International Journal of Social Science, Education and Humanities
ISSN Print: 3104-4239
ISSN Online: 3104-4247
CD ISSN: 3104-4255
https://doi.org/10.63313/IJSSEH.9067

Assessing Intercultural Pragmatic Competence in Chinese EFL Learners: An Integrated Framework Using Written Discourse Completion Tasks and Audio Role-Plays

Wenrui Zhang*
University of Sydney, Camperdown 2050 NSW AU, Australia
Email: wzha0657@uni.sydney.edu.au

How to cite this paper: Zhang, W. (2026). Assessing intercultural pragmatic competence in Chinese EFL learners: An integrated framework using written discourse completion tasks and audio role-plays. International Journal of Social Science, Education and Humanities, 3(2), 66–81. ISSN Print: 3104-4239, ISSN Online: 3104-4247.
https://doi.org/10.63313/IJSSEH.9067
Published: 2026-11-06

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

Abstract

Assessing intercultural pragmatic competence remains a methodological challenge in English as a foreign language (EFL) education. Although written discourse completion tasks (WDCTs) are widely used to investigate learners' pragmatic knowledge, their capacity to represent the interactional and context-dependent nature of communication is limited. Role-play assessments offer opportunities to examine pragmatic performance in interaction, yet introduce challenges related to standardisation, scoring and contextual authenticity. Drawing on research in interlanguage pragmatics, intercultural communication and language assessment, this article proposes an integrated assessment framework combining WDCTs with interactive audio role-plays for Chinese university EFL learners. The framework distinguishes sociopragmatic judgement, pragmalinguistic realisation and interactional management while recognising that these dimensions are not equally observable across assessment modalities. A task specification, analytic scoring framework and staged validation strategy are developed to support future empirical implementation. Particular attention is given to contextual variation, intercultural agency, rating fairness and the risks of treating native-speaker conventions as universal standards. Rather than claiming that the proposed instrument has already been validated, this methodological article identifies the evidence required to evaluate its reliability, construct representation and interpretive usefulness. The framework offers a practical basis for investigating L2 pragmatic performance and for developing context-sensitive assessment practices across EFL educational settings.

Keywords

Intercultural Pragmatic Competence; Chinese EFL Learners; Discourse Completion Tasks; Audio Role-Plays; Pragmatic Assessment; Interactional

1. Introduction

The development of pragmatic competence is a central concern in second language (L2) learning because successful communication requires more than grammatical accuracy and lexical knowledge. Language users must also interpret communicative intentions, negotiate interpersonal relationships and select linguistic resources that are appropriate to particular social contexts. In intercultural encounters, these processes are especially complex because interlocutors may draw on different expectations regarding politeness, directness, interpersonal distance and communicative responsibility.

Within English as a foreign language (EFL) education, pragmatic assessment presents an important challenge. Conventional language tests tend to privilege relatively stable and observable dimensions of language knowledge, whereas pragmatic competence involves context-sensitive choices that may vary across individuals, communities and communicative situations. Roever (2011) argues that established pragmatic assessments have frequently concentrated on speech acts and pragmalinguistic knowledge while insufficiently representing extended discourse and interactive language use.

This problem is particularly relevant to Chinese university EFL learners, whose opportunities for English-language communication may differ substantially according to educational environments, international exposure and individual language-learning experiences. Research conducted in Chinese EFL settings has demonstrated the feasibility of developing pragmatic assessment instruments. Liu (2007), for example, designed a test focusing on apology-related pragmatic knowledge, while Xu and Wannaruk (2016) developed and evaluated a written discourse completion assessment encompassing multiple speech acts.

More recent research has broadened the methodological possibilities of pragmatic assessment. Zhang et al. (2026) investigated pragmatic development among Chinese EFL learners in an English-medium instruction environment using multiple assessment tasks. Lu et al. (2026) developed a computerised dynamic diagnostic approach to assessing pragmatic competence in refusals. Together, these studies demonstrate continued interest in assessing different dimensions of pragmatic knowledge and performance.

