Pragmatic feedback on refusals in a computer-simulated advising session

Paul Richards, Miyazaki Municipal University,
Department of Humanities

Abstract

This study experimentally investigated the effectiveness of feedback on learner refusals in a computer-simulated academic advising session. Ninety participants were assigned to one of three conditions: implicit feedback, explicit feedback, and comparison group. Oral and written discourse completion tasks (DCTs) were administered in a pretest immediate posttest design. Pragmatic development was identified by examining uptake, which was operationalized as incorporation of pragmatic features from the simulation that were absent on the pretests. This analysis, therefore, focuses on the subset of learners who did not make use of the target pragmatic features from the simulation on either pretest (n = 59). Participants in the explicit feedback group showed a significantly greater degree of uptake on the posttest ODCT than the comparison group, while both feedback groups showed significantly greater uptake than the comparison group on the posttest WDCT. The study also used a written retrospective comparison task to examine how learners described changes in their language use following treatment. Responses on this task indicated a greater degree of attention to (im)politeness and evidence of learning among participants in the feedback groups, while participants in the comparison group primarily described their responses as the same as before treatment.

Keywords: L2 Pragmatics, Corrective Feedback, Computer Simulations, Refusals

Language(s) Learned in This Study: English

APA Citation: Richards, P. (2024). Pragmatic feedback on refusals in a computer-simulated advising session. Language Learning & Technology, 28(1), 1–26. https://hdl.handle.net/10125/73549

Introduction

Second language (L2) pragmatics researchers have shown increasing interest in the use of games or simulations for pragmatics instruction as interactions in language classrooms are “limited on the variety of contexts and interlocutors that they can offer” (González-Lloret, 2019, p.115). In response to that, researchers have emphasized the potential for simulations to replicate interactions that can be difficult to recreate in language classrooms. Sydorenko et al. (2020) propose that simulations may prove useful for replicating “very specific pragmatic scenarios” (p. 49) where learners lack the knowledge necessary to perform specialized roles (e.g., medical contexts), and where it is implausible to arrange for learners to interact with relevant speakers through other forms of technology.

Simulations are also expected to ease various challenges surrounding the provision of pragmatic feedback. Holden & Sykes (2013) argue that with simulations it is possible to deliver feedback to learners at precisely the time they need it. However, this can be difficult in classroom contexts, as pragmatic features are often
dispersed throughout discourse, making them difficult to target discretely with verbal feedback (Kasper, 2001). Additionally, it is believed that pragmatic feedback, particularly feedback related to (im)politeness, may be less face-threatening to learners when delivered via computer. Since the earliest stages of the field, L2 pragmatics researchers have recognized that correcting elements of language that are intertwined with cultural and personal values poses greater risks of face-threat than feedback on grammar (Thomas, 1983). For example, language teachers even report being uncomfortable with giving adult learners feedback on issues of (im)politeness, as this feedback-providing practice can make them feel as if they are imposing their cultural values on learners (Takenoya, 2003).

Despite these expectations, little is currently known about the causal relationship between instructional features in simulations and pragmatic development, as most studies in this area have been exploratory in nature. This study aims to address this gap in the literature by experimentally investigating the effectiveness of implicit and explicit feedback delivered in a computer-simulated advising session.

**Instructed Pragmatics Research**

Numerous narrative reviews and meta-analyses demonstrate that pragmatics instruction is more effective than no instruction (Jeon & Kaya, 2006; Plonsky & Zhuang, 2019; Rose, 2005; Taguchi, 2015). In the earliest of these reviews, Rose (2005) categorized instructed pragmatics studies according to three design features: lack of a comparison group (teachability studies), use of a comparison group (treatment versus exposure studies), and use of multiple treatment groups and a comparison one (comparison studies). To date, the field has primarily targeted differences in explicit and implicit instruction. While findings from these studies suggest that explicit instruction is more beneficial than implicit instruction, implicit instruction can also be effective when it appropriately directs learners’ attention toward the instructional target (Taguchi, 2015).

Although the field has trended toward greater emphasis on comparative studies, Taguchi and Kim (2018) argue that greater variable control is needed in instructed pragmatics research, as they observe that previous work has exposed learners to multiple treatments, making it difficult to pinpoint the source of learning and generalize findings across studies. Framing this type of design as the “instructional package” (p.12) approach, Taguchi and Kim propose that previous studies have often implemented this design because they have worked in classroom contexts and therefore have prioritized learning potential over variable control.

**Corrective Feedback in Pragmatics Instruction**

To date, few studies have examined the use of corrective feedback (CF) in pragmatics instruction. The term CF refers to negative feedback given to learners about their language use. This contrasts with positive feedback, which affirms the accuracy of a learner’s utterance. CF is frequently examined in terms of whether it provides learners with input (e.g., recasts), whether it promotes learner output (e.g., elicitations), and whether the feedback is implicit or explicit. Here, explicitness refers to the degree of transparency of the corrective force of the feedback to the learner. Research on CF in language teaching generally shows CF to be more effective than no CF (Lyster & Saito, 2010). Research also suggests that explicit CF may be more effective than implicit CF (Yilmaz, 2012).

