Ghancaran
Jurnal Pendidikan Bahasa dan Sastra Indonesia
P-ISSN 2714-8955 | E-ISSN 2715-9132
https://ejournal.uinmadura.ac.id/index.php/ghancaran/index
https://doi.org/10.19105/ghancaran.v20i2.26078

OPEN ACCESS

RESEARCH ARTICLE

Refusal DCT with content-validity evidence for English-L1 BIPA learners

Raden Yusuf Sidiq Budiawan[1]*, I Dewa Putu Wijana[2], Wira Kurniawati[3]

[1,2,3]Doctoral Program in Humanities, Universitas Gadjah Mada Yogyakarta, Yogyakarta, Indonesia
[1]Universitas PGRI Semarang, Semarang, Indonesia

***Corresponding Author**
r.yusuf.s.b@upgris.ac.id

ABSTRACT
Discourse Completion Tasks (DCTs) are widely used to study refusals, but adaptations of Beebe et al.'s framework are seldom accompanied by transparent validation in the studies we mapped, and we found no documented DCT for English-L1 learners of Indonesian (BIPA). An exploratory mapping of 92 refusal studies found that most adapted Beebe et al.'s framework while only a small minority reported explicit content validation. This study developed a 12-situation refusal DCT, built from four eliciting acts and three levels of relative power, and gathered preliminary evidence of its content appropriateness, field usability, and translation equivalence through readability trials, expert review, field piloting, and blind back-translation. Two experts rated 48 situation-aspect rating units, giving S-CVI/Ave = 0.98 and S-CVI/UA = 0.96, yet two units received the lowest score possible with two raters (I-CVI = 0.50). Piloting with B1 learners flagged one wording problem, and back-translation found six discrepancies, five of them in situations that had earned full expert agreement. All eight revised situations later received full scores. The study contributes a documented, layered procedure for adapting a DCT, with CVI reported for rating units, and an instrument for BIPA pragmatics research. The evidence is preliminary and limited to a two-expert panel and B1 learners.

Keywords: discourse completion tasks; content validity; refusal speech acts; BIPA; interlanguage pragmatics.

ARTICLE HISTORY
Received: September 30, 2026
Accepted: October 28, 2026
Published: November 15, 2026

LICENSE
Copyright © 2026 The Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.
https://creativecommons.org/licenses/by/4.0/.

How to Cite This Article
Budiawan, R. Y. S., Wijana, I. D. P., & Kurniawati, W. (2026). Refusal DCT with content-validity evidence for English-L1 BIPA learners. Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia, Special Edition Lalongét VII, 817-834. https://doi.org/ghancaran/v20i2.26078

INTRODUCTION

Refusing is a face-threatening act for both speaker and listener, so most languages have developed their own ways of softening it. For learners of *Bahasa Indonesia bagi Penutur Asing* (BIPA), these ways do not come automatically with vocabulary and grammar. A learner can be grammatically fluent and still sound blunt, stiff, or even rude when declining a thesis supervisor's request, because the politeness norms of the first language differ from those of Indonesian. Pragmatic learning is also often left to chance. Siddiqa and Whyte (2025), studying English classrooms in French secondary schools, found little pragmatic input in the textbooks and in classroom talk, which is one reason to want good tools for finding out what learners can actually do. Meanwhile BIPA keeps growing: Indonesian language policy treats BIPA teaching as a way to spread the language abroad (Susanto et al., 2024), and research output is rising (Syafruddin et al., 2025).

Research on learners of Indonesian has so far looked mostly at form. Kholiq et al. (2024) traced the roles of L1 and L2 in phonology and vocabulary, and Mahriyuni et al. (2024) classified grammatical, syntactic, and lexical errors in the writing of beginning French-speaking learners. Work on culture has looked at materials and teachers. Mokoginta and Arafah (2022) describe cultural schemas of relationship-building, assertiveness, and communication style in negotiation in Indonesian textbooks, (Widianto and Pusse (2025) found an uneven representation of local, national, and global culture in an official textbook, and Asteria et al. (2023) reported that BIPA teachers mostly treat culture as general knowledge. Pragmatic studies of learners are thin. Primantari and Wijana (2017) looked at requests by Korean-L1 learners, Kurniawati and Ginting (2026) at requests, and Kurniawati (2026) at how Mandarin-L1 learners acquire Indonesian politeness. Refusals are not among them.

The Discourse Completion Task (DCT) entered speech-act research through Blum-Kulka (1982), who used it with learners of Hebrew. For refusals, the framework cited most often is Beebe et al. (1990): 12 situations built by crossing four eliciting acts (requests, invitations, offers, and suggestions) with three levels of relative power, drawing on the variables of Brown and Levinson (1987). The eliciting acts start the exchange, and refusal is the target act in every situation. The grid is still being adapted, for Indonesian EFL students (Dwiana et al., 2021), Hungarian learners of English (Vásári, 2022), and Japanese learners of Spanish (García, 2024).

