A Design-based Research Approach to Developing a Computerised Dynamic Assessment of Oral Pragmatic Competence: Solvable and Persistent Challenges

Allan Nicholas[1[0000−0002−6178−3171]], John Blake[1[0000−0002−3150−4995]], and Maxim Mozgovoy[1[M0000−0003−0684−4012]]

University of Aizu
anich@u-aizu.ac.jp

Abstract. This paper reports on a Design-Based Research (DBR) study investigating how human–AI interactions can be shaped to approximate the ways human assessors administer dynamic language assessments of pragmatic competence. While conversational AI is increasingly integrated into education, most applications remain text-based and overlook the pedagogical challenges of spoken interaction, particularly in pragmatics. Our project addresses this gap through the iterative design and evaluation of a computerised dynamic language assessment system focused on the speech act of requesting. Across six prototype cycles, users interacted with an AI interlocutor in spoken dialogues while an automated tutor agent monitored interactions for perceived instances of pragmatic inappropriateness and delivered graduated feedback. Analysis of iterative testing revealed distinct patterns in the types of issues encountered: surface-level problems in the graphical user interface (e.g., audio handling, text display, visuals) were relatively easy to resolve, whereas dialogue management, particularly phase disambiguation and parsing, proved more persistent and difficult to automate. These findings suggest that while technical refinements can steadily improve the usability of AI-mediated systems, replicating the subtle interactional and mediation strategies of human interlocutors remains a central challenge. By documenting these refinements, this study demonstrates how DBR can expose the boundary between tractable software improvements, prompt engineering fixes and deeper interactional complexities, providing practical insights for the design of AI-mediated spoken language learning environments.

Keywords: Design-Based Research · Dynamic assessment · User Experience · Usability · human–AI Interaction · Intelligent CALL.

1 Introduction

1.1 Motivation

Pragmatic competence is an important aspect of overall communicative competence [6], and can be defined as the ability to adapt one's linguistic realisations to the socio-contextual contexts in which they are embedded [18]. Failure to adhere to the expected pragmatic norms of the relevant community can lead to negative perceptions of the speaker, with potential social consequences [8].

While pragmatic competence is, therefore, important for L2 learners to develop alongside other formal aspects of the L2 (such as grammar or vocabulary), it is typically given relatively little time in the language classroom [21], perhaps due to an inherent degree of subjectivity in perceptions of pragmatic appropriateness. Studies investigating the teachability of L2 pragmatics have primarily focused on the effectiveness of implicit or explicit corrective feedback, with little research exploring alternative approaches to feedback, grounded in sociocultural theory [7, 11]. Further, assessments of L2 pragmatics have typically employed holistic scales for judging perceived pragmatic inappropriateness, lacking the specificity necessary to inform future classroom instruction [19, 23]. Additionally, while learners' needs may vary in relation to specific aspects of pragmatics they find challenging, addressing these needs can be difficult for teachers in large-group contexts [17, 13, 12].

Computerised dynamic assessments (C-DA) [15, 17, 12] offer one potential avenue for addressing the above issues, employing a graduated prompt approach to feedback in which feedback gradually increases in explicitness as needed, diagnosing specific aspects of learner performance amenable to future instruction, and widening access to both opportunities for L2 practice and for individualized feedback based on learner performance [15]. A C-DA administers a language task to a learner and provides feedback designed to simultaneously diagnose a learner's emerging abilities and promote their development [15, 12].

In recent years, a number of C-DA have been developed, primarily focusing on L2 reading or writing abilities [15, 25]. Few C-DA to date have focused on pragmatic competence [17, 13], and none on the pragmatic aspect of spoken interactions. While conversational AI has been widely adopted in text-based learning environments, spoken human–AI interaction focusing on pragmatic competence remains underexplored. This study therefore foregrounds the dual challenge of engineering a C-DA system that approximates the flexibility of human-administered DA, while distinguishing between issues that can be readily resolved through iterative refinement and those that persist despite successive prototypes.