In a recent review of CF studies in pragmatics, Bardovi-Harlig and Yilmaz (2021) identified nine studies with research questions addressing the role of feedback in L2 pragmatics instruction. Of these studies, only one (Fukuya & Hill, 2006) investigated feedback as the primary method of instruction. In this study, Fukuya and Hill found that intonationally enhanced recasts led to greater use of bi-clausal request forms (e.g., Would it be possible to) than a comparison group that did not receive recasts. Each of the other studies compared the effects of instruction plus feedback. By controlling for the instructional variable and providing only one type of feedback per group, researchers (Barekat & Mehr, 2013; Takimoto, 2006) were able to determine whether CF provided learning gains beyond those of instruction alone. In these studies, participants were asked to rate a request in terms of appropriateness. Participants in the instruction + feedback group were told whether their ratings on the task matched the intuition of the researcher and were
also given an explanation. The findings from these studies, however, were mixed: the instruction + feedback group in Takimoto’s study did not outperform the instruction-only condition, but the instruction + feedback group in Barekat and Mehri’s study did. The remaining studies investigated combinations of instruction and feedback but did not isolate the effects of feedback. Together, the role of feedback in pragmatics instruction thus remains largely to be investigated.

**Computer Simulations in L2 Pragmatics Research**

The use of simulations for pragmatics instruction was first explored in a series of pioneering studies by Sykes (2009, 2013). Sykes investigated how Croquelandia, a 3D simulation of study abroad experiences in a Spanish-speaking country, influenced learners’ ability to request and apologize in Spanish. In the simulation, participants interacted with a variety of non-player characters (NPCs) by selecting text responses from a list of options and were given feedback when they selected nontarget responses. Feedback was operationalized as implicit, behavior-based feedback, which involved NPCs responding negatively to participants in the simulation. In addition to the behavioral feedback, participants were also exposed to a variety of other instructional variables, including explicit tips hidden in the simulation, opportunities for peer interaction, and classroom instruction. Further, participants were able to repeat interactions in the simulation as many times as they desired. Unfortunately, Sykes found that participants rarely repeated interactions in Croquelandia. Participants ultimately demonstrated minimal gains related to their production of requests (Sykes, 2009) and moderate gains related to their production of apologies on a written posttest task (Sykes, 2013). Sykes (2013) attributes this difference to the relatively more explicit feedback and tips given for apology scenarios in the simulation.

In a subsequent study, Holden and Sykes (2013) investigated the effects of a mobile, place-based simulation, Mentira, on learners’ pragmatic development. In Mentira, users attempted to solve a murder mystery by interacting with in-game characters. Participants were rewarded with clues when they selected a target response option in the simulation. Participants then needed to share these clues with other students to identify the culprit. To enhance the learners’ connection with the virtual environment, participants visited the setting of the simulation on a class trip. Additionally, to make the feedback more salient, Holden and Sykes programmed NPCs to reprimand users when they selected pragmatically inappropriate responses. Participants, however, showed only minimal gains following treatment.

Drawing inspiration from the work of Sykes (2009, 2013) and Holden and Sykes (2013), Taguchi et al. (2017) investigated the effects of a computer simulation on the learning of Mandarin conventional expressions. In this study, participants selected multiple-choice response options on a computer to respond to video clips of various commercial interactions. Feedback in the simulation was operationalized as leveling, points, and hints. Here, hints (typically described as CF) identified whether the participant’s selection was ungrammatical, inappropriate for the context, or unrelated to the prompt. Points—a type of positive feedback—were also awarded to participants when they selected the target option, with greater values given on higher levels. Participants later completed a more traditional computer learning activity that required them to type conventional expressions missing from a text. After that, participants were shown which of these responses were incorrect and were given an opportunity to make revisions. The authors found that participants were more accurate at writing and recognizing conventional expressions on the post- and delayed posttests.

Each of the studies discussed thus far fall into Rose’s (2005) category of teachability studies, as they investigated the effects of instruction on single groups of learners. Recently, studies on the use of simulations for pragmatics instruction have begun to compare learning outcomes between two groups of participants. For instance, Sydorenko et al. (2020) compared the learning outcomes between a simulation + implicit instruction group against a simulation + explicit instruction group. In the simulation, participants viewed video clips of a professor working in their office. Participants were told that they needed a letter of recommendation from the professor and were instructed to make a recording of this request. After doing this, participants identified the semantic content of their response from a list of options. These selections were then used to determine the next video to play for participants. When participants indicated that they
had recorded an inappropriate response, participants were given implicit oral feedback from the professor in the video. Following this, participants completed the simulation from the perspective of the professor, which served to provide participants with the input of successful request sequences performed by a student.

After completing the simulation, participants in the implicit instruction group viewed two videos of request role-plays involving a student and professor and engaged in awareness-raising activities; participants in the explicit counterpart viewed a video presentation on features of English requests and a video that provided generalized, explicit feedback on frequent issues learners have with making requests. Sydorenko et al. (2020) report subtle differences between groups on the posttests, noting that the implicit instruction group showed slightly greater use of vocabulary from the simulation, while the explicit group showed greater use of pragmatic forms. Both groups, however, were found to perform similarly overall after instruction.