The instrument has known limits. Golato (2003), comparing DCT data with recordings of natural talk on compliment responses, showed that DCT answers mostly reveal what

participants think is appropriate to say, not what they do. Her data concerned compliments, but the point may extend to other speech acts. Kasper and Dahl (1991) found DCT responses shorter and less varied than spoken data, and Kasper (2000) concludes that these limits call for restricting validity claims to pragmalinguistic and sociopragmatic knowledge, not for dropping the DCT.

To see how this plays out in practice, we mapped 92 articles on refusals that used a DCT (an exploratory mapping, not a systematic review; see Method and Supplementary Material 1). Three patterns stood out. Most of the studies adapted Beebe et al.'s (1990) framework. A smaller group acknowledged the DCT's limits, some of them citing Golato's critique. Only a small minority reported an explicit content-validation procedure under a strict coding definition, and a Content Validity Index was not consistently reported even among them. Adaptations of Beebe et al.'s framework to new settings are therefore often used without a demonstration that the adaptation holds. Labben (2022) names the same gap when she argues that a DCT should be treated as a language test with a careful development process. Because the mapping was exploratory and coded by one person, it shows a general pattern and should not be read as a count of the literature.

The gap is wider for Indonesian. Few of the mapped studies had Indonesian as a language of refusal. Gustini (2018) compared Indonesian and Japanese native speakers in workplace situations, and Eviliana (2019) compared the Indonesian and English refusals of Indonesian EFL students. Among the mapped refusal studies, Widanta et al. (2019) is the one we found that involved foreign learners of Indonesian. They reported pragmalinguistic and sociopragmatic errors linked to negative transfer from English, but not how their instrument was developed or validated. Recent work elsewhere in Indonesia follows the native-speaker and EFL line, with students of three ethnic groups (Pasaribu et al., 2021), native speakers of a regional variety (Marniati & Jaafar, 2022), and EFL learners (Rusdi et al., 2022). For English-L1 learners of a language far from English, the closest comparison we found is Morkus (2021), who studied American learners of Arabic through role-plays. We found nothing similar for Indonesian. Boeriswati et al. (2023), in a bibliometric analysis of their own, found no earlier work on developing a BIPA placement test instrument among 117 studies of foreign-language placement tests (2012 to 2022), so well-documented instruments seem scarce in BIPA assessment more broadly.

This study therefore develops a DCT for English-L1 BIPA learners in an academic campus setting and gathers preliminary evidence of its content validity, usability, and

translation equivalence, with relative power as the controlled sociopragmatic variable. It asks three questions. How do two content experts judge the relevance, clarity, naturalness, and contextual cues of each situation? What readability and wording problems appear when intended users work through the instrument? And where do the Indonesian and English versions differ in meaning?

We do not claim novelty only for a new learner group. The contribution has three parts. Methodologically, the study documents a sequence in which content review, piloting, and back-translation are reported as separate layers of evidence, with the Content Validity Index examined for each rating unit as well as in aggregate. Conceptually, it builds the DCT's known limits into the design, in the construct, the claims, and the permitted uses, where they are usually mentioned only at the end. Empirically, within the literature we mapped it offers an English-L1 refusal DCT for BIPA with documented content evidence, which other researchers can examine and build on.

METHOD

Research Design

This was an instrument-development study that gathered preliminary content-validity evidence for a DCT eliciting written refusals from English-L1 BIPA learners. The instrument moved through drafting, readability trials, expert content review, field piloting, and staged revision. Three layers of evidence were kept apart: content evidence from expert review, usability evidence from readability trials and pilots, and translation-equivalence evidence from blind back-translation. Together they do not amount to psychometric validation, which would also need reliability estimates and criterion or construct evidence from substantive use.

Preliminary Literature Mapping

Before designing the instrument, we mapped how refusal DCTs have been adapted and validated, in the spirit of other mapping work in interlanguage pragmatics (Alrefaee & Aljabr, 2025). The mapping had no pre-registered protocol, and the first author did the screening and coding alone, without a second rater. It covered 92 studies that focused on refusal and used a DCT or a DCT-derived written or oral task, including oral variants such as the spoken DCT in Alajmi (2024). A study counted as reporting explicit content validation only when designated reviewers or experts deliberately judged the content or contextual appropriateness of the DCT. Pilot testing for clarity alone, inter-rater checks on coded responses, and bare statements that an instrument was "validated" did not count.

Instrument Design

Construct. We define refusal-related pragmatic competence as the ability to select and formulate a contextually appropriate Indonesian refusal to a request, invitation, offer, or suggestion, both in form (pragmalinguistic ability, such as regret, reasons, or alternatives) and in judging which combination suits the situation (sociopragmatic ability). Because a DCT elicits what participants think is appropriate to say (Golato, 2003; Kasper, 2000), the term covers only elicitable pragmatic knowledge shown in writing, not interactional performance.

Design. The DCT adapts Beebe et al. (1990): 12 situations from four eliciting acts crossed with three levels of relative power (Table 1), set in Indonesian campus life. Power was the only variable manipulated systematically. Social distance and imposition varied within scenarios, as in Beebe et al.'s own design, and the validators rated whether their cues were appropriate. Ecological validity was not tested separately.