Nevertheless, the relationship between assessment modality and the interpretation of pragmatic competence requires further methodological attention. Written discourse completion tasks (WDCTs) allow researchers to present controlled scenarios and systematically elicit linguistic responses. Their relatively low administrative demands make them attractive for educational research. However, they generally provide limited access to real-time interactional processes, including clarification, conversational repair and the negotiation of interpersonal meaning.

Interactive role-plays, by contrast, can create opportunities for learners to respond to evolving communicative circumstances. They may reveal how pragmatic decisions develop across conversational turns rather than appearing as isolated

linguistic products. However, the added interactional complexity also raises concerns about variability in interlocutor behaviour, scoring reliability and the extent to which elicited performances represent naturally occurring communication (Félix-Brasdefer, 2007).

These differences suggest that combining written and oral tasks may broaden construct representation, provided that the two modalities are not treated as interchangeable measures of a single, uniform ability.

The present article therefore proposes an integrated framework for assessing intercultural pragmatic performance among Chinese university EFL learners. Its contribution is methodological rather than empirical: it develops a theoretically informed design for combining WDCTs with interactive audio role-plays and specifies a validation process through which the framework may subsequently be evaluated.

Three questions guide the discussion:

1. How can intercultural pragmatic competence be operationalised in an assessment framework that accommodates both written responses and spoken interaction?
2. What complementary evidence can WDCTs and interactive audio role-plays provide about learners' pragmatic performance?
3. What forms of validity, reliability and fairness evidence are necessary before results from the proposed framework can support educational or research interpretations?

By addressing these questions, this article seeks to contribute to the design of theoretically grounded and practically feasible pragmatic assessment instruments in EFL contexts.

2. Theoretical Background

2.1. Pragmatic Competence and Intercultural Communication

Pragmatic competence can broadly be understood as the ability to interpret and communicate meaning appropriately within social contexts. It involves both the linguistic resources available to language users and their understanding of how communicative choices relate to interpersonal relationships, situations and social expectations.

A useful distinction can be made between pragmalinguistic knowledge and sociopragmatic judgement. Pragmalinguistic knowledge concerns the relationship between linguistic forms and communicative functions. Sociopragmatic judgement concerns the evaluation of contextual and interpersonal circumstances that influence how those functions are expressed.

For example, a learner requesting assistance from a classmate may use a relatively direct expression, whereas a request to a university administrator may require greater contextual explanation or interpersonal mitigation. Successful communication does not necessarily depend on producing the most indirect

formulation. Rather, it depends on selecting and negotiating a communicative strategy that fits the situation and the participants' purposes.

An intercultural perspective further complicates the assessment of pragmatic competence. Deardorff (2006) conceptualises intercultural competence as encompassing attitudes, knowledge, skills and developmental processes involved in effective and appropriate intercultural interaction. Liddicoat and Scarino (2013) similarly emphasise that language users actively interpret and construct meaning through intercultural engagement.

These perspectives caution against reducing intercultural competence to adherence to a fixed collection of linguistic conventions.

Baker (2015) further highlights the importance of intercultural awareness in contexts where English functions as a shared communicative resource among speakers from diverse linguistic and cultural backgrounds. Such contexts challenge simplistic assumptions that effective communication requires the reproduction of a single national or native-speaker norm.

Accordingly, the proposed framework adopts a context-sensitive understanding of intercultural pragmatic competence. Its focus is not whether Chinese EFL learners consistently reproduce predetermined expressions associated with an idealised English-speaking community. Instead, it examines whether learners can make intelligible, contextually defensible and relationally responsive pragmatic choices in specified communicative situations.

Importantly, performance on these tasks cannot establish a learner's overall intercultural competence. The proposed instrument addresses a bounded component of that broader construct: the production and negotiation of pragmatic meaning in simulated intercultural encounters.