1.2 Objectives and Research Questions

In this study, we investigate how a C-DA system can approximate the ways in which human assessors administer dynamic assessment of spoken pragmatic competence. Rather than focusing solely on differences between human–human and human–AI interactions, the study emphasises the types of challenges that arise when attempting to replicate human DA interactions and mediational strategies through iterative prototype design. A key objective is to distinguish between issues that can be readily resolved through technical refinement (e.g., graphical user interface adjustments, audio/text handling) and those that prove more persistent (e.g., phase disambiguation and parsing in dialogue management).

Through this distinction, the study aims to identify where human DA practices are most challenging to automate, and how design-based research (DBR) can guide refinements that bring AI systems closer to pedagogical goals. We aim to answer these research questions:

1. What types of issues arise when iteratively developing an AI-mediated C-DA of spoken pragmatic competence?
2. Which issues can be readily resolved through technical refinement, and which prove more persistent and challenging?
3. How do iterative cycles of design-based research contribute to aligning AI-mediated assessment with the diagnostic strategies employed by human assessors?

2 Dynamic Assessment and Computerised Dynamic Assessment

With the integration of natural language processing, CALL evolved into Intelligent CALL (iCALL), enabling more dynamic and adaptive learning experiences [3]. iCALL systems can analyse learner input, track errors, and provide targeted feedback, thus extending the scope of CALL beyond static materials. Many iCALL projects addressing spoken English have primarily concentrated on pronunciation training, often using speech recognition to detect phonological errors. Others have provided conversational practice through chatbots, simulating basic turn-taking and dialogue. Despite these advances, relatively few iCALL systems have focused on the teaching and assessment of pragmatics, leaving a gap in the development of learners' ability to adapt language use to social and contextual factors. Against this background, Dynamic Assessment (DA) offers a framework that directly addresses limitations in pragmatic assessment.

Two key issues regarding typical assessments of pragmatic competence can be identified: a focus on current ability levels rather than future potential and a lack of diagnostic capacity. [14, 15].

Traditional language assessments (henceforth referred to as non-dynamic assessments, or N-DA), including those assessing pragmatics, focus on independent learner task performance, in which the learner carries out a task unaided. From a sociocultural perspective on learning, such assessments provide useful information on a learner's current level of development; however, it provides limited insight into future potential development [22, 14]. Additionally, pragmatics-focused assessments to date have employed holistic scales to judge the overall pragmatic appropriateness of a learner's performance on a task [19, 23]. While this informs the teacher and learner of their overall level of pragmatic competence on a given task, it does not provide insights into specific aspects of the task the learner found challenging [13, 12].

DA aims to address these two issues. Grounded in a sociocultural-theoretical (SCT) perspective on learning [22], an important concept for a DA is that of the zone of proximal development (ZPD)[22], which can be defined in simple terms as the space between what a learner can achieve by themselves, and what they can achieve with assistance from an expert other [1]. From an SCT perspective, learner development is optimally promoted when assistance falls within this ZPD space [1, 16]. To this end, assistance, or mediation, should be both graded (the most implicit level of assistance given that allows the learner to proceed with the task) and contingent upon need [1].

In a DA, a learner and expert other (mediator) collaborate on a language task, with the mediator providing ZPD-sensitive assistance (mediation) when the learner encounters difficulties. In such instances the mediator and learner will engage in mediation, with the mediator initially providing highly implicit mediation, such as a hint. Should this not be successful in allowing the learner to proceed with the task, the mediator will gradually increase the level of explicitness, until the issue has been successfully resolved. By assessing the frequency and types of mediation engaged in, the mediator is able to diagnose the emerging abilities of the learner, and how close or far the learner is from being able to perform those abilities without assistance. In this way, a DA simultaneously aims to diagnose emerging abilities, and promote their development via mediation that is sensitive to the learner's ZPD.

DA methodology can be categorized into two general types – interactionist and interventionist [10]. With the former, the mediator and learner engage in unscripted mediation, with the mediator responding to the needs of the learner in a flexible manner. With the latter, mediation is standardized and scripted. While interactionist DA has the advantage of flexibility, interventionist DA may require less mediator training. Further, it is amenable to computerization [15].