Versions. BIPA learners received a bilingual Indonesian-English version, with both languages in the instructions, situations, and interlocutor lines, and native speakers a monolingual Indonesian version. Each item gave a situation, one line from the interlocutor, and the prompt *Anda menolak dengan mengatakan:* ("Your refusal in Indonesian:"). Responses were written in Indonesian, with no right or wrong answers, in about 30 to 40 minutes for all 12 situations.

Eliciting act	Equal power	Higher power (refuser)	Lower power (refuser)
Request	Situation 1	Situation 5	Situation 9
Invitation	Situation 2	Situation 6	Situation 10
Offer	Situation 3	Situation 7	Situation 11
Suggestion	Situation 4	Situation 8	Situation 12

Table 1. Crossing of eliciting act and relative power across the 12 DCT situations

Note. "Refuser" is the participant's power position relative to the interlocutor who makes the request, invitation, offer, or suggestion. Social distance and imposition were not crossed systematically.

Development Stages, Participants, and Raters

The instrument went through six stages (Table 2), and what each found is reported under Results and Discussion.

Stage	Activity	Participants	Purpose
1	First draft (three eliciting acts; three variables varied in fixed pairs) and readability trial	3 BIPA learners, English L1, A1 to B1	Readability

Stage	Activity	Participants	Purpose
2	Revision to the 4 × 3 structure of Beebe et al. (1990)	None (desk revision)	Simplify the design
3	Informal pilot of the Stage 2 draft, then a pre-validation BIPA trial	Pilot: 4 BIPA learners, English L1, A2 to B1, and 3 Indonesian speakers. Trial: 6 BIPA learners	Clarity and naturalness
4	First-round expert content validation	2 content experts	Content evidence
5	Pilot of the expert-revised version	6 BIPA learners, English L1, B1 (INCULS placement); 6 Javanese-L1 speakers of Indonesian (3 men, 3 women)	Field usability; naturalness check
6	Back-translation, revision, and revalidation of eight situations	2 content experts; 1 translation specialist	Content and translation-equivalence evidence

Table 2. Development stages and participant groups

We defined L1 by age and order of acquisition: participants were English-L1 if they acquired English from early childhood in the family and community. All six Stage 5 learners met this definition and were placed at B1 by INCULS, Universitas Gadjah Mada, through its standard placement procedure.

Two experts rated every situation on four aspects from 1 to 4: relevance to the BIPA context, clarity, naturalness of the dialogue, and appropriateness of the power structure and of the distance and imposition cues. Senowarsito is a Professor of Applied Linguistics specializing in English pragmatics at Universitas PGRI Semarang. Wira Kurniawati is a BIPA specialist at Universitas Gadjah Mada and a co-author of this article (see Declarations). Each situation-aspect pair was one rating unit, which gave 48 units. Translation equivalence followed Brislin (1970). The first author, who wrote the Indonesian original, translated it into English, and Siti Nur'aini, a translation specialist at Universitas PGRI Semarang who completed her doctorate in New Zealand, translated the English back into Indonesian without seeing the original. Any difference in denotation, register, or pragmatic force counted as non-equivalence and led to a revision of the English.

Research Ethics

The study received research permission from FIB, Universitas Gadjah Mada, on 20 July 2026 (No. 4426.1/UN1/SA.1.1/PASCA/KM.00.00/2026) and ethical clearance from the Research Ethics Committee, Universitas Gadjah Mada, on 22 July 2026 (No. KE/UGM/144/EC/2026). Participation was voluntary, and participants were told they could withdraw at any time without consequence. All participants completed an informed

consent form before taking part in the pilot testing. The expert validators and the translators also gave their consent to take part and to have their ratings, comments, and translations used in this study. Participants were identified by code and not by name in the data, and the data are stored securely and accessible only to the research team.

Data Analysis

Scores of 3 and 4 counted as valid and scores of 1 and 2 as invalid (Lynn, 1986; Polit & Beck, 2006). The I-CVI of a rating unit is the share of validators who rated it valid. With two validators it can only be 0.00, 0.50, or 1.00, so any disagreement gives 0.50 and the cut-offs quoted for larger panels (such as 0.78) cannot be reached; Romero Jeldres et al. (2023), for example, worked with eight experts. We therefore flagged each unit without agreement from both validators to qualitative review and did not judge it against a cut-off. Following Polit and Beck's (2006) advice to report more than one scale-level index, we calculated S-CVI/Ave (the mean I-CVI over the 48 units) and S-CVI/UA (the share of units rated valid by both validators), taking 0.90 as a working reference only. Both describe expert agreement on these 12 situations only. The experts' comments were read descriptively and grouped into narrative logic, clarity of contextual parameters, lexical familiarity, naturalness of dialogue, and gender representation.

RESULTS AND DISCUSSION

We report the findings in the order the evidence arrived, with our reading of each part following it. The last part brings the layers together.