2.2. Written Discourse Completion Tasks

Written discourse completion tasks have been widely employed in interlanguage pragmatics research. Participants are presented with descriptions of social situations and asked to produce responses appropriate to those contexts.

The method offers several practical advantages. First, researchers can systematically manipulate contextual variables, such as relative social status, interpersonal familiarity and perceived imposition. Second, written responses can be collected efficiently across relatively large samples. Third, consistent task conditions facilitate comparisons between participants and communicative situations.

Previous studies support the usefulness of written elicitation methods for investigating pragmatic knowledge. Liu (2007) demonstrated how carefully developed contextual items could contribute to the assessment of apology-related pragmatic competence among Chinese EFL learners. Xu and Wannaruk (2016) subsequently employed written discourse completion tasks in a broader assessment

involving multiple speech acts.

However, these advantages must be considered alongside methodological limitations.

Yuan (2001) compared several methods of collecting speech-act data and demonstrated the importance of data-collection conditions for the linguistic evidence obtained. Sasaki (1998) likewise found that production questionnaires and role-plays elicited different characteristics of learners' speech-act performance.

WDCT responses are therefore best understood as elicited representations of what learners consider possible or appropriate responses, rather than direct reproductions of naturally occurring conversation.

A learner may demonstrate awareness of an appropriate refusal strategy in a written task but experience difficulty producing a similar strategy under the processing demands of spontaneous interaction. Conversely, a learner who produces a brief written response may be capable of negotiating interpersonal meaning effectively in conversation.

For this reason, written performance should not be interpreted as a complete representation of pragmatic ability.

2.3. Role-Plays and Interactional Competence

Role-play assessments are valuable because pragmatic meaning is often constructed across multiple conversational turns.

Rather than responding to a static written prompt, learners must interpret an interlocutor's contribution, formulate an appropriate response and potentially modify their communicative strategy as the interaction develops.

Galaczi and Taylor (2018) describe interactional competence as involving both individual and socially situated dimensions of spoken language use. This perspective suggests that successful performance involves not only the production of linguistic forms but also participation in the collaborative organisation of interaction.

In pragmatic assessment, interactive role-plays may reveal processes such as initiating requests, responding to resistance, negotiating refusals, clarifying misunderstandings and managing interpersonal disagreement.

Félix-Brasdefer (2007) found that open role-plays could provide useful interactional data under controlled contextual conditions, while also identifying differences between role-play discourse and naturally occurring speech.

These findings indicate that role-plays should neither be dismissed as artificial nor treated as straightforward equivalents of authentic communication. Their value depends on the purpose of assessment, the design of the interaction and the interpretation of the resulting performance.

Audio-only role-plays offer a potentially manageable option for educational research. They allow spoken interaction to be recorded for subsequent analysis without requiring video-based observation. Nevertheless, audio-only interaction

excludes visual and embodied resources, including gaze, gesture and facial expression. These limitations should be recognised when interpreting performance.

2.4. The Need for Complementary Assessment Evidence

A methodological argument for integrating WDCTs and audio role-plays follows from their different affordances.

Written tasks may provide relatively controlled evidence of learners' ability to select and formulate pragmatic responses when additional planning time is available. Interactive audio tasks can provide evidence of pragmatic performance under spoken interactional demands.

However, combining the two methods does not automatically produce a more valid measure. Construct representation depends on the relationship among task design, scoring procedures and intended interpretations.

Kane (2013) argues that validation requires evaluating the inferences connecting observed responses to the interpretations and uses of assessment scores. Applying this argument to pragmatic assessment suggests that each task should have a clearly defined evidential function.

The proposed framework therefore treats the two modalities as complementary sources of evidence rather than assuming that they measure identical constructs or should yield equivalent scores.

3. Design Principles of the Integrated Framework

Four principles underpin the proposed assessment framework.

Construct alignment. Each assessment task should be linked to a defined aspect of pragmatic performance. The scoring criteria should reflect what the task can reasonably elicit rather than what the framework ideally wishes to measure.