DA methodology has been the subject of increasing focus in language learning in recent decades. Few, however, have focused on pragmatics [11, 7], with [11], for example, investigating the use of DA in relation to interactive L2 spoken performance of requesting.

While DA addresses a number of key issues relating to pragmatics assessment, it has been criticized for being resource-intensive, being primarily carried out in face-to-face sessions with individual learners [15, 12]. Computerised DA (C-DA) has the potential to widen access to DA methodology, allowing large numbers of learners access simultaneously. However, to date, there have been few pragmatics-related C-DA studies [17, 13, 11] and, to the best of our knowledge, no C-DA studies focusing on interactive spoken L2 performance of speech acts.

In order to address this need for a C-DA of interactive L2 spoken performance, the current study aims to develop a C-DA of L2 spoken requesting among Japanese learners of English at a computer science university in Japan. The C-DA administers a number of spoken requesting tasks, in which the learner interacts with an AI automated agent, with the program engaging in ZPD-sensitive mediation when the learner encounters difficulties. The C-DA follows interventionist DA methodology, with the standardised mediation being both graded and contingent [1].

3 Design-Based Research in Education

Design-Based Research (DBR) is an approach to educational inquiry that emphasises iterative cycles of design, enactment, analysis, and redesign within authentic learning environments [2, 4, 24]. It is guided by two key principles: grounding innovation in real-world practice, and generating both practical solutions and theoretical insights.

DBR typically involves close collaboration between researchers, educators, and learners, ensuring that designs are responsive to contextual needs and pedagogical goals. Unlike controlled laboratory experiments, DBR values ecological validity and acknowledges the complexity of educational settings. High ecological validity means that the conditions and tasks of a study closely resemble authentic learning environments, allowing findings to be generalised to real-world classrooms. In contrast, low ecological validity arises when tasks are highly artificial or detached from the context in which learning normally occurs, making it difficult to transfer findings beyond the experimental setting. By situating prototypes in authentic learning environments, DBR ensures that both the challenges and opportunities of real educational practice inform the design process. This methodology is particularly well suited to the refinement of educational technologies, as it allows systems to be continuously adapted in response to learner feedback and observed use. Each iteration not only improves the design but also contributes to a deeper understanding of how learners engage with the technology. In the context of pragmatics instruction and assessment with AI systems, DBR provides a systematic way to align technological functionality (e.g. speech recognition, dialogue flow, and feedback delivery) with pedagogical objectives. It ensures that system refinements are theory-informed and evidence-driven.

4 System Architecture

The C-DA system is built on a three-layer architecture comprising a dialogue layer, a pragmatic inappropriateness detection layer, and a feedback layer. This architecture reflects the goal of simulating how a human expert assessor administers dynamic assessment: guiding the dialogue, monitoring for perceived pragmatic inappropriateness, and providing developmentally-sensitive feedback. The layers interact in real time, allowing learners to engage in spoken request dialogues with an AI interlocutor while receiving feedback adapted to their performance.

The C-DA administers a set of four oral tasks to a learner, in which the learner is required to make a request to the AI interlocutor as part of a full conversation. Each task comprises a different scenario with varying Power (P; akin to relative social status), Social distance (D; the degree of familiarity between the interlocutors) and Rank of imposition (R; how potentially troublesome the request might be for the interlocutor to carry out) [5]. The learner produces their conversation turns orally, which is then shown on screen as text, allowing the learner to edit their turn before submitting it. The AI interlocutor turns are produced in both audio and text modes. Fig. 1 shows the current implementation of the Graphical User Interface (GUI). On the left is the focus image of the AI interlocutor, at the top is a progress bar showing the completed and current dialogue phase, the chat log is displayed in the centre, the AI tutor feedback is to its right while the microphone button and speech submission form are located at the bottom of the screen.