Early Trials and the Decision to Control Only Power

The first draft varied three sociopragmatic variables at once, in fixed pairs: close distance always came with equal status, and distant relationships always with unequal status. Three BIPA learners found a single situation too heavy to process. In Stage 2 we moved to the 4 × 3 structure of Beebe et al. (1990), added the suggestion act, and kept power as the only variable controlled.

The Stage 2 draft was then tried informally with seven participants. Six gave a clarity rating for each situation, and one gave none. Two situations were consistently hard. In the academic-advisor suggestion scenario (Situation 12), all six respondents reported difficulty: four rated it "unclear", one "not really" clear, and one wrote an answer that agreed with the suggestion, so the scenario did not elicit a refusal at all. In the English-tutor suggestion scenario, which contrasted grammar with conversation (Situation 8), three of

the six rated it “not really” clear. Both were revised before the experts saw the instrument. A pre-validation trial with six BIPA learners then showed that several words were unfamiliar to them, and these were replaced.

The Stage 1 result gave us an empirical reason, as well as a theoretical one, for controlling power alone. It fits Billmyer and Varghese's (2000) finding that the way a DCT prompt is written affects the responses it gets. Distance and imposition may still matter a great deal. Al-Sallal (2024) found that social distance, degree of imposition, and gender each changed which refusal strategies Bahraini and Indian learners of English used. By letting these vary inside scenarios we accepted that later analyses will have to treat them as uncontrolled context. The advisor scenario shows a second risk. A situation that fails to elicit a refusal is easy to mistake, once the data are in, for a learner’s pragmatic choice, and a grid as widely reused as Beebe et al.’s does not guarantee that every cell works in a new language and for a new learner group.

Expert Content Review and Revalidation

Across the 48 rating units, the first-round scores gave an S-CVI/Ave of 0.98 (0.9792) and an S-CVI/UA of 0.96 (0.9583), both above the 0.90 reference. Both validators agreed on 46 units (I-CVI = 1.00) and only one did on 2 (I-CVI = 0.50), which is why the indices differ: 47/48 against 46/48. The two units were the clarity aspect of Situations 1 and 2.

The comments pointed to the same places. Senowarsito questioned the narrative logic of Situation 1, a note-borrowing scenario that still contained an inconsistency, and found the closeness parameter of Situation 2 under-specified. Both situations were revised after this round and before the Stage 5 pilot. The experts also asked for more gender variation among characters (Senowarsito) and for more familiar word choices and smoother dialogue (Wira Kurniawati), for example *mendatangi* in place of *menghampiri* (‘to approach’) in four situations.

After the Stage 5 pilot and the translation review, further situations were revised (Table 3), and eight (1, 2, 3, 5, 6, 7, 11, and 12) went back to both validators. Situation 12 had by then been revised twice. Both validators gave full scores (4/4) on every aspect of all eight, noting that the revisions did not change the context or substance of any situation, stayed consistent with the intended power structure and the distance and imposition conditions, and in some cases made the English clearer. This is agreement between two validators on revised content, and we do not offer it as evidence of psychometric properties. On this basis the instrument was finalized for the main data collection, which

will be reported separately. The final 12-situation instrument, in bilingual and monolingual forms, is provided as a supplementary file.

An average of 0.98 coexisted with two units at the lowest score two raters can produce. Reported alone, S-CVI/Ave would have read as "nothing to fix"; looking at rating units and reading the comments beside the numbers showed where the work was. The same setup limits what the index can say. With two raters, one dissent is enough to give 0.50, and the number cannot separate a real flaw from a difference in judgment, so the written comments did that. Rating contextual cues as an aspect of its own also has support. Hosseinpur et al. (2022) found that richer contextual detail in DCT prompts brought elicited requests closer to classroom data for head acts and internal modifiers, though not for external modifiers. Finally, expert feasibility review already has a place in BIPA materials development (Muliastuti et al., 2023). We carried the same logic to an elicitation instrument, where the experts judge a situation, not a lesson.

Situation	Aspect flagged	I-CVI (Round 1)	Expert concern	Field or translation evidence	Revision
1	Clarity	0.50	Narrative logic inconsistent (note-return timing unclear)	Back-translation: time marker non-equivalent	Scenario rewritten (missed-class framing); time marker later set to "At noon". Revalidated, 4/4
2	Clarity	0.50	Closeness parameter under-specified	Not separately tested	Wording revised; *menghampiri* changed to *mendatangi*. Revalidated, 4/4
3	None	1.00	None at Round 1	Back-translation: *mengantuk* returned as "tired"	English corrected to "sleepy". Revalidated, 4/4
5	None	1.00	None at Round 1	Back-translation: *sore* returned as "evening"	English time marker corrected to "afternoon". Revalidated, 4/4
6	None	1.00	None at Round 1	Back-translation: *pelantikan* returned as "handover ceremony"	English corrected to "inauguration ceremony". Revalidated, 4/4
7	None	1.00	Word familiarity questioned by BIPA validator	Field pilot: one BIPA participant marked the wording unclear. Back-translation: *tidak sengaja* returned as "totally my fault"	Indonesian revised (*tersenggol* to *tertabrak*); English revised to "It was an accident." Revalidated, 4/4
8	Clarity (informal pilot)	n/a (revised before Round 1)	None	Informal pilot: 3 of 6 respondents rated it "not really" clear	Scenario simplified before expert validation
11	None	1.00	None at Round 1	Back-translation: *Ketua Program Studi* returned as "Department Chair"	English corrected to "Head of your Study Program". Revalidated, 4/4