Contextual comparability. Tasks should represent a range of interpersonal circumstances while controlling unnecessary differences in topic familiarity, linguistic difficulty and situational complexity.

Interactional representation. Audio role-plays should include opportunities for response, clarification and negotiation, rather than simply requiring learners to deliver a spoken version of a written answer.

Interpretive fairness. Pragmatic performance should be evaluated with reference to communicative effectiveness and contextual justification, not automatic conformity to a single culturally prescribed expression.

These principles inform the instrument specifications presented below.

4. Proposed Assessment Protocol

4.1. Target Population and Intended Use

The proposed framework is designed for adult Chinese university EFL learners, particularly those enrolled in general English, English for Academic Purposes or

intercultural communication programmes.

It is intended primarily for research and low-stakes diagnostic assessment, rather than immediate use in high-stakes proficiency certification.

The framework may be used to investigate differences in performance across assessment modalities, examine context-sensitive pragmatic choices and identify potential areas for pedagogical development.

For an initial empirical validation study, a sample representing more than one English proficiency level would be desirable. Information regarding participants' language-learning experiences, international contact and educational contexts should also be collected.

Such information would help researchers avoid interpreting differences between participants exclusively in terms of nationality or presumed cultural background.

4.2. Task Selection and Contextual Variables

The initial framework focuses on three communicative functions: requests, refusals and disagreements.

These functions were selected because each requires learners to coordinate communicative goals with interpersonal considerations and may involve negotiation when expectations differ.

The proposed assessment includes twelve elicitation tasks: six WDCT items and six interactive audio role-plays. Each communicative function is represented by two tasks in each modality.

Table 1. Proposed Assessment Blueprint

Communicative function	WDCT items	Audio role-plays	Primary assessment focus
Requests	2	2	Contextual appropriateness, mitigation, negotiation
Refusals	2	2	Relational management, explanation, alternative proposals
Disagreements	2	2	Stance expression, interpersonal sensitivity, negotiation
Total	6	6	Complementary pragmatic evidence

Task development should consider the interaction of social variables, including relative power, social distance and degree of imposition.

For example, one request scenario may involve two familiar university classmates, while another may involve communication with an unfamiliar person occupying an institutional role.

However, the framework does not assume that these variables operate identically across cultural contexts. Their interpretation should be examined during scenario development rather than assigned exclusively through researcher intuition.

An initial task pool should contain more scenarios than the final instrument requires, allowing poorly functioning or culturally ambiguous items to be revised or removed.

4.3. Written Discourse Completion Task Design

Each WDCT item should provide sufficient contextual information for participants to understand the communicative situation while avoiding excessive detail that restricts their available responses.

A proposed example is provided below.

Illustrative WDCT: Request

You are working on a university group assignment with a classmate whom you know only slightly. You need clarification about a section of the assignment that your classmate has already completed. Your classmate appears busy preparing for another deadline.

What would you say to ask for clarification? Write your response in English as naturally as possible.

This scenario is intended to elicit a request under conditions involving limited familiarity and potential competing demands.

The assessment would not assume that one particular opening, modal expression or degree of indirectness constitutes the only appropriate answer.

Instead, the response would be examined in relation to its communicative clarity, linguistic choices and sensitivity to the described circumstances.

4.4. Interactive Audio Role-Play Design

Audio role-plays should involve a trained interlocutor who follows a structured scenario while allowing limited, meaningful interactional variation.

Each task should begin with a standardised contextual prompt and continue through a short exchange. The interlocutor should have a set of anticipated follow-up prompts and conditional responses that preserve opportunities for negotiation without making the task wholly unpredictable.

Illustrative audio role-play: Refusal

Situation: A university student has been asked by an international classmate to help organise an event during a period when the student has significant academic commitments.

The interlocutor begins:

“Hi! We're organising an event this weekend, and I was wondering whether you could help us with the preparations.”