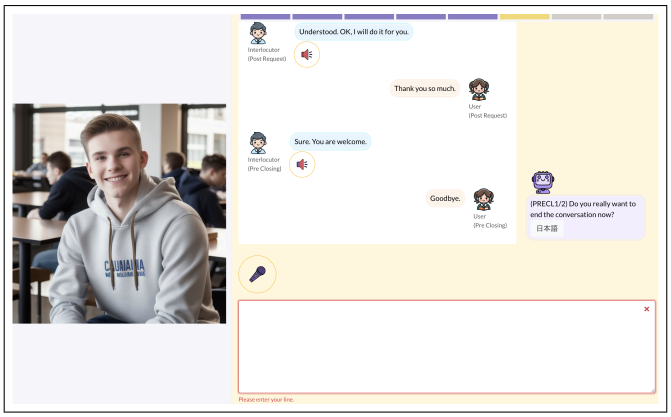

Fig. 1: C-DA User Interface.

For each task, the learner initiates the interaction, produces a request and closes the interaction. Conversation analysis literature has identified common patterns in the ways interlocutors typically co-construct request-based conversations (sequential organisation; [20]). These conversations commonly consist of an opening phase, a pre-request phase in which the requestor signals that a request is upcoming, the actual request, a post-request (such as an expression of gratitude), a pre-closing in which a space is created for ending the conversation, and the closing [20].

Fig. 2 shows this dialogue flow within the C-DA. Instances of pragmatic inappropriateness are identified by identifying the stage of dialogue the interaction is currently in (such as the closing), and identifying either the absence of obligatory pragmatic moves (such as a closing phrase to end the conversation), or the presence of an inappropriate expression. Required expressions or moves are those typically identified in request-based interactions in the literature, and are thus identified as being pragmatically expected.

When the C-DA identifies an instance of perceived pragmatic appropriateness, a feedback message is shown in text form.

Learner speech is recognised by Whisper ASR, chosen for its robustness in handling non-native English accents such as those of Japanese learners, while spoken output is generated through the Piper text-to-speech system. Transcription accuracy is critical, as errors at this stage directly affect the detection of

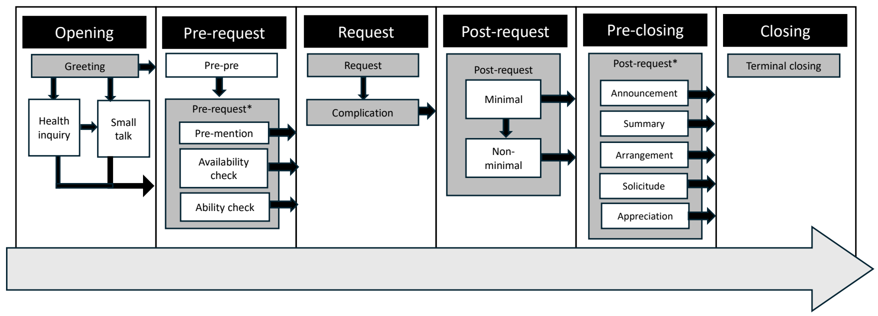

Fig. 2: Dialogue flow.

Note: Grey blocks are obligatory; White blocks are optional. When white blocks are within a grey block at least one must be completed. Key: *One or more in any combination.

pragmatic inappropriateness and the subsequent feedback. Iterative prototyping highlighted particular challenges in recognising reduced forms and disfluent speech, leading to refinements in noise filtering and language model selection. Conversation flow is managed by a large language model (LLM), specifically Google Gemma 3:12b, which was prompted to follow the sequential organisation of requesting as described in conversation analysis research. The LLM interlocutor is tasked with simulating realistic request-based dialogues, ensuring that the learner is pushed to perform target pragmatic moves. Initial versions of the system tended to fall into one of two extremes: they allowed learners to exploit loopholes in the dialogue, or they imposed overly rigid conversational structures. Refinements included constraining the range of system moves and explicitly modelling the core conversational phases, namely: opening, pre-request, request, post-request, pre-closing and closing.

