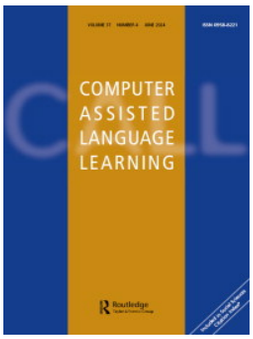

Computer Assisted Language Learning

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncal20

Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests

Allan Nicholas, John Blake, Jeremy Perkins & Maxim Mozgovoy

To cite this article: Allan Nicholas, John Blake, Jeremy Perkins & Maxim Mozgovoy (10 Jul 2024): Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests, Computer Assisted Language Learning, DOI: 10.1080/09588221.2024.2374775

To link to this article: https://doi.org/10.1080/09588221.2024.2374775

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2024.2374775

Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests

Allan Nicholas, John Blake, Jeremy Perkins and Maxim Mozgovoy

University of Aizu, Tsuruga, Ikki-machi, Aizu-wakamatsu, Fukushima, Japan

ABSTRACT

Automated writing evaluation can be effective in providing support for L2 English learners. However, little research to date has investigated its use in the teaching of pragmatics in relation to L2 email writing, grounded in a sociocultural perspective on learning. We employ a quasi-experimental approach, investigating the effectiveness of a computerised dynamic assessment programme (C-DA) of L2 email writing, focusing on pragmatic development. The C-DA employs a developmentally sensitive approach, providing automated, immediate mediation to learners. The C-DA was administered twice – with a two-week delay between administrations – to a single group of 17 Japanese English L2 learner participants at a Japanese university; participants were 19–22 years of age with intermediate L2 proficiency levels. Text, identified pragmatic failure and mediation data were elicited from both administrations. Analysis of pragmatic failure frequency and explicitness of the mediation necessary for successful resolution of the identified pragmatic failure shows both frequency and explicitness decreasing not only within a round of administration, but also between rounds. Evidence of learner development was found across all types of pragmatic failure, including the requesting head act, email openings and closings. The study provides evidence to support a sociocultural approach to assessment and learning with regards to the pragmatics aspect of L2 email writing, in which mediation is sensitive to individual learners' developmental needs. Further, findings support the use of a computerised approach to DA focusing on pragmatic competence, allowing for wider access to DA methodology among large learner group contexts.

ARTICLE HISTORY
Received 7 February 2024
Accepted 26 June 2024

KEYWORDS
Automated mediation; computerised dynamic assessment; L2 email writing; L2 pragmatics

Introduction

Automated Writing Evaluation (AWE) systems, which aim to provide corrective feedback (henceforth, we use the term *feedback* for non-sociocultural studies, and *mediation* for sociocultural-related

CONTACT Allan Nicholas anich@u-aizu.ac.jp University of Aizu, Tsuruga, Ikki-machi, Aizu-wakamatsu, Fukushima, 965-8580, Japan

contexts) simultaneously to large groups of learners, have garnered increased interest due to their potential to support second language (L2) learners in text creation, a process where the provision of corrective mediation is crucial for enhancing L2 English learner performance (Fu et al., 2024; Link et al, 2020). AWE allows teachers to focus on teaching 'higher order' aspects of writing, delegating mediation on more mechanistic aspects of texts to a computer-based system (Link et al., 2020).

The employment of AWE has been found to be effective in supporting student revision of text drafts (Li et al., 2015). In terms of leading to improvements in writing accuracy, studies have generally shown the use of AWE to have a positive effect (Li et al., 2015; Li et al., 2017). However, AWE studies to date have focused primarily on formal aspects of writing, such as grammar and vocabulary (Chapelle et al., 2015), and in relation to academic writing specifically (Cotos et al., 2017; Li et al, 2015). There has been little research in relation to email writing, or L2 email writing pragmatics, in which learners must adapt their language choices to social context. Additionally, there has been little exploration of situating AWE within the theoretical-developmental perspective of sociocultural theory (SCT; Vygosky, 1978).

SCT foregrounds the role social interaction plays in the learning process, with a learner's abilities appearing initially through collaboration with another, before being gradually internalised, in a shift from the 'intermental to the intramental' (Lantolf & Appel, 1994, p.11). The *zone of proximal development* (ZPD; Vygotsky, 1978) is a key concept here and can be defined for operational purposes as the space between what a learner can achieve via collaboration with another and what they can achieve independently (Vygotsky, 1978). In SCT, a collaborator can promote development by engaging in mediation – assistance – that falls within this space (Aljaafreh & Lantolf, 1994; Poehner, 2008). Dynamic assessments (DA) are grounded in this perspective on learning, employing mediation to both simultaneously assess a learner and promote their development via mediation that is sensitive to a learner's ZPD (Poehner, 2008; Qin & van Compernolle, 2021).

In this study, we investigate the effectiveness of a computerised dynamic assessment (C-DA) approach to AWE in regards to Japanese EFL learners' L2 English email writing. The C-DA interface presents learners with four sequential email tasks, with each task presenting a meaningful scenario in which the learner must carry out a request. The program elicits email texts from users for each task and uses pattern-matching algorithms to identify specific instances of pragmatic failure relating to email openings, closings and request directness. The C-DA offers four levels of mediation, ranging from highly implicit to highly explicit with the purpose of being sensitive to a learner's ZPD.

Following mediation, learners are able to revise their texts until either the maximum level of mediation has been reached, or the issue has been resolved. All interactions are stored in the database. In doing so, the program aims to promote improvement in pragmatic performance, and simultaneously allow for assessment, extending the employment of a C-DA approach to the pragmatics of L2 English email writing for the first time. Therefore, we contribute to the need for effective L2 pragmatics instruction and mediation in the language classroom generally (Roever, 2022; Taguchi & Roever, 2017); pragmatics instruction focusing on email writing specifically (Nguyen, 2018), and the frequent need to be able to deliver such mediation in an efficient manner (Poehner et al., 2015).

Pragmatics and L2 email

Email remains an important mode of computer-mediated communication (Chen, 2015) in academia, allowing students to request and receive mediation on assignments in a less pressured manner than face-to-face interactions (Economidou-Kogetsidis, 2011). Pragmatic competence is one aspect of overall communicative competence (Celce-Murcia, 2007), and can be seen as the intersection of the social contexts in which we communicate and the linguistic choices we make (Roever, 2022). Composing an email, the sender must take into consideration relevant social variables, and make appropriate language choices. Failure to adhere to pragmatic norms may lead to negative perceptions of the sender by the receiver (Economidou-Kogetsidis et al., 2020; Hendriks et al, 2023).

However, as a 'hybrid medium', being interactive like oral communication, yet asynchronous and in the written mode, composing pragmatically appropriate L2 emails can be challenging (Chen, 2015; Economidou-Kogetsidis, 2016; Nicholas et al., 2023), with unique pragmatic norms. Email compositions typically consist of two primary 'moves' – 'framing' moves and 'content' moves (Chen, 2015). The former refers to the opening and closing portions of email texts; openings typically consist of a greeting (*Dear…*), the name of the intended recipient, and – if appropriate – a self-introduction identifying the sender (Chen, 2015). Closing framing moves include a *pre-closing*, in which the sender signifies the upcoming end of the email (*I look forward to hearing from you…*), and a closing, consisting of complementary closing (*Best regards…*) and signature elements (Chen, 2015; Uso-Juan, 2022).

Content moves, on the other hand, refer to elements of the text that address the email's purpose, such as a request. These moves can be categorised in terms of *internal* and *external modification strategies* (Chen, 2015; Nicholas et al., 2023). For request-based emails, the former refer to strategies within the request head act – the portion of text that contains

the request – while external modification strategies occur outside of it, in the surrounding text. Both strategy types serve to mitigate the request's directness, indicating awareness of the potential imposition that the request may impose upon the receiver (Economidou-Kogetsidis, 2016; Nicholas et al., 2023). Internal strategies within the head act range from highly direct imperatives ('*send me the paper*') to conventionally indirect strategies ('*could you let me know what time it arrives?*') and want or need statements ('*I need you to get it to me by this afternoon*'). External modification strategies may appear before or after the head act, and may include *grounders* – reasons for the upcoming request – ('*I'm not feeling well today, so…*'), *preparators* that hint at the upcoming request ('*you know that assignment we have?*'), apologies ('*I'm really sorry to ask, but…*'), and others (see Figure 1 for a full scheme of strategies relevant to this study).

In choosing moves, a learner must take into account the social context in which the communication is taking place. Relevant factors include *power* (P), relating to the relative social status of the sender and receiver; *social distance* (D) – how well the sender and receiver know each other, and *rank of imposition* (R), accounting for the potential level of imposition of a request upon the receiver (Brown & Levinson, 1987).

Composing pragmatically appropriate L2 emails, therefore, is challenging for many learners. With requesting-focused emails, L2 English users have been found to differ from L1 English users in their use of head act request strategies, with requests frequently being more direct, and in their use of openings and closings (Economidou-Kogetsidis, 2011, Economidou-Kogetsidis, 2016). Such divergences from L1 community conventions may lead to negative social consequences for learners (Economidou-Kogetsidis, 2015; Hendriks et al., 2023; Savic, 2018).

Addressing this challenge, L2 pragmatics instruction and mediation has been found to be effective in leading to change in learner performance (Plonsky & Zhuang, 2019; Ren et al., 2022). There is a need for instruction and mediation in regards to L2 email writing in academic contexts (Nguyen, 2018), as without such guidance, pragmatic development may not take place (Rau & Rau, 2016). Research, however, has primarily focused on oral communication (Plonsky & Zhuang, 2019; Ren et al., 2022); with relatively little investigation of pragmatics instruction effectiveness on L2 email writing (Nguyen et al., 2019; Uso-Juan, 2022).

