmatics and Interaction: A Practical Guide*. Routledge.
- Roever, C., & Ellis, R. (2021). 7 Testing of L2 pragmatics: The challenge of implicit knowledge. In *New Directions in Second Language Pragmatics* (pp. 142-155). De Gruyter Mouton.
- Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL Sociopragmatics: Development and Validation of a Web-Based Test Battery*. Germany: Peter Lang.
- Roever, C., Higuchi, Y., Sasaki, M., Yashima, T., & Nakamuro, M. (2023). Validating a test of L2 routine formulae to detect pragmatics learning in stay abroad. *Applied Pragmatics*, 5(1), 41-63.
- Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33(3), 385-399.
- Röver, C. (2005). *Testing ESL Pragmatics: Development and Validation of a Web-Based Assessment Battery*. Germany: Peter Lang.
- Soler, E. A., & Hernández, A. S. (2017). Learning pragmatic routines during study abroad: A focus on proficiency and type of routine. *Atlantis*, 191-210.
- Su, Y. (2021). Accepting invitations and offers in second language Chinese: Effect of proficiency on pragmatic competence in interaction. *Journal of Pragmatics*, 180, 131-149.
- Tabatabaei, S., & Tabatabaei, S. (2019). Language proficiency and appropriateness of using refusal speech acts by Iranian EFL learners. *Applied Linguistics Research Journal*, 4(1), 35-45.
- Taguchi, N. (2009). Corpus-informed assessment of comprehension of conversational implicatures in L2 English. *TESOL quarterly*, 43(4), 738-749.
- Taguchi, N. (2018). Contexts and pragmatics learning: Problems and opportunities of the study abroad research. *Language Teaching*, 51(1), 124-137.
- Taguchi, N. (2022). Teaching and learning Pragmatics. In *Handbook of practical second language teaching and learning* (pp. 499-512). Routledge.
- Taguchi, N., & Roever, C. (2017). *Second Language Pragmatics*. Oxford University Press.
- Taguchi, N., & Yamaguchi, S. (2020). Intercultural pragmatics in English as a lingua franca. In *Pragmatics pedagogy in English as an international language* (pp. 76-94). Routledge.
- Taguchi, N., Chen, Y., & Qin, Y. (2023). Metapragmatic knowledge and transfer of learning across speech acts. *Applied Linguistics*.
- Taguchi, N., Hirschi, K., & Kang, O. (2022). Longitudinal L2 development in the prosodic marking of pragmatic meaning: Prosodic changes in L2 speech acts and individual factors. *Studies in Second Language Acquisition*, 44(3), 843-858.
- Trebits, A. (2021). Degree of multilingualism, code-switching and intensity of target language contact predict pragma-linguistic awareness in an English as a foreign language context. *International journal of multilingualism*, 18(3), 475-490.
- Umale, J. (2011). Pragmatic failure in refusal strategies: British versus Omani interlocutors. *Arab World English Journal*, 2(1), 18-46.
- Vidal, C. P., & Shively, R. L. (2019). L2 pragmatic development in study abroad settings. In *The Routledge handbook of second language acquisition and pragmatics* (pp. 355-371). Routledge.
- Ward, G. L. (2016). *The Semantics and Pragmatics of Preposing*. Routledge.
- Yule, G. (2022). *The Study of Language*. Cambridge University Press. <https://doi.org/10.1017/9781009233446>

APPENDIX

Sample Items of the Test

SECTION 1: Multiple choice speech act

Instruction- In each item, a situation will be described along with the following conversation. You will be asked to

DOI: <http://dx.doi.org/10.30998/scope.v8i1.17418>

rate the certain utterances based on the given criteria. Choose one of the answers in the multiple choices!

S1. You and your friends booked a study room at the library for two hours today, from 2 pm – 4 pm. When you arrived at the library, it was already 2.10 pm but you found a professor sitting by himself in the room. You knocked on the door and opened it.

You said: "Excuse me, Sir. I am really sorry to bother you, but could you please kindly get out of this room now? We have made the booking from 2 pm to 4 pm."

Look at the utterance. Do you think it is

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

S2. Your lecturer announced that it was strongly recommended to work in pairs for the final project. After class, a friend came and invited you to work with him, but you had already agreed to work with another friend.

Friend: "Hi, have you got a topic you want to work on for the final? Do you want to do it together?"

You: "Thank you so much for asking me, I really appreciate it. It means a lot to me. But I'm truly sorry that I already have a partner".

Look at the response to the friend, do you think it is

- A. very impolite/very harsh
- B. not quite polite
- C. completely appropriate
- D. a little too polite
- E. far too polite

SECTION 2: Implicature

Instruction- In each item, a situation will be described along with the following conversations. In each situation you will be asked to choose the best answer from the four choices.

S7. Jack and Rosie are working on a group project at the library. The assignment is due today at midnight but they haven't finished yet.

Rosie: "I think we should send an email to Professor Baker. Do you think it will affect our score if we submit the paper tomorrow morning?"

Jack: "Is the sky blue? "

What is the point of Jack's question?

- A. The professor will not reduce their score if they submit tomorrow morning
- B. They have to submit the paper before midnight to avoid deduction
- C. He doesn't have an answer to Rosie's question. She should ask another one that he can answer.
- D. It will not cause a problem to submit the paper tomorrow morning as other students also do it

S8. Pete has just got his driving licence, and Jose was the first person who's been riding with him. Their friend Frank asked Jose how Pete's driving was. Jose answered,

Jose: "Well, it was an interesting experience. I'd say it's safe".

What does Jose probably mean?

- A. Pete is a careful driver, and he made him feel safe.
- B. Pete is not very skilled at driving, but he could still manage it.
- C. Pete drives slow and safe.
- D. Jose greatly enjoyed riding with Pete.

SECTION 3

Instruction- this section tests if you know how people express themselves in everyday English conversation. In each item, you will see a short description of a situation. Read the situation and write what the person would probably say. While there are several possibilities for each answer, you should keep in mind that your answers must make sense into the situations.

S13. Lizzy is at a party. Her friend offers her some cake, but she is not a sweet tooth. What would Lizzy probably say?

S14. You made an appointment with a professor to discuss your paper. Unfortunately, you forgot the meeting, and you didn't show up. You just remembered that when you saw the professor at a seminar the next day. You went to him. What would you probably say?