The detection layer analyses each learner utterance for potential pragmatic inappropriateness. As in earlier written-request systems [13], instances of pragmatic inappropriateness are defined in terms of categories such as greeting, head act (the actual request turn in the conversation), and closing. For spoken interaction, these categories are adapted to capture issues such as phase disambiguation (e.g., unclear transition between opening and request), overly direct or indirect requests, and inappropriate lexical choices. Detection relies on a hybrid approach: surface-level pattern matching (e.g., absence of expected greeting tokens) combined with Part-of-speech-tag-based rules to identify more complex phenomena such as directness of requests. The system also incorporates a precedence hierarchy to avoid overlapping feedback; for instance, “greeting inappropriate” is not triggered if “greeting absent” has already been identified.

The feedback engine delivers responses according to a graduated prompt system [15]. On the first attempt, feedback is highly implicit, such as asking the learner to think again. The learner is given an opportunity to try the conversation turn again; should they fail to resolve the issue identified by the system, feedback becomes more explicit. There are four levels of feedback organised from implicit to explicit in total. Should the learner fail to resolve the issue after receiving level four feedback, the C-DA simply moves on, continuing the conversation. This graduated and contingent approach aims to optimally promote learner development within the learner's ZPD [1].

5 Method

To understand the extent to which iterations of prototypes improved the system, issues occurring in each testing phase were analyzed. Email communications between the research team regarding issues were collected and classified by prototype version. Template analysis [9], which stands midway between thematic and content analysis, was then used to categorise and classify the issues mentioned. The initial categories focused on the location of the issue while the final categories focus on the cause of the issue. Issues may be resolved by addressing the underlying cause or ameliorating their effect.

6 Results

Table 1 shows that a total of 68 issues were raised across the six prototype cycles. The most frequent categories were *Phase* and *GUI*, each accounting for approximately 45% of all reported issues, while problems with *AI conversation* were far less common.

Categories	**Prototypes**						**Total**
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	
AI conversation	2	1	1	0	0	1	5
GUI	9	8	7	3	0	4	31
Phase	5	9	5	0	9	4	32
Total	**16**	**18**	**13**	**3**	**9**	**9**	**68**

Table 1: Issues classified by category across six prototypes.

Table 2 breaks down the issues into eight subcategories, showing their distribution across the six prototypes. The most frequent subcategory was *Phase disambiguation*, which occurred 25 times (approximately 37% of all issues).

Subcategories	Prototypes						Total
	P1	P2	P3	P4	P5	P6	
Marked AI conversation	1	0	1	0	0	1	3
Conversation Parameters ignored	1	1	0	0	0	0	2
Audio in GUI	2	2	2	2	0	3	11
Behaviour in GUI	1	0	2	1	0	1	5
Text in GUI	5	5	1	0	0	0	11
Visuals in GUI	1	1	2	0	0	0	4
Phase disambiguation	5	8	5	0	5	2	25
Phase parsing	0	1	0	0	4	2	7
Total	**16**	**18**	**13**	**3**	**9**	**9**	**68**

Table 2: Issues classified by subcategory across six prototypes.

This reflects a recurring difficulty in identifying the learner's intended move within the sequential organisation of a request dialogue; for example, whether a learner utterance functioned as an opening, a pre-request, or the request itself. Such ambiguity often led to misclassification or conversational breakdowns, making this a persistent challenge.

Two GUI-related subcategories, *Audio in GUI* and *Text in GUI*, were the next most common, each occurring 11 times (16%). Audio-related issues included inconsistent playback, speech recognition delays, and unclear voice prompts. Text-related issues covered problems of legibility, formatting, and alignment that occasionally hindered learner interpretation of system feedback.

Less frequent GUI concerns nevertheless affected usability. *Behaviour in GUI* (5 issues) referred to unresponsive buttons or unexpected interface behaviour, while *Visuals in GUI* (4 issues) concerned layout, iconography, or visual consistency.

Interactional issues were less common overall. *Marked AI conversation* (3 issues) highlighted instances where the AI's responses sounded obviously artificial, reducing perceived authenticity. *Conversation parameters ignored* (2 issues) occurred when contextual factors such as power or social distance were omitted, producing pragmatically inappropriate replies.