If the participant refuses, the interlocutor may continue with a standardised follow-up:

“That's understandable. Would you be able to help for just an hour?”

This second turn creates an opportunity to observe whether the participant maintains, modifies or negotiates the refusal.

Relevant evidence may include the ability to communicate a clear position, acknowledge the interlocutor's request and manage subsequent interaction without unnecessary misunderstanding.

These examples are illustrative materials developed for the present framework.

They have not undergone participant testing or empirical validation.

4.5. Administration and Counterbalancing

A future administration procedure should use clear instructions and a consistent task sequence.

Participants would first complete a background questionnaire and receive a brief orientation to both assessment modalities. They would then undertake written and oral tasks under standardised conditions.

To minimise practice and task-order effects, the study should use matched scenario sets and counterbalance modality order across participants.

The same scenario should not simply be repeated verbatim in both modalities for the same participant. Such repetition could allow responses produced in the first task to influence performance in the second.

Instead, two scenario sets should be constructed to be comparable in communicative function and contextual characteristics. Assignment of the scenario sets to written and oral modalities should be counterbalanced across participants.

Interlocutors should receive training in prompt delivery, permissible follow-up responses and the handling of unexpected participant contributions.

Audio recordings should be retained securely, subject to participant consent and relevant ethical approval.

The feasibility of the administration schedule, including task duration, participant fatigue and task-order effects, would need to be established through pilot testing.

5. Proposed Scoring Framework

5.1. Shared and Modality-Specific Criteria

The proposed framework distinguishes between dimensions observable in both modalities and those that require modality-specific evidence.

Three shared dimensions are proposed:

1. Sociopragmatic fit: The extent to which the response takes account of contextual and interpersonal circumstances.

2. Pragmalinguistic realisation: The effectiveness of linguistic resources used to express the intended communicative function.

3. Relational management: The extent to which the response recognises and manages relevant interpersonal concerns.

For written tasks, an additional criterion would examine discourse organisation and the clarity of the formulated response.

For audio role-plays, an additional criterion would examine interactional responsiveness, including response to subsequent turns, clarification and conversational repair.

Table 2. Proposed Analytic Rating Dimensions

Dimension	WDCT	Audio role-play
Sociopragmatic fit	Assessed	Assessed
Pragmalinguistic realisation	Assessed	Assessed
Relational management	Assessed	Assessed
Written discourse organisation	Assessed	Not applicable
Interactional responsiveness	Not applicable	Assessed

Each criterion could initially be rated on a four-point scale from 0 to 3.

A provisional general interpretation of the scale is:

Table 3. Provisional Four-Point Rating Scale

Score	General descriptor
0	Relevant communicative function is absent or cannot be interpreted
1	Limited demonstration, with substantial contextual or communicative difficulties
2	Generally effective demonstration, with some limitations
3	Effective, contextually responsive demonstration with clear communicative purpose

These descriptors should be treated only as an initial rating structure.

Before implementation, dimension-specific descriptors and annotated examples must be developed through expert consultation and pilot response analysis.

The shared dimensions may permit cautious comparison of performance patterns across modalities. However, modality-specific dimensions should be reported separately rather than combined into an unqualified total score.

Such separation is necessary because differences in spoken and written performance may reflect processing conditions, interactional demands or task characteristics rather than differences in a single underlying proficiency trait.

5.2. Rating Fairness

A central concern is how pragmatic appropriateness is defined.

A scoring system that automatically rewards high levels of indirectness, elaborate politeness markers or conformity to a single native-speaker model risks misrepresenting legitimate pragmatic variation.

For example, direct communication may be effective and appropriate in situations where interlocutors have established expectations of efficiency or familiarity. Conversely, linguistic elaboration does not necessarily indicate greater sensitivity to context.

Consequently, raters should consider whether a response is intelligible, interactionally defensible and responsive to the stated circumstances.

Rater training should include examples of pragmatically successful responses employing different communicative strategies.