Of the pragmatics-focused L2 email writing research to date, learners have generally been found to be responsive to pedagogic intervention (Chen, 2015; Uso-Juan, 2022). Studies have primarily focused on *explicit* forms of pragmatics instruction and feedback, typically involving presenting learners with metapragmatic information relating to L1 norms and

the ways in which a learner can reflect those norms through their language choices (Ren et al., 2022). This may, for example, involve raising learner awareness of P, D and R factors, and the linguistic resources available when making a request. Implicit pragmatics instruction, however, typically does not present metapragmatic information; rather, learners may be encouraged to reflect on text examples, and ways in which they may reflect pragmatic norms (Nguyen et al., 2015; Ren et al., 2022).

Chen (2015), investigating the effects of explicit pragmatics instruction on request-based email writing of Taiwanese university learners, found instruction to lead to changes in performance. Participants were presented with information relating to openings, closings, request strategies, and were given practice opportunities. Pre- and post-instruction data were collected and assessed by raters on a 4-point scale for pragmatic appropriateness, with additional qualitative analysis. Results show framing moves to have responded to instruction, while content moves were less responsive, presumably due to their less formulaic nature (Chen, 2015).

Nguyen et al. (2015) investigated the effects of varying feedback types on Vietnamese EFL learners' request email production and ability to judge pragmatic appropriateness. A control group received no feedback; a second group received feedback suggesting corrections with no explanations provided, while a third group received metapragmatic information on pragmatically inappropriate aspects of their texts, but with no suggested corrections. Elicited email text data were collected, along with pragmatic appropriateness multiple choice questions. Results found both treatment groups to perform better than the control. In terms of email production, both treatment groups performed similarly; the metapragmatic group performed better in judging pragmatic appropriateness.

Nguyen et al. (2019) investigated the effects of combining explicit feedback with opportunities to revise email drafts on Vietnamese EFL learners' emails. A control group received feedback but no opportunities for revision; a second group received feedback with one opportunity for revision, while a third group received two cycles of feedback and revision opportunities. Pre- and post-treatment data were assessed holistically for pragmatic appropriateness, with results finding all three groups to perform similarly, regardless of revision opportunities.

Unlike the studies above, Uso-Juan (2022) investigated the effects of instruction on authentic request email data from students to university faculty. Pedagogic intervention involved participants analysing sample emails for typical language use and formality; encouragement to use various resources and tools as composition aids, and opportunities to practice. Results found both framing and content moves to change post-treatment, with framing moves increasing in appropriateness.

Evidence was also found of content moves being employed to reduce the directness of requests.

While the above studies shed light on the effectiveness of explicit pragmatics instruction and feedback on learner email requests, there is less insight into the role implicit forms may play; there is also little research investigating the effectiveness of a sociocultural, dynamic assessment approach, or the automatisation of such mediation.

Dynamic assessment and L2 pragmatics

There has been increasing interest in recent years in situating evaluation and mediation through the lens of SCT (Vygotsky, 1978). A key concept is the ZPD (Vygotsky, 1978), which can be defined as the space between what can be achieved with mediation (*other-regulation*), and what can be achieved independently (*self-regulation*; Aljaafreh & Lantolf, 1994). From this perspective, only assistance or mediation that falls within this space can reliably promote development (Poehner, 2008; Vygotsky, 1978), and thus should be *graded* and *contingent* upon need (Aljaafreh & Lantolf, 1994; Qin & van Compernolle, 2021), providing the minimal amount that allows a learner to resolve an issue.

Dynamic assessment (DA) employs this approach to mediation in order to unify instruction and assessment (Poehner, 2008), involving the provision of mediation to both promote and evaluate development, and is an attempt to move away from a focus on the isolated performance of an individual, towards an approach that takes into account this social, mediated view of learning. Lidz and Gindis (2003) identify three defining characteristics of DA: a view of development that recognises the potential for learner change via intervention; the importance of both assessment and learning, and a diagnostic function informing future instruction.

In a DA, a mediator and learner collaborate on a task. If a learner struggles, the interlocutors can initiate mediation; initially, highly implicit forms of mediation are given, with explicitness gradually increased as necessary, until the issue has been resolved. By providing ZPD-sensitive assistance, development is promoted. Also, by analysing the frequency and explicitness of the mediation, we can assess the extent to which a learner is close to self-regulation.

Lantolf and Poehner (2004) describe two categories of DA – interactionist and interventionist. With the former, mediation is unscripted, allowing the mediator to respond flexibly, in the moment. Here, the emphasis is typically oriented towards development, with less emphasis on reliability (Poehner, 2008). While, arguably, interactionist DA allows for greater sensitivity to a learner's ZPD, it is also time and labour-intensive,

limiting its applicability to some educational contexts. Interventionist DA, however, employs standardised, scripted mediation, systematically organised by explicitness. This allows for increased reliability and the ability to scale assessment by reducing the amount of mediator training required (Poehner, 2008; Qin & van Compernolle, 2021). To date, interactionist DA studies have entailed a human expert mediator, who is able to respond flexibly to learners in a face-to-face context; the non-standardised nature of this approach, however, makes it challenging to implement in computerised DA forms (C-DA; Poehner et al., 2015; Qin & van Compernolle, 2021). Interventionist DA, on the other hand, due to its standardised, scripted approach, is amenable for use in C-DAs, and has, thus been the DA form employed in C-DA studies to date (Poehner et al., 2015; Qin & van Compernolle, 2021). C-DAs also address the practical concern of efficiency (Poehner et al., 2015).

In developing a C-DA, Poehner et al. (2015) put forward a number of principles. Mediation should be *developmentally sensitive to a learner's ZPD*, with mediation graded and contingent (Aljaafreh & Lantolf, 1994). By examining mediation frequency and explicitness, conclusions can be drawn regarding learner proximity to other or self-regulation. Further, a C-DA should be *efficient*, administrable to large groups of learners simultaneously, while still retaining sensitivity to each learner's ZPD (Poehner et al., 2015; Qin & van Compernolle, 2021).

A third feature is *transfer,* or *transcendence*, in which a learner is challenged to extend their abilities beyond one particular task at hand, and apply their knowledge to a variety of tasks in differing contexts (Feuerstein et al., 1979). Both face-to-face DA and C-DA studies have typically incorporated transfer tasks (Ableeva, 2010; Nicholas, 2020; Poehner, 2008) into procedures. Poehner et al., (2015), in a C-DA of L2 Chinese reading and listening comprehension, incorporated transfer items into sets of multiple-choice items with the aim of evaluating participants' ability to adapt their knowledge to differing contexts. Similarly, Qin and van Compernolle (2021) incorporated transfer items into their C-DA of implicature comprehension, increasing the challenge due to their use of less conventional language, or less formulaic expressions.

Few L2 DA studies to date have focused on pragmatics. Van Compernolle and Kinginger (2013) applied an interactionist face-to-face DA approach to L2 French learners' use of *tu* and *vous* pronouns, and the influence of P, D and R factors on their spoken usage. Interactionist methodology was employed, aiming to promote participant understanding and ability and simultaneously assess their abilities, analysing mediation explicitness and frequency. The approach was found to effectively promote participants' understanding of the target concepts. Similarly, Nicholas (2020) employed an in-person interactionist DA approach to

the development of Japanese EFL learners' ability to produce spoken requests in L2 English. The researcher engaged in mediation to promote understanding of oral requesting, finding evidence of changes in both participant performance in response to mediation, and also of their increased understanding of the target concept's relationship with social context.

Recently, a number of studies have investigated C-DA, in which evaluation and mediation is automated, including those focusing on L2 Chinese listening comprehension (Zhang & Lu, 2019), and L2 English grammar among Taiwanese junior high school learners (Tang & Ma, 2023). However, there has been little research applying it to pragmatics. Qin and van Compernolle (2021) applied a C-DA approach, focusing on implicature comprehension (understanding intended meaning) in EFL learners. Employing multiple-choice items, participants received three levels of mediation if they did not initially answer a question correctly, ranging from highly implicit to explicit. Analysis assigned two scores – one for 'independent performance' (first attempts to answer items, before mediation), and another for mediated performance, with the number depending on the amount of mediation provided. This allowed for insights into participants' maturing abilities, signified by the degree of explicitness of mediation needed. Results found the participants to perform better with mediation than independently.

Research questions

In this study, we extend for the first time a C-DA approach to L2 email requesting pragmatics among Japanese EFL learners, investigating its effectiveness in leading to changes in learner performance. We address the following two questions:

> RQ1: Does the C-DA lead to a decrease in instances of pragmatic failure and a decrease in mediation explicitness within a single administrative round? Further:
>
> RQ2: Do we see such decreases over time? To address this, we administered the C-DA twice, with a two-week delay between administrations.

Material and methods

Participants

The study took place at a Japanese computer science university. Approximately, 40% of faculty members are non-Japanese, with English being the lingua franca on campus. The study's voluntary participants ($n = 17$; of the initial 20 participants, three were absent in the second

administration, and were thus excluded from analysis) were undergraduate Japanese L1 users majoring in computer science and engineering-related areas, and were 19–22 years of age. Test of English for International Communication (TOEIC) proficiency levels were in the 450–600 range.

Participants completed a pre-treatment survey regarding their English L2 email writing experiences (see Appendix A). Of the 20 respondents, 18 had prior experience writing English emails, primarily in academic contexts. Only eight had written English emails five times or more. Thirteen of the 20 considered their English email writing ability to be 'not good', with all respondents finding it challenging, frequently citing a lack of knowledge regarding email-specific conventions, and appropriate formality.

Requesting email tasks

Requesting scenarios were based on those developed by Nicholas et al. (2023) specifically to elicit L2 English email request text data. We focus on request-based email scenarios, being a common email type in student-faculty email exchanges, and challenging due to the need to attend to directness and formality (Economidou-Kogetsidis, 2015; Nguyen, 2018).

Following Liu (2007), scenarios were initially elicited from a sample of the student population ($n = 108$) via a questionnaire. Elicited scenarios were ranked by frequency, with the most frequent ones serving as templates for task creation. From this, eight task scenarios were created and used, and each assigned P, D and R values by the researchers. Expert English users then moderated the items for validity – ensuring they elicit the requesting act– efficiency of language and agreement on assigned P, D, and R values (see Appendix B for definitions). Selected scenarios include academic contexts; however, personal and business-related scenarios were also included in order to vary P, D, and R levels and adhere to the concept of *transfer* in C-DA methodology (Poehner et al., 2015) (Appendix C).