Situation	Aspect flagged	I-CVI (Round 1)	Expert concern	Field or translation evidence	Revision
12	Clarity (informal pilot)	n/a (revised before Round 1)	None	Informal pilot: all 6 respondents reported difficulty (4 "unclear", 1 "not really", 1 non-refusal response)	Revised twice, including replacement of the academic-advisor deadline scenario by an internship-supervisor note-taking scenario. Revalidated, 4/4

Table 3. Revision history and later validation of the DCT situations that changed

Note. I-CVI (Round 1) refers to the rating units from the first formal expert round (Stage 4). Situations 8 and 12 were revised during the informal pilot stage, before that round, so their earlier forms were not scored ("n/a"). One of seven pilot participants gave no clarity rating. Situations 4, 9, and 10 raised no concern at Round 1 or in back-translation and are left out.

Field Testing with B1 Learners

The expert-revised version was piloted with six B1-level BIPA learners at INCULS and with six Javanese-L1 speakers of Indonesian. The Javanese-L1 group served only as a check on local naturalness and comprehensibility. It was not evidence that the instrument suits English-L1 BIPA learners, and it was not a comparison group in any analysis here. Across 72 situation exposures (six learners by twelve situations), eleven situations raised no notable comprehension problem. In the exception, the janitor offer-of-replacement scenario (Situation 7), one participant marked the word for the coffee-spilling incident as unclear. Wira Kurniawati had raised the same word in the first round, so the agreement between expert and learner settled the decision to change it.

One flag in 72 exposures is encouraging, but six learners at one proficiency level cannot carry much, and proficiency seems to matter for this kind of instrument. Morkus (2021) found that intermediate American learners of Arabic transferred negatively from English more often than advanced learners. García (2024) reported that Japanese learners' linguistic ability correlated with pragmatic transfer, and Al-Gahtani and Roever (2018) found that proficiency affects how learners organize misreferred responses. If this DCT is used to compare proficiency groups, its readability at A2 and at higher levels should be piloted first, and so should the amount of English support in the bilingual version. Morkus (2021) also used role-plays at lower, equal, and higher status, which suggests one way to triangulate the written data (see Conclusion).

Translation Equivalence

Blind back-translation found six discrepancies, in Situations 1, 3, 5, 6, 7, and 11 (Table 3). They ranged from word meaning (*mengantuk*, 'sleepy', returned as "tired") to

institutional terms (*Ketua Program Studi*, 'Head of Study Program', returned as "Department Chair", which is closer to *Ketua Jurusan*) and emotional register. *Tidak sengaja* ('unintentionally', a mild register) returned as "totally my fault", a full admission of fault. The time markers *siang* and *sore* in Situations 1 and 5 were also not equivalent. For Situation 1 we first considered keeping the English marker because its range seemed wide enough, and then replaced it with "At noon", which is clearer and matches the other situations. Situation 7 had two separate problems: the familiarity of one Indonesian word, found in the field pilot, and the register of the English, found here. Both were revised and confirmed in the revalidation. Of the six discrepancies, five led to final-stage revisions, and the sixth, in Situation 7, had been handled earlier.

Five of the six (all but Situation 1) sat in situations that had received full agreement from both experts in the first round, and four of these (Situations 3, 5, 6, and 11) were not flagged in the field pilot either. High content validity for the Indonesian version did not guarantee that the English said the same thing, and a learner reading the English side would have been answering a slightly different situation. For this reason, we treat back-translation (Brislin, 1970) as a layer of evidence in its own right, not an administrative step. Neither the experts nor the field participants were placed to see these shifts, because each mainly engaged with one language version. Back-translation does not establish equivalence by itself, so we paired it with expert revalidation of every revised situation. The question is not peculiar to BIPA. Vásári (2022) studied refusals in Hungarian and English with an adapted DCT, and cross-language DCT use is routine. In the refusal studies we mapped, equivalence between versions of a bilingual instrument received little attention.

What the Three Layers Add

Table 4 sets out what each layer caught and what it missed. None would have produced the final instrument alone.