Finally, *Phase parsing* accounted for 7 issues, showing that even when the system identified the correct phase, it often misanalysed the pragmatic function of learner utterances within that phase. In practice, accurate identification of pragmatic success or failure depends on situating each utterance within the appropriate conversational phase, since the appropriateness of a move can only be judged relative to its sequential context. To achieve this, the system attempts to track state changes across a series of compulsory phases, using specific triggers—such as the successful completion of a prior phase—to advance the dialogue state. *Small talk* is a notable exception: it is optional and can legitimately occur in limited positions within the dialogue structure, which increases the risk of misclassification. As a result, errors often arose not from isolated recognition mistakes, but from difficulties in dynamically tracking the learner's progression through sequential phases and correctly mapping utterances to the relevant pragmatic expectations.

Overall, GUI-related issues were frequent but generally solvable through incremental design refinements, while dialogue management issues—especially phase disambiguation—proved both common and persistent across iterations.

7 Discussion

The results demonstrate that issues emerged across both interface-related and dialogue-management dimensions. On the interface side, problems with *Audio in GUI* and *Text in GUI* were among the most frequent, together accounting for nearly one-third of all recorded issues. These included audio playback inconsistencies, difficulties in handling voice input/output, and text formatting or legibility concerns that affected learners' interpretation of feedback. Visual and behavioural aspects of the GUI also generated recurring but less frequent problems. On the dialogue side, *Phase disambiguation* and *Phase parsing* were the most salient, representing the system's difficulty in identifying the learner's conversational moves and interpreting them appropriately within the sequential structure of a request. Smaller categories such as *Marked AI conversation* and *Conversation parameters ignored* pointed to authenticity gaps, where the AI interlocutor failed to adhere to a task scenario's expected pragmatic norms. Taken together, the prototyping process revealed a broad landscape of challenges, ranging from surface-level usability issues to deep limitations in AI dialogue management.

The iterative design cycles showed a clear divide between solvable and persistent categories of issues. GUI-related problems were generally tractable. For instance, adjustments to font size, colour, and alignment improved text legibility, while modifications to button responsiveness and audio playback reduced user frustration. Such changes could be implemented quickly between prototypes, and once corrected, they rarely resurfaced. In contrast, dialogue management challenges, particularly *Phase disambiguation*, remained stubbornly persistent, accounting for over one-third of all issues even after multiple design iterations. Unlike GUI refinements, which could be addressed through straightforward coding or interface design, phase disambiguation required advances in natural language understanding, pragmatic inference, and turn-taking logic. Even with incremental improvements, the system continued to misclassify utterances, fail to recognise pre-requests, or mishandle ambiguous learner moves. Similarly, *Phase parsing* highlighted the limitations of pattern-matching approaches when confronted with non-typical learner language choices. These findings suggest that while technical refinements steadily improved surface-level usability, approximating the diagnostic subtlety of human assessors in real-time conversation remains an open challenge.

The DBR approach proved essential for making progress toward human-like C-DA interactions. By conducting multiple rounds of testing, we were able to identify which problems affected learner experience most acutely and which resisted easy solutions. Iterative refinements addressed GUI-related challenges that might otherwise have undermined learner engagement, thereby stabilising the foundation of the system. At the same time, repeated encounters with dialogue-management failures underscored the gap between human and AI capabilities. Human DA assessors routinely manage ambiguity, draw on paralinguistic cues, and flexibly negotiate meaning when administering a DA. In contrast, the AI interlocutor often faltered when faced with the same ambiguity, revealing where automation falls short of human practice. By systematically documenting these contrasts, DBR helped to map the boundary between what can be engineered with existing tools and what still requires more advanced AI modelling.

This study contributes a novel account of the integration of oral AI into a pragmatics-focused C-DA of L2 English, within a design-based research framework. By tracing six prototypes, it identifies which challenges in C-DA are solvable through technical refinement and which remain resistant, particularly in dialogue management. The work demonstrates how iterative prototyping can bring AI-mediated assessment closer to human practices, while also mapping the boundaries of current capabilities. Future research should build on these insights to develop more human-like, pedagogically aligned systems for pragmatic learning, as well as explore applications to other speech acts.