With regards to task order, a DA typically incorporates transfer items towards the end of an administration to evaluate participants' ability to apply knowledge to novel or more complex contexts. Poehner et al. (2015) employed multiple-choice items in a C-DA to assess reading and listening comprehension, with transfer items involving more linguistic complexity. Similarly, Qin and van Compernolle (2021), also using multiple-choice items to assess implicature comprehension, incorporated transfer items in the latter half of the C-DA with fewer conventionalised linguistic expressions to increase difficulty.

In the current study, however, determining the difficulty level of email tasks is challenging. While relatively simple multiple choice-type items

assessing comprehension can be varied in terms of linguistic complexity or use of conventionalised expressions, for example, we administer tasks that elicit whole email texts from participants, including both framing and content-type moves. Further, each task scenario is assigned varying P, D and R values, adding to task complexity. This makes it difficult to determine which scenarios participants may find particularly challenging. Previous studies have primarily focused on learner emails to university faculty, and have found participants to frequently employ relatively direct head act request formulations (Economidou-Kogetsidis, 2011, 2016), and to struggle with email openings and closings. These studies, however, do not compare email compositions with varying P, D and R values, as in the current study. Nicholas et al.'s (2023) investigation of email requests among Japanese L2 English learners found a high frequency of pragmatic failure across all email tasks with varying P, D and R values, and for all types of pragmatic failure – head act-related, openings and closings. In the current study, therefore, we did not distinguish between 'standard' tasks and 'transfer' tasks as such; rather, the challenge is varied in each task, with differing P, D and R combinations requiring learners to account for formality and directness in their language choices. See Table 1 for an overview of administered task scenarios.

Automated identification of pragmatic failure

The C-DA's automated pragmatic failure identification system is based on Nicholas et al.'s (2023) Japanese EFL L2 English request-based email corpus research, in which a coding scheme was developed for identifying specific instances of pragmatic failure in learner texts (see Figure 1). The scheme is based on the foundational work of Blum-Kulka and Olshtain (1984) and Economidou-Kogetsidis (2011, 2016), who adapted the former for use with email text data. Nicholas et al.'s (2023) coding scheme adapts the latter further, for identification of specific instances of pragmatic *failure*, rather than *features*. This allowed the annotators a high degree of specificity, tagging particular parts of text, compared with previous studies' use of questionnaires to elicit holistic perceptions of learner emails.

Table 1. Administered tasks for first and second round administrations.

	Recipient		Language choices		
Task colour	Round 1	Round 2	Power (P)	Social distance (D)	Rank of imposition (R)
Blue	Local business owner	Professor	+	+	+
Red	University administrative officer	Local business owner			–
Green	Friend	Friend	–	–	+
Yellow	Friend	Friend			–

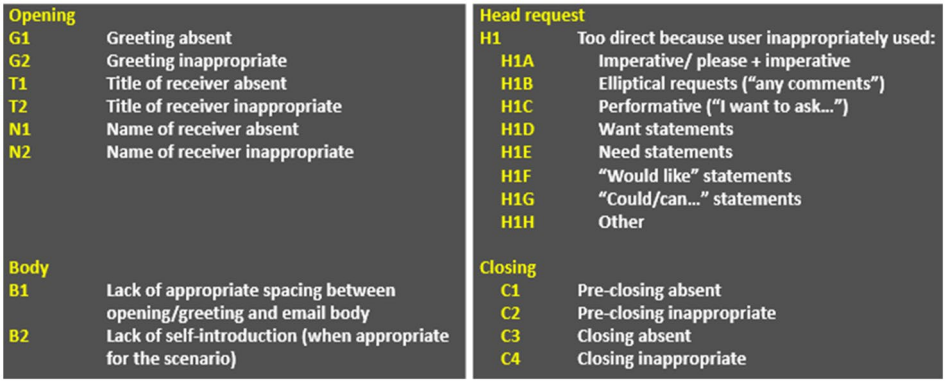

Figure 1. Coding scheme employed for pragmatic failure identification (Nicholas et al, 2023; adapted from Blum-Kulka & Olshtain, 1984; Economidou-Kogetsidis, 2011).

The coding scheme was applied to a corpus of approximately 1,300 elicited Japanese EFL learner L2 English email texts ($n = 426$). Following previous studies (Economidou-Kogetsidis, 2011; Savic, 2018), expert English users were recruited for manual annotation of specific perceived instances of pragmatic failure. To ensure intra- and inter-rater reliability, guidelines were created, a training course and benchmark testing implemented. Ten percent of the data were annotated by multiple annotators to ensure a reasonable degree of reliability. Inter-rater reliability checking followed the procedure set out by Campbell et al. (2013), in which an initial agreement check was followed by a negotiation phase in cases of disagreement. Following previous studies of perception data on pragmatic appropriateness (Economidou-Kogetsidis, 2016; Savic, 2018), annotators functioned as proxy email receivers, evaluating emails. The annotators were relevant members of the local community, being Japanese university lecturers. To further strengthen the relevance of the annotators' pragmatic appropriateness judgements to the study participants, other relevant community members – an administrator and a student at the institution – were recruited as specialist informants. A sample of the corpus data was shown to them, and their assessments of perceived instances of pragmatic failure elicited. This was then compared with the annotators', to ensure annotator coding aligned with the pragmatic norms of the study participant community.

The annotated corpus informed development of the C-DA's automated pragmatic failure detection system. Unlike Nicholas et al. (2023), however, the C-DA does not identify external modification strategies, as it was found during development that their general nature made it difficult for the system to identify appropriate strategies. The C-DA, therefore, limits identifiable content moves to internal modification strategies within the head act.

The C-DA was developed within the web framework Django (Dauzon et al., 2016), which runs on a web server. To identify instances of pragmatic failure, *regular expressions* – symbols and characters used to look for text patterns – were used to identify specific instances of failure. An iterative approach was used, cycling through evaluation and development stages, achieved through numerous releases and testing, enabling continuous incremental improvements. During evaluation, multiple trials and pilot tests were used to assess system accuracy and reliability, comparing system pragmatic failure identification with human annotators'. Generally, increases in accuracy were due to refinements in the matching power of the regular expressions, with a dual accuracy focus: (1) reducing the number of false positive results, and (2) increasing the number of true positives. Once the system accuracy threshold of approximately 90% was achieved in two consecutive tests, the software was deemed ready for use.

C-DA mediation

A C-DA typically aligns with the ZPD concept by providing mediation only when necessary (an issue is identified) and by providing graded mediation (Aljaafreh & Lantolf, 1994). Regarding the latter, this is operationalised by initially providing implicit forms of mediation, encouraging the learner to take on as much responsibility for carrying out the task as possible. Only if this does not allow the learner to resolve the issue does the system then gradually increase the level of explicitness until either the issue is successfully resolved, or the maximum level of explicitness has been reached. While previous C-DAs have employed three levels of mediation (Poehner et al., 2015; Qin & van Compernolle, 2021), we apply four, ranging from highly implicit to highly explicit, to each pragmatic failure type. For all categories of failure, we operationalise this in a similar manner, shown in Figure 2. Level one mediation (highly implicit) draws the learner's attention to the relevant portion of the text by highlighting it. Level 2 (implicit) – in addition to highlighting – alludes to the general nature of the issue; level three (explicit) is more specific about the issue, while level four (highly explicit) explains the issue, with possible resolutions. In this way, we aimed to strike a balance between allowing for sensitivity to a participant's ZPD, while at the same time seeking to avoid learner fatigue.

C-DA administration

Figure 3 shows the study phases. The C-DA was administered twice to an in-tact group of 20 voluntary participants in a L2 English language classroom setting, with a two-week delay between administrations (no

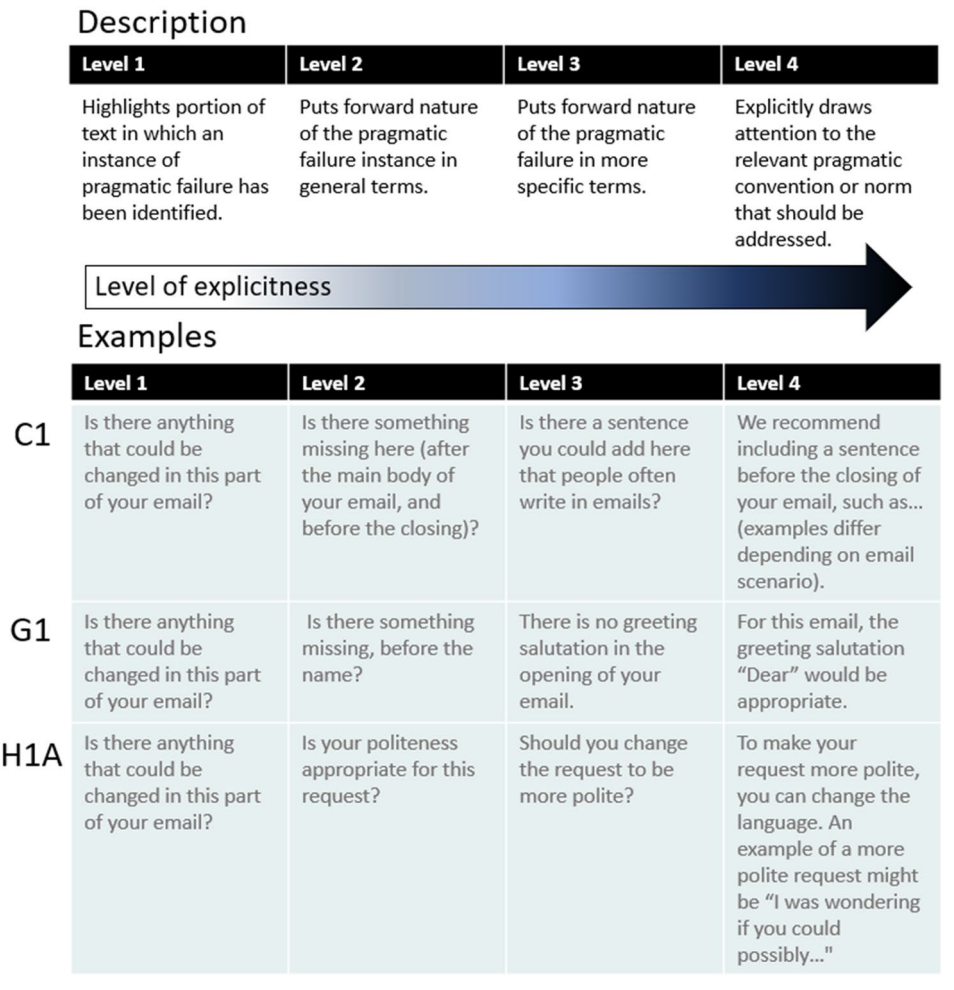

Figure 2. Descriptions and examples of the four mediation levels.

instruction given), to investigate sustained learner development across time. Of the 20 participants who completed Round 1C-DA, 17 also completed round 2; the three absentees were excluded from results.