Layer	What it caught in this study	What it did not catch
Informal pilot and pre-validation trial (Stage 3)	A scenario that did not elicit a refusal (Situation 12); unclear framing (Situation 8); unfamiliar words	Whether situation cues were pragmatically appropriate
Expert review (Stages 4 and 6)	Narrative inconsistency (Situation 1); under-specified closeness (Situation 2); gender variation; word familiarity	Meaning shifts in the English version (Situations 3, 5, 6, 7, and 11 all scored I-CVI = 1.00 in Round 1)
Field pilot (Stage 5)	A word unfamiliar to B1 learners (Situation 7)	Register shifts in the English; readability at other proficiency levels
Back-translation (Stage 6)	Six equivalence discrepancies (Situations 1, 3, 5, 6, 7, and 11)	Whether the revised wording was appropriate, which needed revalidation

Table 4. What each evidence layer caught and missed in this study

The first contribution is conceptual. Golato (2003) and Kasper (2000) tell us that DCT responses reflect what participants judge appropriate more than what they do, and most studies mention this near the end as a limitation. Here it was put in at the start. It shaped the construct (elicitable pragmatic knowledge), the claims we make (content appropriateness, usability, translation equivalence, but not psychometric validity), and the limits on use. Labben (2022) argued that a DCT should be treated as a language test with a careful development process, and this study applies that argument to BIPA. A bounded claim also leaves room for other data. Al-Sallal (2024), who used role-plays and a DCT, found that learners used different strategies under the two methods, and El Hammouchi (2025) paired a DCT with focus-group interviews, which put learners' own reasons for refusing beside what they wrote.

The second contribution is methodological and offers three things others can reuse. Content validity was reported for rating units as well as at scale level, which showed that a 0.98 average can hide two units at 0.50 in a two-rater panel. Translation equivalence was kept as a layer of its own. The number of controlled variables was settled by an early readability trial. Set against the small minority of mapped studies that reported any explicit content validation, a documented sequence of situation-level review, piloting with the intended learners, and equivalence checking before substantive data collection is a small but practical addition.

The third is practical, for BIPA. The introduction noted that research has described forms, materials, and teachers, but not what learners do when they have to refuse. The DCT reported here is built to ask that. It is provided as a supplementary file, with a stated intended range: English-L1 learners at B1 in university settings.

CONCLUSION

This study set out to develop a 12-situation refusal DCT for English-L1 BIPA learners and to gather preliminary evidence of its content appropriateness, usability, and translation equivalence. The study produced three main findings. A high aggregate agreement (S-CVI/Ave = 0.98) hid two rating units at the lowest score two raters can give, so reading the units and the experts' comments together is what showed where revision was needed. Content validity of the Indonesian version did not guarantee that the English version said the same thing: back-translation found six discrepancies, five of them in situations that had earned full expert agreement. No single layer of evidence would have

produced the final instrument, because the early trials, expert review, field pilot, and back-translation each caught problems the others missed. After revision, both validators gave full scores to all eight revalidated situations.

These findings carry three implications. For refusal research, adapting Beebe et al.'s framework to a new language and learner group should come with situation-level content review, piloting with the intended users, and an equivalence check before main data collection, and reports should say what was done. For BIPA, there is now a documented instrument for studying how English-L1 learners refuse in Indonesian, a question that research on forms, materials, and teachers has left open. For the DCT debate, bounding the claims from the start, to elicitable pragmatic knowledge rather than interactional behavior, lets the instrument be used honestly for what it can do.

The evidence is preliminary, and several limits apply. The literature mapping was exploratory and coded by one person, so it shows a general pattern and cannot be read as a count of the literature. The panel had two experts, one of them a co-author, and with two raters an I-CVI can only be 0.00, 0.50, or 1.00. The pilot groups were small, and the field pilot used a single proficiency level (B1). Written responses show what participants think is appropriate to say, not what they do in conversation. For these reasons the instrument should be used only with English-L1 BIPA learners in comparable academic settings and at comparable proficiency, and wider use needs further piloting.

The next steps follow from these limits: a larger and more varied expert panel, piloting at other proficiency levels, and triangulation with role-plays or retrospective interviews. The main data collection with this instrument will be reported separately.

ACKNOWLEDGMENTS (OPTIONAL)

The authors would like to thank INCULS, Universitas Gadjah Mada; Universitas Gadjah Mada; Universitas PGRI Semarang who supported the completion of this study through academic, technical, or administrative assistance. The authors also sincerely appreciate the constructive comments and suggestions provided by the anonymous reviewers and editors, which have substantially improved the quality of this manuscript. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DECLARATIONS

Ethical considerations

Research permit: INCULS, Universitas Gadjah Mada, 20 July 2026, No. 4426.1/UN1/SA.1.1/PASCA/KM.00.00/2026. Ethical clearance: Research Ethics Committee, Universitas Gadjah Mada, 22 July 2026, No. KE/UGM/144/EC/2026. Informed consent was obtained as described under Research Ethics in the Method section.

Use of artificial intelligence (AI)

The authors declare that the generative artificial intelligence (AI) tool Claude (Anthropic) and Grammarly were used exclusively for language editing and/or grammatical improvement. The use of AI did not influence the scientific content, study design, data analysis, data interpretation, results, or conclusions of the manuscript. Full responsibility for the content remains with the authors.