Where feasible, the rating process should involve individuals with experience in different English-using and intercultural communication contexts. Disagreements among raters may provide useful information about the assumptions embedded within scoring criteria.

6. Validation Strategy

The proposed framework requires empirical validation before it can support substantive claims about learners' pragmatic competence.

Following an argument-based approach to validation (Kane, 2013), the process should examine the evidential links between task performance, observed scores and intended interpretations.

6.1. Content and Construct Representation

Initial validation should involve expert review of task specifications and scoring dimensions.

Experts in interlanguage pragmatics, intercultural communication and language assessment should examine whether the scenarios represent plausible communicative situations and whether the scoring criteria reflect the intended construct.

Particular attention should be paid to the cultural assumptions underlying scenario descriptions.

The expert review should also evaluate whether the proposed range of requests, refusals and disagreements provides adequate representation of the narrower construct targeted by the instrument.

It should not be assumed that three communicative functions comprehensively represent intercultural pragmatic competence.

6.2. Cognitive and Response-Process Evidence

A subsequent stage should examine how learners interpret task instructions and situational information.

Cognitive interviews and retrospective verbal reports could be used to investigate whether participants understand scenarios as intended, notice relevant contextual variables and employ the communicative reasoning anticipated by the task design.

These procedures would also help identify unintended sources of difficulty, such as unfamiliar institutional terminology or ambiguity in the participants' assigned roles.

For audio role-plays, response-process evidence should include attention to participants' interpretation of interlocutor prompts and to the consistency of interlocutor behaviour.

6.3. Pilot Testing and Reliability

A pilot study should be conducted before any larger-scale administration.

Pilot testing would allow researchers to identify unsuitable scenarios, excessive task demands and weaknesses in the provisional scoring rubric.

At least two trained raters should independently evaluate an appropriate subset of performances.

Inter-rater agreement should be examined using statistics suited to the scale and analysis design. For example, weighted agreement coefficients may be appropriate

for individual ordinal ratings, whereas intraclass correlation coefficients may be considered for certain aggregated scores when their assumptions are met.

Where sufficient data and an appropriate design are available, many-facet Rasch modelling could provide further evidence concerning participant ability, item difficulty and rater severity.

Nevertheless, reliability statistics alone would not establish validity. A consistently applied scoring rubric may still privilege inappropriate communicative norms or inadequately represent interactional competence.

6.4. Cross-Modality Evidence

One purpose of the integrated framework is to investigate whether written and oral tasks provide complementary rather than redundant information.

Future studies could examine associations among shared scoring dimensions, identify recurring patterns of performance and compare the kinds of pragmatic difficulties observable in each modality.

For instance, a learner may formulate a contextually appropriate refusal in a written task but encounter difficulty maintaining that refusal during a subsequent spoken turn.

Such a pattern would be consistent with differences in the processing and interactional demands of the two tasks, although alternative explanations, including scenario differences and task familiarity, would also need to be examined.

Because the proposed tasks are not psychometrically equated, numerical differences between modalities should not automatically be interpreted as gains or deficits in pragmatic competence.

6.5. External Evidence and Consequences

The validity argument should eventually consider how assessment results relate to communicative performance beyond the instrument itself.

With appropriate ethical safeguards, researchers could compare elicited task performance with independent evidence from authentic or more naturalistic communication.

Such evidence might include consenting participants' academic interactions, collaborative tasks or other communicative activities relevant to the intended assessment purpose.

The consequences of assessment use must also be examined.

If an instrument encourages teachers to view culturally different pragmatic strategies as deficiencies, it may undermine intercultural learning even when the instrument demonstrates acceptable statistical reliability.

The 2014 Standards for Educational and Psychological Testing emphasise the importance of evidence supporting intended test-score interpretations and uses. Accordingly, the educational value of the proposed framework should be considered

alongside its measurement properties.