The C-DA administered four email tasks in each round. Tasks were administered in the same order for all participants in both rounds, due to system limitations (see Appendix C for tasks and task order), with task order being decided randomly. Figure 4 shows a screenshot of the user interface. Participants were allowed five attempts for each task and received mediation on the first four attempts, with the level of explicitness increasing for each repeated failure instance.

Figure 5 shows the progression of participants within each round.

Participants initially registered and completed ethical approval procedures (see Appendix D), before reading the task scenario in both the L1 and L2 and submitting their attempt. Attempts were automatically

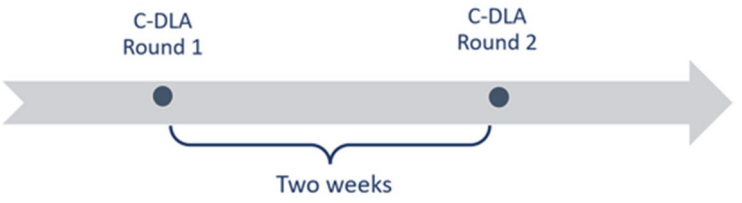

Figure 3. C-DA stages of administration.

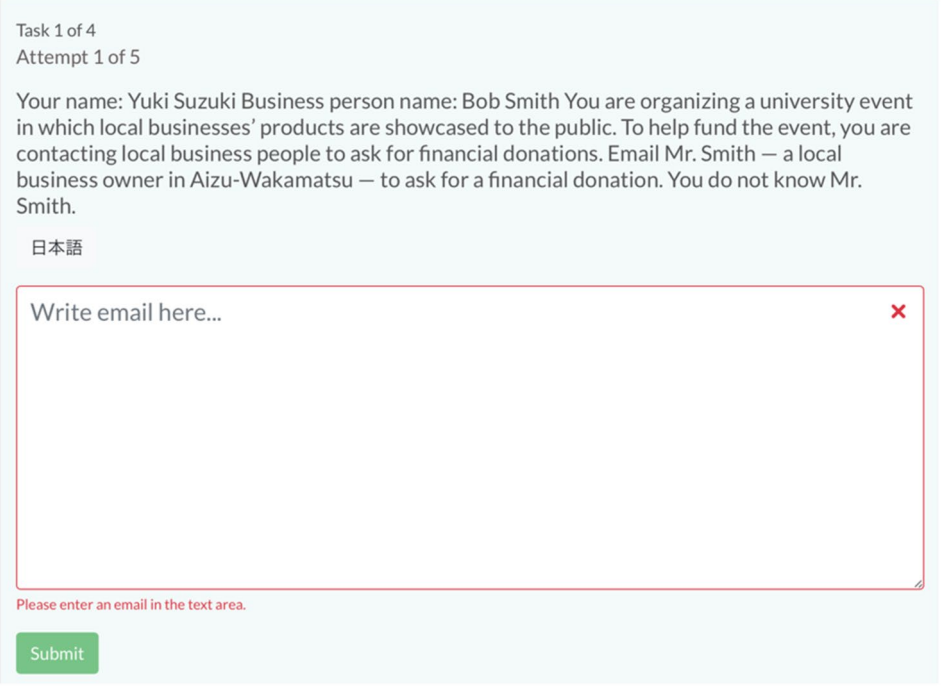

Figure 4. Example screenshot of C-DA task, as seen by participants.

checked; if failure was identified, the area of text was highlighted and mediation messages displayed. Level one mediation was shown upon submission of the first draft, after which participants were able to revise and resubmit. For each resubmission, the mediation level of explicitness increased until either the problem had been resolved, or the participant completed 5 submitted drafts. This process was repeated for all four tasks. Figure 6 shows an example screenshot displaying level 2 mediation for a submitted text.

Analysis

To test whether pragmatic failure frequency and mediation levels required to resolve a given instance of failure type decreased over time, mixed

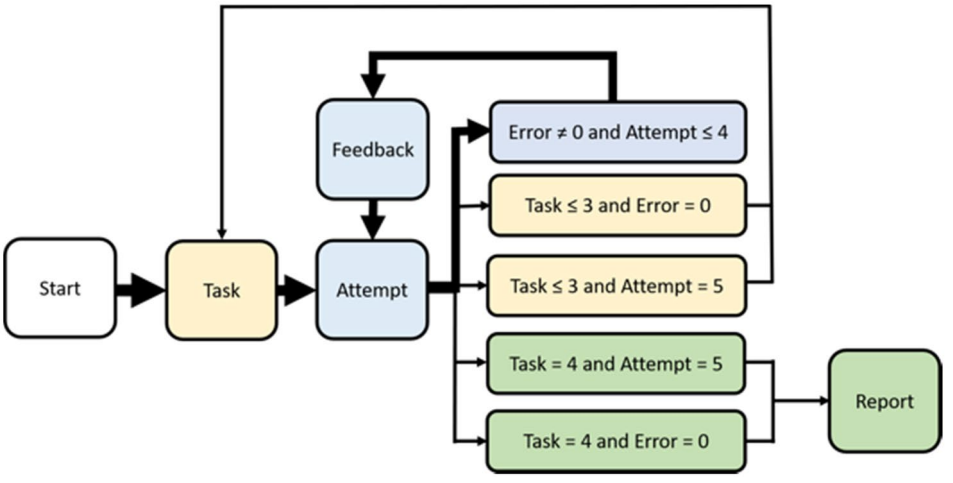

Figure 5. Diagrammatic representation of user actions and internal working of the C-DA.

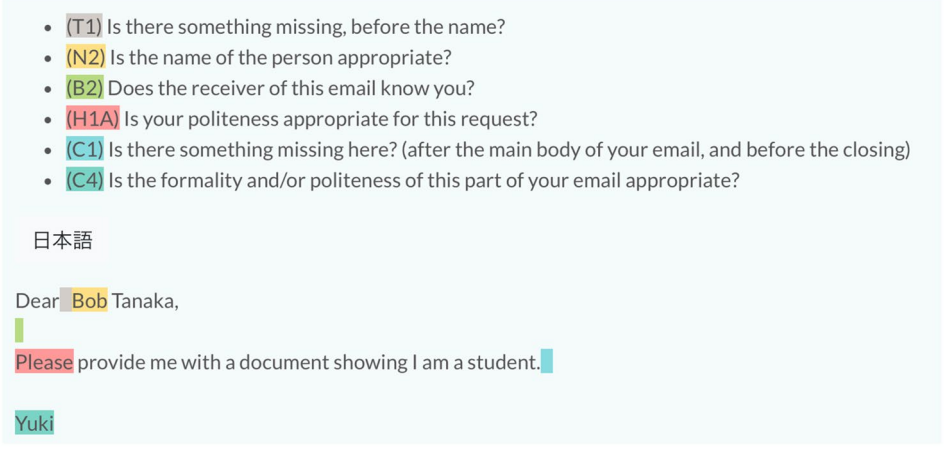

Figure 6. Demonstration screenshot of level 2 mediation shown to user. Note: L1 translation is available to users.

effects models with a cumulative link function were built using the ordinal package (Christensen, 2022) in R. The models set an ordinal factor for the number of instances of failure for a given participant in each task as the dependent variable. Because this dependent variable is ordinal and not continuous, a cumulative link function was required to properly perform a regression analysis. The number of instances of failure is equal to the mediation level, except that level-four instances could be split into two additional categories. Unresolved level-four pragmatic failure was coded with a total of five instances of failure and resolved level-four failure was treated as having four instances of failure.

An initial mixed model included ordinal within-subject factors for Pragmatic Failure Type (opening-type, head-act-related or closing-type), Round (1 or 2) and Task (blue, green, red and yellow). A random unique starting point, or intercept, was added for Participant to compensate for different participant ability levels. Next, to explore differences between pragmatic failure types, three separate models were built for each pragmatic failure type. Post-hoc tests were done using the Emmeans package in R (Lenth, 2023), which obtains the estimated marginal means for both linear and mixed models. The Akaike Information Criterion (AIC), which penalizes complexity to avoid overfitting (Akaike 1973), was used to balance the goodness-of-fit and simplicity of the models.

Results

The research question asked whether the participant group showed evidence of pragmatic development. To answer this, we analyze whether the *frequency of a given failure type* and the *mediation level required to resolve instances of failure* decrease in later tasks compared with earlier ones both *within* a single round and *between rounds*. Table 2 shows a summary of overall mediation levels and failure frequency by failure type in Rounds 1 and 2 for all participants combined.

Note that, while the C-DA was able to identify B-type instances of failure, their very low frequency in the data led us to exclude them from analysis. There is a clear difference between Round 1 and Round 2 in terms of both mediation level and instances of failure decreasing. Across all categories, Round 2 shows a decrease in comparison with Round 1. Head act-related instances of failure were the most frequent in participants' email texts, followed by closing and pre-closing type instances. Opening-related failure types were the least frequent in both rounds. An initial cumulative link mixed model that included a factor for these pragmatic failure types confirmed these observations (head act > closing: ES (effect size) = 1.53, Z-ratio = 7.447, $p < .001$; head act > opening: ES =

Table 2. Overall mediation levels and pragmatic failure frequency by pragmatic failure type.