Conflict of Interest

One of the two content validators for this instrument, Wira Kurniawati, is also a co-author of this article. She is a recognized BIPA specialist and currently chairs APPBIPA Yogyakarta, which is why she was invited as a validator. To reduce the risk of influence from this authorship relationship, her content-validation scores and qualitative notes were recorded independently and without access to the other validator's ratings, using the same evaluation procedure applied throughout the study; this procedure reduces, but does not eliminate, the potential influence of the relationship. Readers should note that this validator's judgment is not fully independent of the authorship team, and this is reflected as a limitation in the Conclusion. Beyond this disclosed relationship, the authors declare no other conflict of interest.

Funding

This research did not receive any financial support.

REFERENCE

Alajmi, N. M. (2024). Analysing refusal strategies in Kuwaiti Arabic and British English: Cross-cultural pragmatics. *English Language Teaching, 17*(9), 107-122. https://doi.org/10.5539/elt.v17n9p107

Al-Gahtani, S., & Roever, C. (2018). Proficiency and preference organization in second language refusals. *Journal of Pragmatics*, *129*, 140-153.

https://doi.org/10.1016/j.pragma.2018.01.014

Alrefaee, Y. M. H. A., & Aljabr, F. (2025). Evolution and trends in interlanguage pragmatics research: A bibliometric analysis (1973–2024). *Cogent Arts and Humanities*, *12*(1). https://doi.org/10.1080/23311983.2025.2567505

Al-Sallal, R. E. (2024). Pragmatic analysis of refusal strategies in spoken English of Bahraini and Indian L2 learners. *Cogent Arts & Humanities, 11*(1), Article 2321682. https://doi.org/10.1080/23311983.2024.2321682

Asteria, P. V., Rofiuddin, A., Suyitno, I., & Susanto, G. (2023). Indonesian-based pluricultural competence in BIPA teachers' perspective. *Eurasian Journal of Applied Linguistics, 9*(1), 190-201. https://doi.org/10.32601/ejal.901016

Beebe, L. M., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. C. Scarcella, E. S. Andersen, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55-73). Newbury House.

Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics, 21*(4), 517-552. https://doi.org/10.1093/applin/21.4.517

Blum-Kulka, S. (1982). Learning to say what you mean in a second language: A study of the speech act performance of learners of Hebrew as a second language. *Applied Linguistics, 3*(1), 29-59. https://doi.org/10.1093/applin/III.1.29

Boeriswati, E., Ningsih, R. Y., & Rahayu, W. (2023). Web-based design of BIPA placement test instrument for foreign speakers. *International Journal of Learning, Teaching and Educational Research*, *22*(2), 392-412. https://doi.org/10.26803/ijlter.22.2.22

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185-216. https://doi.org/10.1177/135910457000100301

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

Dwiana, N. R., Basthomi, Y., Anugerahwati, M., & Syahri, I. (2021). "I'm afraid I can't": Initiating acts on refusal strategy realization. *Indonesian Journal of English Language Teaching and Applied Linguistics, 5*(2), 297-308. https://doi.org/10.21093/ijeltal.v5i2.716

El Hammouchi, A. (2025). Saying 'no' politely: The use of refusal strategies by Moroccan EFL university learners. *Studies in Pragmatics and Discourse Analysis, 6*(2), 17-30. https://doi.org/10.48185/spda.v6i2.1732

Eviliana, N. (2019). Pragmatic transfer in refusals realized by proficient EFL learners English Education Program Sriwijaya University. *Mabasan, 9*(1), 43-54. https://doi.org/10.62107/mab.v9i1.155

García, I. P. (2024). Refusals in Japanese and Spanish: Pragmatic transfer in L2. *Acta Linguistica Asiatica, 14*(2), 93-119. https://doi.org/10.4312/ala.14.2.93-119

Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics, 24*(1), 90-121. https://doi.org/10.1093/applin/24.1.90

Gustini, M. (2018). Contrastive analysis of refusal expressions in Japanese and Bahasa at work place. *JAPANEDU: Jurnal Pendidikan dan Pengajaran Bahasa Jepang, 3*(2), 132-139. https://doi.org/10.17509/japanedu.v3i2.13501

Hosseinpur, R. M., Lowni, A., & Lowni, M. (2022). Does interlanguage pragmatic measure enhancement change the data pattern? A mixed method approach. *Issues in Language Teaching, 11*(2), 1-32. https://doi.org/10.22054/ilt.2022.68389.704

Kasper, G. (2000). Data collection in pragmatics research. In H. Spencer-Oatey (Ed.), *Culturally Speaking Second Edition: Culture, Communication and Politeness Theory* (pp. 316-369). Continuum International Publishing Group. https://doi.org/10.5040/9781350934085.ch-014

Kasper, G., & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition*, *13*(2), 215-247. https://doi.org/10.1017/S0272263100009955

Kholiq, A., Yulianto, B., & Suhartono. (2024). Role of L1 and L2 in the acquisition of Bahasa Indonesia as a third language. *Issues in Language Studies, 13*(1), 413-431. https://doi.org/10.33736/ils.5567.2024

Kurniawati, W. (2026). "Permisi, Bu!" Acquisition of Indonesian politeness for BIPA learners. *Tabasa: Jurnal Bahasa, Sastra Indonesia, Dan Pengajarannya*, *7*(01), 319-336. https://doi.org/10.22515/tabasa.v7i01.14493