7. Discussion

7.1. The Value of Integrating Written and Oral Evidence

The principal contribution of the proposed framework lies in its explicit differentiation of task affordances.

WDCTs and interactive audio role-plays generate different types of pragmatic evidence. The former permit the relatively controlled elicitation of formulated responses, whereas the latter introduce interactive demands that may reveal processes of negotiation and adjustment.

These differences are theoretically important because pragmatic competence involves more than knowing which linguistic expressions may be suitable in particular circumstances.

Pragmatic performance also depends on the capacity to interpret and respond to an interlocutor whose behaviour cannot be entirely anticipated.

By using both modalities, researchers may investigate not only learners' linguistic choices but also the relationship between contextual judgement and interactional performance.

However, the integrated approach should not be interpreted as an inherently superior assessment simply because it contains more tasks. Additional tasks can increase participant burden and introduce new sources of measurement error.

Its potential advantage depends on whether the additional modality supplies meaningful evidence that is insufficiently represented by the original assessment.

7.2. Standardisation and Communicative Authenticity

A persistent challenge concerns the tension between standardisation and interactional authenticity.

Highly standardised assessments facilitate consistency, but excessive control may restrict the variability through which pragmatic competence becomes observable.

Interactive tasks, by contrast, may provide richer evidence but can introduce differences in interlocutor conduct and conversational opportunities.

The proposed framework addresses this tension through structured but responsive interlocutor protocols.

Standardised opening prompts and controlled follow-up options can improve comparability, while limited opportunities for negotiation can preserve selected interactional features.

Nevertheless, this represents a design compromise rather than a complete solution.

The success of the compromise would need to be assessed by examining whether participants receive comparable opportunities to demonstrate interactional competence and whether interlocutors follow the protocol consistently.

7.3. Intercultural Agency and Assessment Norms

A further contribution concerns the relationship between pragmatic assessment and intercultural agency.

Traditional approaches to second language assessment may implicitly position learners as reproducing established target-language norms.

Yet English is used in communicative environments that are linguistically and culturally diverse. The suitability of pragmatic choices may therefore depend on the participants' negotiated expectations rather than a fixed national standard.

The proposed assessment recognises that learners may communicate effectively through different combinations of directness, explanation, mitigation and relational acknowledgement.

This does not imply that all responses are equally effective.

Some communicative choices may create ambiguity, fail to achieve an intended purpose or disregard information clearly established in the scenario. Such difficulties remain legitimate areas for assessment.

The challenge is to distinguish observable communicative problems from differences that arise because raters prefer particular cultural conventions.

A context-sensitive rubric, supported by diverse expert review and transparent rating rationales, may help address this distinction.

7.4. Pedagogical Implications

If empirically validated, the framework could support pragmatic instruction in university EFL programmes.

Teachers could use written scenarios to promote reflection on form–function relationships and situational expectations, while audio role-plays could create opportunities for practising spontaneous interactional responses.

Comparison between modalities might help identify learners who demonstrate relevant pragmatic knowledge but require additional support in applying that knowledge under interactional demands.

The assessment process could also encourage reflective discussion of intercultural variation.

For example, learners might examine several possible refusal strategies and discuss how the strategies may be interpreted differently according to interpersonal relationships, institutional expectations and communicative purposes.

Such activities could move pragmatic instruction beyond the memorisation of formulaic expressions towards more flexible communicative judgement.

However, these applications remain prospective. Their effectiveness would require investigation through appropriately designed classroom research.

8. Limitations and Future Research

Several limitations of the proposed framework must be acknowledged.

First, the article presents a methodological proposal rather than an empirically evaluated instrument. No claims can yet be made regarding its reliability, validity or comparative effectiveness.

Second, the focus on requests, refusals and disagreements captures only part of the broader domain of pragmatic competence. Other relevant abilities, such as interpreting implicatures, managing humour or navigating extended institutional discourse, are not directly assessed.

Third, simulated tasks cannot fully reproduce the social consequences, histories and contextual complexity of naturally occurring intercultural communication.