Round	Item	Greetings (G1 and G2)	Titles (T1 and T2)	Names (N1 and N2)	Head acts (H)	Pre-closing (C1 and C2)	Closing (C3 and C4)	Total
1	Mediation	64	47	74	220	117	158	689
	Failure	66	49	79	245	126	177	752
2	Mediation	40	23	48	186	90	107	504
	Failure	42	24	52	206	95	121	551
Total	Mediation	104	70	74	406	207	265	1193
	Failure	108	73	131	451	221	298	1303

Note. In this Table, if a given failure type is identified in a participant's fifth, final email draft, after receiving the level 4 mediation, a failure count of 5 in total is recorded for that task.

2.99, Z-ratio = 8.327, $p < .001$; closing > opening: ES = 1.46, Z-ratio = 13.602, $p < .001$).

To assess whether the mediation level required to resolve pragmatic failure types decreased over time, cumulative-link mixed effects models were built. The results for each failure type, including opening, head-act-related and closing-types, as well as across all failure types, are discussed next.

All instances of pragmatic failure in aggregate

First, across all failure types, the number of instances of failure decreased in later tasks within rounds, and also decreased generally in Round 2, illustrated in Figure 7.

For Task, the model showed both a significant linear decrease in instances of failure in later tasks (α = −1.65, $p < .001$) and also a quadratic decrease (α = −0.72, $p < .001$), indicating that later tasks saw a larger learning effect than earlier ones. This is shown in the downward-shaped curves for both rounds. For Round, fewer instances of failure were made in Round 2, and the mixed model confirmed this *via*

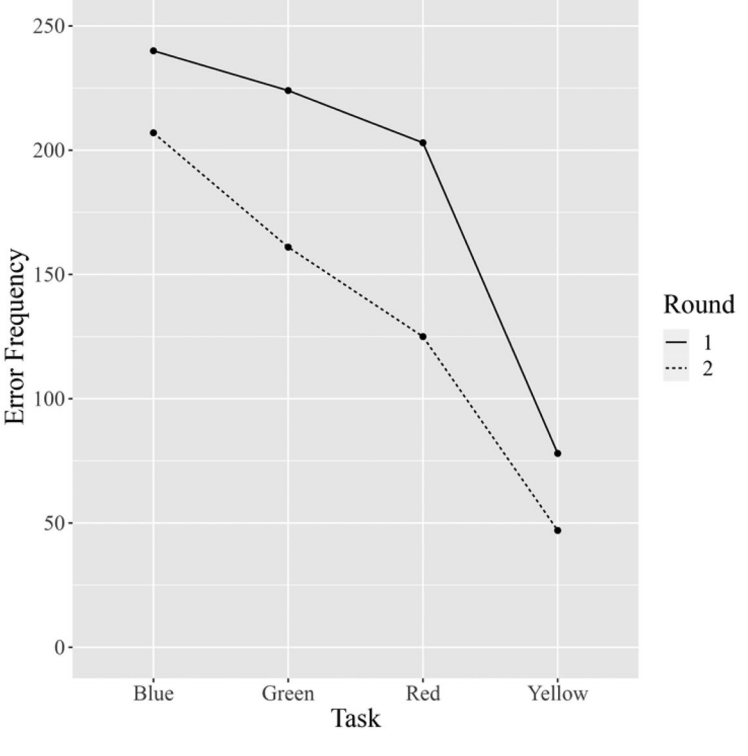

Figure 7. Pragmatic failure frequency by Round and Task across all failure types.

a significant negative linear trend (α = −0.47, $p < .001$, where α is the coefficient in the model for a given factor).

Next, we asked whether each successive task saw lower failure rates than the task preceding it, within each round. To answer this, a second regression model, mirroring the first but incorporating an interaction term for Task and Round, was created. However, the model with the interaction had a higher AIC value (1807) than the original model (1802), indicating that the increase in model fit was insufficient to warrant the increase in complexity in the second model. From this, we concluded that task effects did not differ between Rounds 1 and 2, and so the initial model without the interaction effect was adopted.

Post-hoc tests with adjustments using the Tukey method were carried out to check for pairwise significant differences in failure rate among the four tasks. The Yellow task had significantly fewer instances of failure than each of the three preceding tasks (blue: ES = 2.36, $p < .001$; red: ES = 1.75, $p = .001$; ES = 2.05, green: $p < .001$). In addition, the Red task had fewer instances of failure than the blue task (ES = 0.61, $p = .01$). However, no significant differences among failure rates were found between any other pairs of the blue, green and red tasks.

Opening-type pragmatic failure

Having established that overall pragmatic failure frequency significantly decreased both within and between rounds, we focus on instances of failure related to email openings. Among opening-type failure, the number of instances decreased in later tasks within rounds, as well as between Round 1 and 2, illustrated in Figure 8.

A linear decrease failure with later tasks was confirmed (α = −1.21, $p < .001$). Significant quadratic (α = −0.56, $p = .030$) and cubic (α = −0.50, $p = .028$) effects were also found. The quadratic effect indicates that later tasks saw larger decreases failure and the cubic effect may indicate that the number of instances in the Red task was relatively high, producing an inflection in the curve. Again, instances of failure were significantly lower in Round 2 (α = −0.47, $p = .005$), providing evidence of a general learner development effect.

Like the model across all failure types, there was no evidence of an interaction between Task and Round as a model containing the interaction had a higher AIC value (785) than the basic model (780). As such, the simple model was adopted and the task effect was analyzed independent of round. As with the case across all pragmatic failure types, the Yellow task saw lower failure rates than all other tasks (blue: ES = 1.85, $p < .001$; green: ES = 1.42, $p = .004$; red: ES = 1.56, $p = .001$). However, no other significant differences were found among the other tasks.

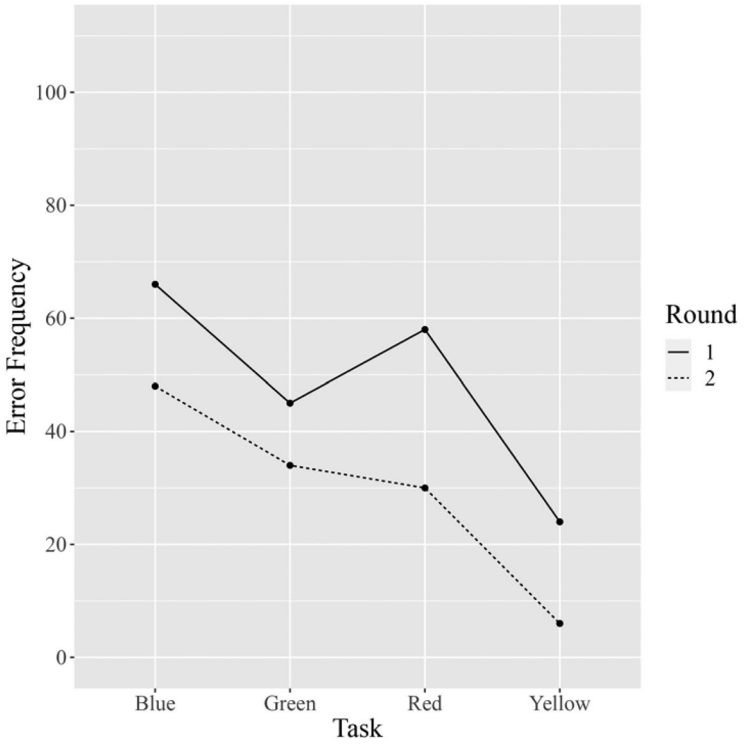

Figure 8. Pragmatic failure frequency by Round and Task among opening-type categories.

To demonstrate participant responsiveness to mediation, we provide qualitative excerpts from the email text data. Focusing specifically on openings, Excerpt 1 shows participant 10's Round 1 first attempt at the Blue task, requesting help from a local business person. G1 (absent greeting), T1 (title absent) and N2 codes (inappropriate name) were identified, leading to the provision of level 1 (highly implicit) mediation, in which the opening portion of text is highlighted and the participant asked to consider if they might want to change anything in the text. The second email draft in Excerpt 2 does not show any change to the opening, triggering level 2 mediation (implicit). For the G1 code, the level 2 mediation asks the participant to consider if something is missing, before the receiver's name. The participant responds to this in Excerpt 3 by revising 'To' to 'Hello'; the system does not consider this to be appropriate, and still identifies a G1 code in attempt 3, providing level 3 (explicit) mediation, informing the participant about the lack of a suitable greeting salutation. The participant responds in Excerpt 4, providing an appropriate greeting (*Dear*); however, T1 and N2 codes are still identified, triggering level 4 (highly explicit) mediation, providing appropriate titles and name conventions for the scenario. This leads to the participant's final

draft in Excerpt 5, in which the system does not identify any opening-related failure codes.

Participant 10, Round 1 Blue task

(1) Attempt 1; G1, T1, N2 pragmatic failure codes identified
To Bob Smith
(2) Attempt 2; G1, T1, N2
To Bob Smith
(3) Attempt 3; G1, T1, N2
Hello Bob Smith
(4) Attempt 4; T1, N2
Dear Bob Smith
(5) Attempt 5; no codes identified
Dear Mr. Smith

Excerpt 6 shows participant 10's first attempt at the Blue task in Round 2, in which the participant is required to email faculty, requesting help with a research project.

Participant 10, Round 2 Blue task

(6) Attempt 1; G1 and N2
Mr Bob Johnson
(7) Attempt 2; G1 and N2
Mr Bob Johnson
(8) Attempt 3; G1 and N2
Dear Mr Johnson

The system identified two instances of pragmatic failure in this opening – G1 and N2, prompting level 1 mediation. Excerpt 7 shows the participant's second draft, showing no response to mediation, triggering level 2 mediation. Excerpt 8 shows the participant's response; a greeting (*Dear)* is now present, and the receiver's first name is now absent. The system, therefore, did not identify any failure. Comparing Participant 10's performance in Round 1 with Round 2, therefore, shows evidence of some movement within their ZPD, requiring less explicit mediation to successfully resolve opening-related issues.