Acknowledgments. This study was funded by XXXXXX.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Aljaafreh, A., Lantolf, J.P.: Negative feedback as regulation and second language learning in the zone of proximal development. The Modern Language Journal **78**, 465–483 (1994)
2. Anderson, T., Shattuck, J.: Design-based research: A decade of progress in education research? Educational Researcher **41**(1), 16–25 (2012)
3. Blake, J.: Intelligent call: Individualizing learning using natural language generation. In: Tso, A., Ng, S., Law, L., Bai, T. (eds.) Annual conference of Hong Kong Association for Educational, Communications and Technology. pp. 3–18. Springer (2022)
4. Brown, A.L.: Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences **2**(2), 141–178 (1992)
5. Brown, P., Levinson, S.C.: Politeness: Some Universals in Language Use. Cambridge University Press (1987)
6. Celce-Murcia, M.: Rethinking the role of communicative competence in language teaching. In: Alcon-Soler, E., Safont, M.P. (eds.) Intercultural language use and language learning, pp. 41–57. Springer (2007)

7. van Compernolle, R.A., Kinginger, C.: Promoting metapragmatic development through assessment in the zone of proximal development. Language Teaching Research **17**(3), 282–302 (2013)
8. Economidou-Kogetsidis, M.: Teaching email politeness in the EFL/ESL classroom. ELT Journal **69**(4), 415–424 (2015)
9. King, N.: Template analysis. In: Symon, G., Cassell, C. (eds.) Qualitative methods and analysis in organizational research: A practical guide, pp. 118–134. Sage Publications Ltd (1998)
10. Lantolf, J.P., Poehner, M.: Dynamic assessment: Bringing the past into the future. Journal of Applied Linguistics **1**, 49–74 (2004)
11. Nicholas, A.: Dynamic assessment and requesting: Assessing the development of Japanese EFL learners' oral requesting performance interactively. Intercultural Pragmatics **17**(5), 545–575 (2020)
12. Nicholas, A., Blake, J.: Profiling learner development with a computerized dynamic assessment of a Japanese learner's L2 email writing. Research Methods in Applied Linguistics **3**(3), 100164 (2024). https://doi.org/https://doi.org/10.1016/j.rmal.2024.100164
13. Nicholas, A., Blake, J., Perkins, J., Mozgovoy, M.: Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests. Computer Assisted Language Learning pp. 1–33 (2024). https://doi.org/https://doi.org/10.1080/09588221.2024.2374775
14. Poehner, M.: Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting L2 Development. Springer Science & Business Media (2008)
15. Poehner, M., Zhang, J., Lu, X.: Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. Language Testing **32**(3), 337–357 (2015)
16. Poehner, M.E., Lantolf, J.P.: Sociocultural Theory and Second Language Developmental Education. Cambridge University Press (2024)
17. Qin, T., van Compernolle, R.A.: Computerized dynamic assessment of implicature comprehension in L2 Chinese. Language Learning & Technology **25**(2), 55–74 (2021)
18. Roever, C.: Teaching and testing second language pragmatics and interaction: A practical guide. Routledge (2022)
19. Savic, M.: Lecturer perceptions of im/politeness and in/appropriateness in student e-mail requests: A Norwegian perspective. Journal of Pragmatics **124**, 52–72 (2018)
20. Sidnell, J.: Conversation Analysis: An introduction. Wiley-Blackwell (2010)
21. Taguchi, N., Roever, C.: Second Language Pragmatics. Oxford University Press (2017)
22. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press (1978)
23. Winans, M.D.: Email requests: Politeness evaluations by instructors from diverse language backgrounds. Language Learning and Technology **24**(2), 104–118 (2020)
24. Xing, W., et al.: Development of a generative ai-powered teachable agent for secondary school using design-based research. British Journal of Educational Technology **56**(5), 2043–2077 (2024)
25. Yang, Y., Qian, D.D.: Promoting L2 English learners' reading proficiency through computerized dynamic assessment. Computer Assisted Language Learning **33**(5-6), 628–652 (2019). https://doi.org/https://doi.org/10.1080/09588221.2019.1585882