Fourth, audio-only assessment omits nonverbal resources that may be fundamental to interpersonal meaning-making.

Finally, the category of Chinese EFL learners encompasses considerable variation in linguistic experience, educational background and communicative orientation. Findings obtained from one institutional population should not automatically be generalised to all Chinese learners or to other EFL settings.

Future empirical research should therefore prioritise task validation, the investigation of response processes, evaluation of rater consistency and comparison with external communicative evidence.

Cross-contextual replication would also be valuable for determining which aspects of the framework are transferable beyond the population for which it was initially designed.

9. Conclusion

Assessing intercultural pragmatic competence requires careful attention to the relationship between linguistic knowledge, contextual judgement and interactional performance.

This article has proposed an integrated framework combining written discourse completion tasks and interactive audio role-plays for the assessment of Chinese university EFL learners.

The framework is based on the premise that different elicitation methods provide different forms of pragmatic evidence. Rather than treating written and oral performance as interchangeable, it distinguishes shared assessment dimensions from modality-specific features and proposes procedures for task development, rating and validation.

Its central methodological contribution is an explicit account of how complementary task evidence can be collected and interpreted without assuming that pragmatic competence consists of a fixed set of culturally prescribed expressions.

Although the framework remains to be empirically validated, it offers a basis for future research into the assessment of pragmatic performance in intercultural EFL communication.

More broadly, it highlights the importance of developing assessment practices that

are not only technically defensible but also responsive to the linguistic and cultural diversity of contemporary English use.

References

[1] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

[2] Baker, W. (2015). Research into practice: Cultural and intercultural awareness. Language Teaching, 48(1), 130–141. https://doi.org/10.1017/S0261444814000287

[3] Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. Journal of Studies in International Education, 10(3), 241–266. https://doi.org/10.1177/1028315306287002

[4] Félix-Brasdefer, J. C. (2007). Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish. Spanish in Context, 4(2), 159–185. https://doi.org/10.1075/sic.4.2.03fel

[5] Galaczi, E. D., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. Language Assessment Quarterly, 15(3), 219–236. https://doi.org/10.1080/15434303.2018.1453816

[6] Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73. https://doi.org/10.1111/jedm.12000

[7] Liddicoat, A. J., & Scarino, A. (2013). Intercultural language teaching and learning. Wiley-Blackwell. https://doi.org/10.1002/9781118482070

[8] Liu, J. (2007). Developing a pragmatics test for Chinese EFL learners. Language Testing, 24(3), 391–415. https://doi.org/10.1177/0265532207077206

[9] Lu, Q., Chen, Y., & Yang, L. (2026). Developing and validating a computerized dynamic diagnostic assessment of pragmatic competence for Chinese learners of English. Language Testing. Advance online publication. https://doi.org/10.1177/02655322261438154

[10] Roever, C. (2011). Testing of second language pragmatics: Past and future. Language Testing, 28(4), 463–481. https://doi.org/10.1177/0265532210394633

[11] Sasaki, M. (1998). Investigating EFL students' production of speech acts: A comparison of production questionnaires and role plays. Journal of Pragmatics, 30(4), 457–484. https://doi.org/10.1016/S0378-2166(98)00013-7

[12] Xu, L., & Wannaruk, A. (2016). Testing university learners' interlanguage pragmatic competence in a Chinese EFL context. PASAA, 52, 209–236. https://doi.org/10.58837/CHULA.PASAA.52.1.8

[13] Yuan, Y. (2001). An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral DCTs, field notes, and natural conversations. Journal of Pragmatics, 33(2), 271–292. https://doi.org/10.1016/S0378-2166(00)00031-X

[14] Zhang, L., Aubrey, S., & He, H. (2026). L2 pragmatic development of Chinese EFL learners at an EMI transnational university: A longitudinal study. System, 139, Article 104028. https://doi.org/10.1016/j.system.2026.104028