Head act-related pragmatic failure

Regarding head-act-related failure, Figure 9 shows a similar downward trend in failure rate in later tasks and administrations.

The logistic mixed model confirmed this with a significant negative linear trend for Task ($\alpha = -2.76$, $p < .001$). It also showed a significant negative quadratic trend ($\alpha = -1.36$, $p < .001$), confirming that failure rates decreased even more in later tasks. A downward linear trend for

Round was also confirmed (α = −0.63, $p = .018$), showing that fewer instances of failure occurred in Round 2.

There was no evidence for Task differing across rounds as a model that included an interaction between Task and Round had a higher AIC value (311) than the simple model (307). Post-hoc tests confirmed that the Yellow task had fewer instances of failure than the three preceding tasks (Blue: ES = 3.54, $p < .001$; Green: ES = 3.99, $p < .001$; Red: ES = 2.27, $p < .001$). In addition, this time, the red task had fewer instances than the green task (ES = 1.71, $p = .01$), but not the Blue task ($p = .07$). No other significant differences were discovered.

Excerpts 9, 10 and 11 show participant 15's requesting head acts in Rounds 1 and 2, respectively.

Participant 15, Round 1 Yellow task

(9) Attempt 1; H1A

So, please spend five minutes of talking with you in English.

(10) Attempt 5; no codes identified

So, could you spend five minutes of talking with you in English.

Participant 15, Round 2 Yellow task

(11) Attempt 1; no codes identified

Could you lend me the note from class.

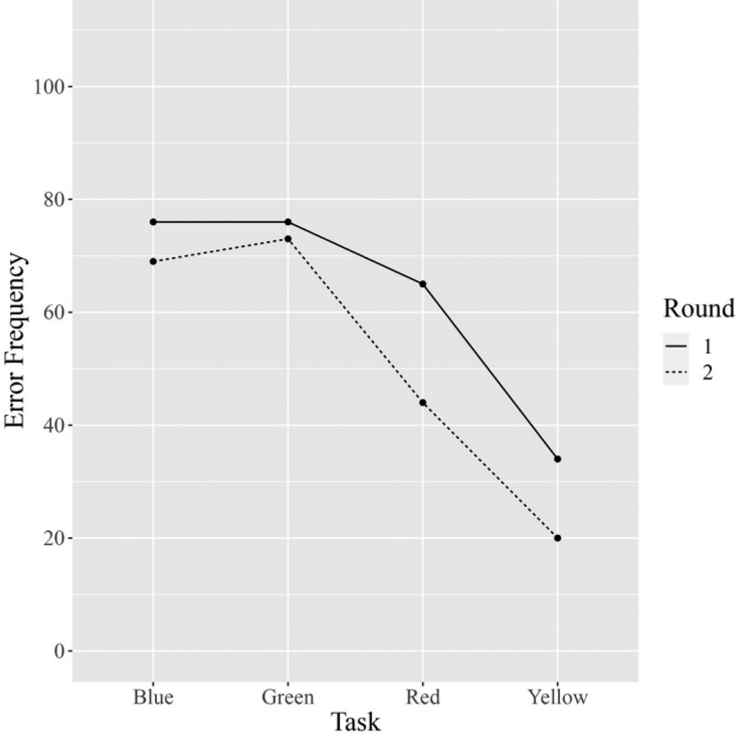

Figure 9. Pragmatic failure frequency by Round and Task among head-act-related instances of failure.

Excerpt 9 shows the Round 1 Yellow task first attempt, with the participant being required to request help with English conversation practice. The participant produces an imperative, triggering level 1 mediation. This does not lead, however, to the participant changing the head act; mediation levels 2 and 3 also do not lead to any changes. It is not until they receive highly explicit level 4 mediation that they successfully revise the request (Excerpt 10). In Round 2, however, Participant 15 does not require any mediation to produce a pragmatically appropriate head act for the Yellow task, employing the same requesting formulation as in Excerpt 10 (see Excerpt 11).

Closing-type pragmatic failure

Finally, considering closing-type pragmatic failure, Figure 10 shows fewer instances of failure being produced in later tasks and rounds.

The cumulative link mixed model confirmed both linear ($\alpha = -2.05$, $p < .001$) and quadratic ($\alpha = -0.75$, $p = .005$) decreases for failure frequency by Task. Again, the later tasks seem to produce even larger learning effects. Round 2 also had fewer instances of failure ($\alpha = -0.57$, $p = .002$).

Again, there was no evidence of differences by task across rounds with a model containing an interaction term having a higher AIC value (658) than the simple model (656). Post-hoc tests confirmed that the Yellow task had fewer instances of failure than the three preceding tasks (Blue: ES = 2.94, $p < .001$; Green: ES = 2.39, $p < .001$; Red: ES = 2.05, $p < .001$). In addition, the Red task had fewer instances of failure than the Blue task this time (ES = 0.88, $p = .045$).

Excerpts 12–15 show Participant 13's Round 1 Blue task attempts, focusing on closings. A C1 code indicates a missing pre-closing (*I look forward to hearing from you*, for example), while C4 refers to an inappropriate closing. Excerpt 13 shows the learner not responding to level 1 mediation; however, they do respond to level 2 mediation, shown in Excerpt 14. The closing is revised to *sincerely*, deemed by the system to be appropriate. To resolve the C1 code, however, level 3 mediation was required (excerpt 15).

Participant 13, Round 1 Blue task

(12) Attempt 1; C1 and C4

Regards,
Yuki Suzuki

(13) Attempt 2; C1 and C4

Regards,
Yuki Suzuki

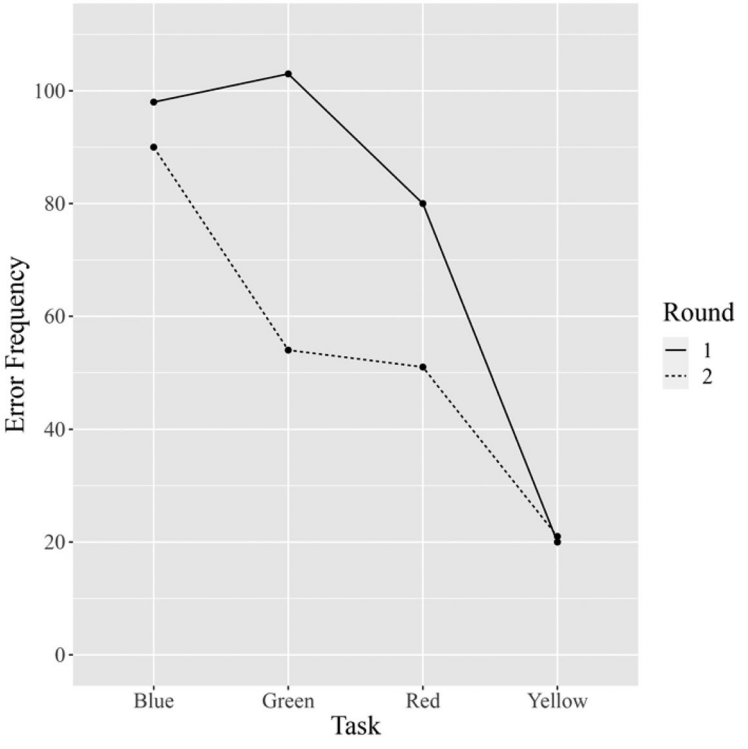

Figure 10. Pragmatic failure frequency by Round and Task among closing-type failure.

(14) Attempt 3; C1
Sincerely,
Yuki Suzuki

(15) Attempt 4; no codes
I am sorry to bother you but I appreciate it you send me email.
Sincerely,
Yuki Suzuki

In Round 2, Participant 13 composes a pragmatically appropriate email closing in the first attempt, with no codes identified (see Excerpt 16).

(16) Participant 13, Round 2 Blue task; attempt 1; no codes identified
I am sorry to bother you. I appreciate it you send email.
Sincerely,
Yuki Suzuki

Discussion

Analysis shows evidence of learner development both overall, and also within the specific categories of openings, the requesting head act and closing-related instances of pragmatic failure. We also note that overall

instances of failure decrease in Round 2 compared with Round 1 when comparing tasks of the same colour (the same combination of P, D and R values). The figures demonstrate that, for example, the frequency of pragmatic failure in the Round 2 Blue task was lower than in the Round 1 Blue task. This also holds true for particular categories of pragmatic failure, such as those relating to head acts. The one exception is seen in the Yellow task with regards to closing-type instances of pragmatic failure, in which frequency levels are relatively low in both rounds. This suggests that learner development is taking place across the range of task scenarios, and not just in tasks in which, for example, the R value is relatively low.

Figures 7 to 10 also show that the downward trend of failure frequency does not simply continue in Round 2 from the low point seen at the end of Round 1. Rather, following the two-week delay between rounds, failure frequency in early tasks in Round 2 is reset to some extent. However, it is not as high as it was at the beginning of Round 1, suggesting there was some sustained development among participants following the first round.

Regarding categories of failure relating to framing moves, results show overall frequency levels to be higher for pre-closing and closing type categories than for opening-related ones, with participants requiring more explicit levels of mediation to resolve those issues. This aligns with previous findings indicating closing-related moves to be more challenging for learners than openings, due to their less formulaic nature (Nguyen, 2018); this is particularly the case with pre-closings (Nicholas et al., 2023), which can be produced in a variety of ways.

That being said, significant decreases in frequency were found in relation to both opening-type and closing-related categories of pragmatic failure, within and between rounds. This aligns with previous research finding framing moves to be amenable to instruction (Chen, 2015; Uso-Juan, 2022). The significant effect of the programme on openings and closings may be due to the relatively formulaic nature of these elements in comparison with content moves (Chen, 2015). Despite pre-closings being challenging (Nicholas et al., 2023), this category of failure also proved amenable to mediation, with instances of failure and mediation levels decreasing across time.