Kurniawati, W., & Ginting, S. R. (2026). Requesting with an appropriate degree: The request speech act competence of BIPA learners. *Bahastra*, *46*(1), 196-207. https://doi.org/10.26555/bs.v46i1.1982

Labben, A. (2016). Reconsidering the development of the discourse completion test in interlanguage pragmatics. *Pragmatics*, *26*(1), 69-91. https://doi.org/10.1075/prag.26.1.04lab

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*,

35(6), 382-385. https://doi.org/10.1097/00006199-198611000-00017

Mahriyuni, M., Pramuniati, I., & Sitinjak, D. R. (2024). Interlanguage development among the learners of Indonesian language in Paris. *Indonesian Journal of Applied Linguistics, 14*(1), 206-219. https://doi.org/10.17509/ijal.v14i1.70394

Marniati, M., & Jaafar, M. F. (2022). The realisation of the refusal speech act in Tanjung Tanah dialect of Kerinci. *Jurnal Arbitrer, 9*(2), 94-106. https://doi.org/10.25077/ar.9.2.94-106.2022

Mokoginta, K., & Arafah, B. (2022). Negotiation in Indonesian culture: A cultural linguistic analysis of Bahasa Indonesia textbooks. *Theory and Practice in Language Studies, 12*(4), 691-701. https://doi.org/10.17507/tpls.1204.09

Morkus, N. (2021). Negative pragmatic transfer and language proficiency: American learners of Arabic. *The Language Learning Journal, 49*(1), 41-65. https://doi.org/10.1080/09571736.2018.1474247

Muliastuti, L., Mayuni, I., Nurhaina, A., & Saddhono, K. (2023). Tailoring CEFR to BISOL (Bahasa Indonesia for Speakers of Other Languages): A model for integrative language teaching materials. *International Journal of Language Education, 7*(4), 590-601. https://doi.org/10.26858/ijole.v7i4.53219

Pasaribu, A. N., Sinambela, E., Pasaribu, T. K., Sihite, J. E., & Sembiring, R. A. (2021). The strategies of refusing requests by students of different ethnic groups. *Syntax Literate ; Jurnal Ilmiah Indonesia*, *6*(1), 67-73. https://doi.org/10.36418/syntax-literate.v6i1.2178

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, *29*(5), 489-497. https://doi.org/10.1002/nur.20147

Primantari, A. N., & Wijana, I. D. P. (2017). Tindak tutur meminta oleh pembelajar BIPA dari Korea: Kajian pragmatik bahasa antara (interlanguage pragmatics). *Jurnal Penelitian Humaniora*, *18*(1), 27-40. https://doi.org/10.23917/humaniora.v18i1.3638

Romero Jeldres, M., Díaz Costa, E., & Faouzi Nadim, T. (2023). A review of Lawshe's method for calculating content validity in the social sciences. *Frontiers in Education, 8*, Article 1271335. https://doi.org/10.3389/feduc.2023.1271335

Rusdi, R. P., Purwanti, I. T., & Dahnilsyah, D. (2022). Refusal strategies used by EFL learners of English Study Program in Riau, Indonesia. *IDEAS: Journal of Language Teaching and Learning, Linguistics and Literature, 10*(1), 175-186.

https://doi.org/10.24256/ideas.v10i1.2581

Siddiqa, A., & Whyte, S. (2025). Classroom learning of English L2 requests: Input and interactional opportunities in French secondary schools. *Language Teaching Research, 29*(2), 535-559. https://doi.org/10.1177/13621688211064932

Susanto, G., Pickus, D., Espree-Conaway, D., Suparmi, Rusiandi, A., & Noviya, H. (2024). Indonesian language policy and perspectives on its implementation in promoting Bahasa Indonesia as an international language. *Cogent Arts & Humanities, 11*(1), Article 2364511. https://doi.org/10.1080/23311983.2024.2364511

Syafruddin, Pratiwi, B., Izzati, A. N., & Sianipar, I. F. (2025). Teaching Indonesian for foreign speakers (BIPA): A bibliometric analysis. *Issues in Language Studies, 14*(1), 55-73. https://doi.org/10.33736/ils.7551.2025

Vásári, N. (2022). Pragmatic transfer in the refusal strategy choices of Hungarian EFL students. *Alkalmazott Nyelvtudomány, 22*(1), 208-222. https://doi.org/10.18460/ANY.2022.1.013

Widanta, I. M. R. J., Hudiananingsih, P. D., Sitawati, A. A. R., & Ardika, I. W. D. (2019). Pragmatic errors and transfer of foreign learners of Indonesian: The case of refusals. *Journal of Language Teaching and Research, 10*(3), 501-508. https://doi.org/10.17507/jltr.1003.13

Widianto, E., & Pusse, T.-K. (2025). Cultura franca: Investigating cultural gradations of Indonesian as a foreign language textbook. *Asian-Pacific Journal of Second and Foreign Language Education, 10*(1), Article 30. https://doi.org/10.1186/s40862-025-00335-y