While the overall trajectory for framing moves in Round 1 was towards fewer instances of failure, one nuance to note is the increase in opening-related failure in the third task (Red) relative to task two (Green). In this Red task scenario, the participant is required to request documents from an administrative officer of the university, with + P, +D and − R values assigned. In aggregate, the participants frequently struggled to produce appropriate email openings for this situation. This, we suggest, may be related to the transactional nature of the scenario, with

the receiver expected to grant the request as part of their work responsibilities (Nicholas et al., 2023). The fact that we do not see this increase in the Round 2 Red task (emailing a local business owner) lends support to this argument.

We also see a somewhat non-linear progression of responsiveness to mediation among closing-related failure in Round 1, with Figure 10 showing a slight increase in failure in the second task (Green; P–, D–, R+), relative to the first (Blue, P+, D+, R+). It is possible the participants initially struggled to adapt their language choices to the less formal expectations of the Green task. In contrast, we see a substantial decrease in failure in the third and fourth tasks, suggesting the participants' responsiveness to the explicit levels of mediation they had received earlier in the round. In Round 2, we see a more linear trajectory, with participants needing considerably less mediation in the Round 2 Green task relative to the corresponding task in Round 1, again suggesting development in participant understanding of the effect of context on language choices.

Interestingly, we see a significant decrease of pragmatic failure related to the requesting head act, within and between rounds. This is encouraging, as previous research found content moves less amenable to instruction than framing moves, due to being less formulaic and more 'idiosyncratic' (Chen, 2015). It is worth noting that, although a significant change was found, the frequency counts of head act-related failure were high in both rounds, relative to other categories. This supports previous findings relating to content moves, in which learners were found to struggle with formulating appropriate requesting head act strategies, relying on direct formulations to some extent (Chen, 2015; Economidou-Kogetsidis, 2016; Nicholas et al., 2023). Further, while Round 2 saw less content move-related pragmatic failure than in Round 1, the difference was less than with framing moves.

Although not a primary focus of this study, we note variation in responsiveness to different levels of mediation between framing and content moves. For opening and closing-related failure, responsiveness to the four levels of mediation was somewhat linear, with each successive level of mediation leading to a decrease in instances of failure. For head act-related failure, however, we see a more gradual decrease in failure until level four mediation, after which the failure frequency decreases considerably. This suggests that participants were overall less responsive to implicit mediation for content moves than framing moves. It is also worth noting that participants were frequently unable to resolve head act-related instances of failure even after level 4 mediation, indicating that the effectiveness of the mediation may need to be improved upon further to facilitate participant understanding.

From a sociocultural perspective, the variation in aggregate learner responsiveness to mediation provides insights into learner development for different aspects of email writing, with differing ZPDs for framing and content moves. Regarding head act-related failure, the frequent necessity for explicit mediation suggests the participants in aggregate are closer to other-regulation – in which they are reliant on mediation – than self-regulation, in which an ability has been internalised (Vygotsky, 1978). At the same time, we do see responsiveness to mediation regarding the head act, indicating progression towards the ability to perform independently. In contrast, we see less need for explicit mediation for opening-related failure, indicating participants may be closer to self-regulation.

As noted above, at times we also see a non-linear progression within rounds (e.g. with opening-related failure in the Round 1 second task versus task one). We suggest that this does not necessarily contradict the proposed effectiveness of the contingent and graded mediation provided by the C-DA. Rather, it may align with Vygotsky's observation that learner development may not always follow a consistent trajectory but may entail periods of seeming regression as the learner undergoes an internalisation process regarding the target ability (Vygotsky, 1978).

With regards to head act-related failure specifically, the relatively high frequencies of failure overall in both rounds suggest the possible limitations of the interventionist DA approach to mediation employed here. The fact that a number of participants were unable to successfully resolve issues even after highly explicit mediation suggests that a more flexible, interactionist approach may have been able to more successfully adapt mediation to the particular needs of each participant.

The limitations of the C-DA system in comparison with more traditional forms of in-person interactionist DA should also be noted in relation to ZPD conceptualisation. In an interactionist DA, the ZPD is typically conceived as being co-constructed by both the mediator and learner, with both interlocutors' contributions being important in this process (Poehner, 2008). On the other hand, the constraints of the interventionist C-DA system inevitably foregrounds the mediating moves of the C-DA system. While the C-DA responds to a learners' text production, it is unable to respond flexibly to learners' moves in the ways an interactionist DA can.

Regarding the high frequency of head act-related pragmatic failure, the limitations of the programme should be observed. Testing of the C-DA has found the programme's overall accuracy to be approximately 90% in appropriately identifying instances of failure, and for head act detection specifically, to be in the 85%–90% range. If we exclude unexpected L2 English usage from analysis, the accuracy rate increases to approximately

90%. However, typos, spelling mistakes or punctuation errors may lead to the programme not recognising a head act. Additionally, non-typical head act formulations that may nonetheless be viewed as pragmatically appropriate may not be recognised. This is due to the large range of possible non-typical formulations that may occur, which cannot be easily accounted for when programming the software. Therefore, it is possible that there were occasions in which the programme incorrectly identified a head act, thus leading to the participant being confused or unable to successfully resolve an issue.

An additional limitation of the study relates to possible task order effects. Within rounds, we see a significant decrease in failure in the later Yellow task, compared with earlier tasks. It is possible that we see this because the participants found the Yellow task's combination of P, D and R values to be relatively less challenging. However, it should be noted that we also see a significant decrease in the Red task in Round 2 as well, in comparison with earlier tasks. We also note that the general trend from task 1–4 in each round is downwards, suggesting a learning effect taking place. As mentioned above, we also see fewer instances of pragmatic failure when comparing same-colour tasks between rounds, providing evidence of development. Future studies may address this by assigning tasks in different orders to each participant.

Development was seen despite there being no additional treatment phase between Rounds 1 and 2. A complete version of the programme will include integration of further instruction, based on the individualised mediation a participant receives. Implementation of this, in combination with the automatised mediation, may further promote pragmatic development in learners.

Further, while the programme allows for the tracking of learner development across time, making visible movement within a participant's ZPD, the programme alone does not provide insight into the degree of conceptual understanding a participant has regarding the relationship between socio-contextual factors and language choices. Further insights here may be gained if the programme were combined with interviews, for example.

Conclusions

By analysing and tracking the frequency of pragmatic failure in participants' email writing texts, evidence is provided for the effectiveness of the C-DA in promoting learner development, and that this was sustained across time. We see evidence of the learner participants responding to the programme's mediation, shown in the decreasing frequency of pragmatic failure and decreasing levels of explicitness in the hints required to

successfully resolve failure. From a sociocultural perspective, this provides evidence of the participants moving from *other-regulation* in their email writing, towards a degree of *self-regulation*.

We put forward that this study addresses a number of key theoretical and practical issues for L2 pragmatics mediation and assessment. By employing a methodology grounded in sociocultural theory and interventionist DA, we provide evidence of the effectiveness of an alternative approach that avoids the bifurcation of mediation into implicit and explicit forms. The C-DA approach places primary importance on offering mediation that is sensitive to the ZPD of each individual learner. By initially providing highly implicit assistance and then gradually more explicit mediation contingent upon need, the C-DA aims to reliably promote learner development (Poehner, 2008). Being grounded in a theory of development, it also has the advantage of offering a systematic perspective on learner development, and allows for the tracking of development across time. This study's findings provide evidence for this approach's effectiveness in regards to pragmatics mediation. It also offers rich data for educators, as the program identifies specific instances of pragmatic failure in email writing for learners, and allows instructors to base further instruction on these insights.

Further, DAs in language learning are still relatively new, with relatively little research carried out into their efficacy. In particular, there have been few such assessments in relation to pragmatics, and those that have, have been conducted in in-person contexts (van Compernolle & Kinginger, 2013, Nicholas, 2020). To the best of our knowledge, this study is the first to focus on the pragmatics of L2 English email writing, and the first to employ an interventionist DA approach to the provision of mediation. Here, we provide evidence that a C-DA of email writing can be effective in promoting pragmatic development in learners, and that that development is sustained across time. This has positive implications for the language classroom, as such a program allows for large groups of learners to be administered a program that simultaneously promotes pragmatic development, and also assesses their written performance.

Future research will involve integrating additional instructional materials into the C-DA, tailored to the specific needs of individuals, based on C-DA task performances. DA studies typically incorporate an *enrichment programme* phase, in which learners' conceptual understanding of target abilities is developed through pedagogical approaches grounded in sociocultural theory (Poehner, 2008). Such a conceptual approach may be fruitful here as well, and would offer opportunities to garner insights into learners' understanding of the pragmatics aspects of English email writing. To this end, research incorporating participant interviews would be useful.

Additionally, a comparison of participants using the C-DA version of the programme with groups receiving explicit-only mediation and control groups receiving no mediation will allow for insight into the comparative effectiveness of the current approach. Also important is continuing work on improving the accuracy of the programme's detection of pragmatic failure, and in particular relation to head acts, which will further add to its effectiveness.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Japan Society for the Promotion of Science 19K00826.

Notes on contributors

Allan Nicholas is a senior associate professor at the Center for Language Research, University of Aizu, Japan. His research interests include sociocultural theory, dynamic assessment, computer-assisted language learning, and L2 pragmatics learning.

John Blake is a senior associate professor at the Center for Language Research, University of Aizu, Japan. His research interests include computer-assisted language learning, corpus linguistics, and second language acquisition.

Jeremy Perkins is a senior associate professor at the Center for Language Research, University of Aizu, Japan. His research interests include phonetics, second language acquisition and language learning.

Maxim Mozgovoy is a senior associate professor in the Department of Computer Science and Engineering, University of Aizu, Japan. His research interests include computer-assisted language learning, natural language processing and artificial intelligence.

ORCID

Allan Nicholas http://orcid.org/0000-0002-6178-3171
John Blake http://orcid.org/0000-0002-3150-4995
Jeremy Perkins http://orcid.org/0000-0002-7800-4234
Maxim Mozgovoy http://orcid.org/0000-0003-0684-4012

References

Ableeva, R. (2010). *Dynamic assessment of listening comprehension in second language learning*. Pennsylvania State University.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, *78*, 465–483. https://doi.org/10.2307/328585

Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies1: A Cross-Cultural Study of Speech Act Realization Patterns (CCSARP) 1. *Applied Linguistics*, *5*(3), 196–213. https://doi.org/10.1093/applin/5.3.196

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

Campbell, J. L., Quincy, C., Osserman, J., & Pederson, O. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. Alcon-Soler & M.P. Safont Jorda (Eds.), *Intercultural language use and language learning.* (pp. 41–57). Springer.

Chapelle, C. A., Cotos, E., & Lee, J. Y. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385–405. https://doi.org/10.1177/0265532214565386

Chen, Y. (2015). Developing Chinese EFL learners' email literacy through requests to faculty. *Journal of Pragmatics*, *75*, 131–149. https://doi.org/10.1016/j.pragma.2014.05.009

Christensen, R. H. B. (2022). Ordinal—Regression models for ordinal data. R package version 2022.11.16. https://CRAN.R-project.org/package=ordinal.

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genere learning. *Special Issue in Language Learning and Technology*, *21*(3), 104–130.

Dauzon, S., Bendoraitis, A., & Ravindran, A. (2016). *Django: Web development with Python*. Packt Publishing Ltd.

Economidou-Kogetsidis, M. (2011). "Please answer me as soon as possible": Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics*, *43*(13), 3193–3215. https://doi.org/10.1016/j.pragma.2011.06.006

Economidou-Kogetsidis, M. (2015). Teaching email politeness in the EFL/ESL classroom. *ELT Journal*, *69*(4), 415–424. https://doi.org/10.1093/elt/ccv031

Economidou-Kogetsidis, M. (2016). Variation in evaluations of the (im)politeness of emails from L2 learners and perceptions of the personality of their senders. *Journal of Pragmatics*, *106*, 1–19. https://doi.org/10.1016/j.pragma.2016.10.001

Economidou-Kogetsidis, M., Woodfield, H., & Savvidou, C. (2020). Non-native EFL teachers' email production and perceptions of e-(im)politeness. *Journal of Politeness Research*, *17*(2), 155–187. https://doi.org/10.1515/pr-2020-0046

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments and techniques*. University Park Press.

Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning*, *37*(1-2), 179–221. https://doi.org/10.1080/09588221.2022.2033787

Hendriks, B., van Meurs, F., & Kakisina, B. (2023). The effects of L1 and L2 writers' varying politeness modification in English emails on L1 and L2 readers. *Journal of Pragmatics*, *204*, 33–49. https://doi.org/10.1016/j.pragma.2022.12.006

Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics (Technical Report 7)*. University of Hawai'i.

Lantolf, J. P., & Appel, G. (1994). *Vygotskian approaches to second language research*. J. P. Lantolf & G. Appel (Eds.). Ablex Publishing.

Lantolf, J. P., & Poehner, M. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, *1*, 49–74. https://doi.org/10.1558/japl.1.1.49.55872

Lenth, R. V. (2023). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.5. https://CRAN.R-project.org/package=emmeans.

Lidz, C., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, Gindis B., Ageyev, V.S., & Miller, S.M. (Ed.), *Vygotsky's educational theory in cultural context* (pp. 99–116). Cambridge University Press.

Li, Z., Feng, H., & Saricaoglu, A. (2017). The short-term and long-term effects of AWE feedback on ESL students' development of grammatical accuracy. *CALICO Journal*, *34*(3), 355–375. https://doi.org/10.1558/cj.26382

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605–634. https://doi.org/10.1080/09588221.2020.1743323

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

Nguyen, T. (2018). Pragmatic development in the instructed context. *Pragmatics*, *28*(2), 217–252. https://doi.org/10.1075/prag.00007.ngu

Nguyen, T., Do, T., Nguyen, A., & Pham, T. (2015). Teaching email requests in the academic context: A focus on the role of corrective feedback. *Language Awareness*, *24*(2), 169–195. https://doi.org/10.1080/09658416.2015.1010543

Nguyen, T. T. M., Do, T. T. H., Pham, T. T. T., & Nguyen, T. A. (2019). The effects of corrective feedback with and without revision on enhancing L2 pragmatic performance. *Applied Pragmatics*, *1*(1), 1–25. https://doi.org/10.1075/ap.00001.min

Nicholas, A. (2020). Dynamic assessment and requesting: Assessing the development of Japanese EFL learners' oral requesting performance interactively. *Intercultural Pragmatics*, *17*(5), 545–575. https://doi.org/10.1515/ip-2020-5002

Nicholas, A., Blake, J., Mozgovoy, M., & Perkins, J. (2023). Investigating pragmatic failure in L2 English email writing among Japanese university EFL learners. *Register Studies*, *5*(1), 23–51. https://doi.org/10.1075/rs.20016.nic

Plonsky, L., & Zhuang, J. (2019). A meta-analysis of second language pragmatics instruction. In *Routledge handbook of SLA and pragmatics* (pp. 287–307). Routledge.

Poehner, M. (2008). *Dynamic assessment- A Vygotskian approach to understanding and promoting L2 development*. Springer Science & Business Media.

Poehner, M., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, *32*(3), 337–357. https://doi.org/10.1177/0265532214560390

Qin, T., & van Compernolle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology*, *25*(2), 55–74. http://hdl.handle.net/10125/73433.

Rau, D.-H, Rau, G. (2016). Negotiating personal relationship through email terms of address. In: Y. Chen, D.-H. Rau, & G. Ray (Eds.), *Email Discourse Among Chinese using English as a Lingua Franca* (pp.11–36). Springer. https://doi.org/10.1007/978-981-287-888-5_2

Ren, W., Shaofeng, L., & Xiaoxuan, L. (2022). A meta-analysis of the effectiveness of second language pragmatics instruction. *Applied Linguistics*, 1010–1029. *amac055* https://doi.org/10.1093/applin/amac055

Roever, C. (2022). *Teaching and testing second language pragmatics and interaction: A practical guide*. Routledge.

Savic, M. (2018). Lecturer perceptions of im/politeness and in/appropriateness in student e-mail requests: A Norwegian perspective. *Journal of Pragmatics*, *124*, 52–72. https://doi.org/10.1016/j.pragma.2017.12.005

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.

Tang, Y., & Ma, X. (2023). An interventionist dynamic assessment approach to college English writing in China. *Language Assessment Quarterly*, *20*(1), 44–65. https://doi.org/10.1558/lst.v2i2.25956

Uso-Juan, E. (2022). Exploring the role of strategy instruction on learners' ability to write authentic email requests to faculty. *Language Teaching Research*, *26*(2), 213–237. https://doi.org/10.1177/13621688211066997

van Compernolle, R. A., & Kinginger, C. (2013). Promoting metapragmatic development through assessment in the zone of proximal development. *Language Teaching Research*, *17*(3), 282–302. https://doi.org/10.1177/1362168813482917

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Zhang, J., & Lu, X. (2019). Measuring and supporting second language development using computerized dynamic assessment. *Language and Sociocultural Theory*, *6*(1), 92–115. https://doi.org/10.1558/lst.31710

Appendix A. Summarised results of participant survey, administered prior to round 1 C-DLA administration

Question	Responses					
	Yes	No				
At [this institution], have you ever written an email in English?	18	2				
Have you ever written an English email other than at [this institution]?	4	16				
How many times have you written an English email?	Never	Only once	2–3 times	3–5 times	5–10 times	More than 10 times
	2	1	4	5	2	6
Who have you written English emails to?	English-speaking professor	International student	University administration	Other		
	16	0	0	3		
How would you rate your English email writing ability?	Very good	Okay	Not good			
	1	6	13			
Do you find writing English emails	Easy	Medium	Difficult			
	0	14	6			

Appendix B. Definitions of power (P), social distance (D) and rank of imposition (R), with values (adapted from Hudson et al., 1995)

P	Meaning	D	Meaning	R	Meaning
+	Receiver has higher rank, title or social position.	+	Sender and receiver do not know, or identify with, each other.	+	Great expenditure of goods or energy by receiver to carry out request.
–	Receiver has lower rank or social position.	–	Sender or receiver know, or identify with, each other.	–	Small expenditure of goods or energy by receiver to carry out request.

Appendix C. Full task scenarios administered in round 1 and round 2 administrations

Task colour	Round	Scenario
Blue	1	You are organizing a university event in which local businesses' products are showcased to the public. To help fund the event, you are contacting local business people to ask for financial donations. Email Mr. Smith – a local business owner in Aizu-Wakamatsu – to ask for a financial donation. You do not know Mr. Smith.
	2	You are working on a research project as part of your university studies. Professor Johnson, who works at a different university, is an expert in the field. You do not know Professor Johnson. Email Professor Johnson to ask him/her to help you with analyzing your data and giving advice.
Green	1	You need to go to Sendai for an academic conference next week, but the train there is too expensive. You email your friend (who has a car) asking them to drive you there. It takes about 3 hours to drive from your home to Sendai by car. Your friend will be busy next week, so this will be inconvenient for him/her.
	2	You must buy an expensive textbook for a university course, but you do not have enough money to pay for it. Email your good friend (he/she lives in a different city to you) asking them to lend you the money.
Red	1	You must submit a document to the local government office in Aizu-Wakamatsu proving that you are a student at the University of Aizu. Email the manager of the Student Affairs Office at the university asking them to provide you with the document you need.
	2	You are organizing an event in which local businesses' products are showcased to the public. You would like to use a photograph (you already have the photograph) of one local businesses' product in the event flyer. Email Mr. Blair – the business owner – to ask for permission.
Yellow	1	You have a close friend who is an international student at university. You want to practice your English conversation skill next week, so you email him/her asking if they can spend 5 minutes of their time talking with you in English.
	2	You are taking a class in math at university, but you missed a recent lesson. Your classmate (a friend) has notes from the class. Email your classmate asking him/her to lend you their notes from that class.

Appendix D. Ethical procedures

The ethics procedures of the institution in which the study took place were fully complied with. Participants were provided with full details of the study, in both first and second languages, and voluntary informed consent was provided. Participants were able to withdraw their consent at any time, and were able to contact the researchers at any point. All data was kept securely in accordance with the ethics procedures of the institution.