Most recently, Tang and Taguchi (2021) compared the learning of Mandarin conventional expressions in the simulation Questaraunt, which incorporated several gaming features (narrative, characters, and the inclusion of a minigame), against a group using a minimalistic computer interface without gaming elements. Participants in the gaming condition played as a robot that needed to learn conventional expressions appropriate for working in a restaurant, while participants in the non-gaming group were not presented with a narrative and were shown stock images rather than animated characters. Participants in both groups completed the same interactions by selecting responses from an on-screen menu and were given positive or negative feedback depending on whether they selected the target response option. The negative feedback explained whether the response was ungrammatical, grammatical but unconventional, or inappropriate to the context. Feedback in the gaming condition was also enhanced with sound effects and character animations to reinforce whether a selection was correct or incorrect. The study, however, found no significant differences between these groups in terms of production or recognition of conventional expressions on post- or delayed posttests, indicating that the gaming elements did not contribute to learning beyond that of the feedback provided in the more simplistic, non-gaming condition.

In summary, each of the previous studies on simulations for pragmatics instruction has incorporated feedback as an instructional feature, but none have attempted to isolate the effects of feedback. Tang and Taguchi’s (2021) study is also the only one where learners were not exposed to instructional variables external to the simulation (i.e., classroom instruction or more conventional computer-based exercises). Additionally, except for Sydorenko et al. (2020), previous research has relied on written/text-based assessments to identify learning from computer simulations.

**Academic Advising Sessions**

This study investigates the effectiveness of feedback on refusals in a computer-simulated advising session. Advising sessions were chosen because they are goal-oriented tasks that all undergraduates at the host institution must complete to enroll in courses. The use of a computer simulation was deemed appropriate for this study because: (a) learners are generally unable to observe others in advising sessions (Bardovi-Harlig & Hartford, 1996), (b) previous research indicates that advisors may respond to the intent of the learner without providing feedback on language use (Bardovi-Harlig & Hartford, 1996), and (c) international undergraduate students would typically be unfamiliar with the institutional requirements of the role of the advisor, thus making roleplays a challenge to implement in the classroom. As suggested by Sydorenko et al. (2020), it is assumed in the present study that computer simulations may prove particularly useful in situations where opportunities for naturalistic input and feedback are limited, and in cases where learners may be unfamiliar with the duties of one of the roles and where that role is not one that the learners need to perform (i.e., that of the advisor).

**Assessing Pragmatic Development**

To demonstrate a causal relationship between instruction and learning outcomes, L2 pragmatics researchers increasingly use a battery of assessments and multi-level analyses to measure learning. Sydorenko and Tuason (2016), for instance, identified development through uptake, which they defined as the first-time use of a pragmatic feature from treatment that was absent from a participant’s pretest production.
Particularly, they examined changes in terms of semantic formulas (or pragmatic strategies), the building blocks that make up speech acts (see Fraser, 1980; Olshtain & Cohen, 1983) and lexical and syntactic forms used in realizing semantic formulas, such as the bi-clausal request form in the current study. The authors then compared changes in relation to notes taken by participants during the study and retrospective interviews with participants. In doing this, Sydorenko and Tuason were able to convincingly demonstrate the nature of development from the instruction provided in the study. Further, by identifying development in terms of an expansion of the pragmatic repertoire, they were able to measure learning in a manner consistent with a view of L2 pragmatics as “the study of the development of alternatives” (Bardovi-Harlig, 2017, p. 230).

**Purpose of the Present Study**

The current study compares the effectiveness of implicit and explicit feedback delivered in a computer-simulated advising session. Although feedback has been included as an instructional feature of each of the game/simulation studies conducted to date, no study has attempted to isolate the effects of feedback. The implicit feedback in this study targeted the overall politeness of the response selected by the participant in the simulation, and the explicit feedback addressed both the overall politeness of the response and the forms that contributed to the advisor’s assessment of the response as impolite.

This study used both oral and written discourse completion tasks (DCTs) to determine whether learning from a text-based simulation is evident orally and under time pressure, or if gains are only evident on an untimed written posttest. Previous studies on the use of simulations in pragmatics instruction have primarily relied on text-based assessments for measuring development, although oral discourse completion tasks (ODCTs) have been shown to be more representative of natural language use (Yuan, 2001).

A retrospective comparison task was included to investigate how learners evaluated changes in their language following treatment. Responses to the comparison task were then examined in relation to the language used by participants on the pretests and posttests and their selections in the simulation. The following research questions are addressed in the present study:

1. Do implicit and explicit feedback lead to differential degrees of uptake (i.e., first-time use of a feature from the treatment following instruction) on spoken assessment tasks?
2. Do implicit and explicit feedback lead to differential degrees of uptake (i.e., first-time use of a feature from the treatment following instruction) on written assessment tasks?
3. How do participants describe differences in their pretest and posttest performance following treatment?

**Method**

**Procedure**

Participants completed elicitation and treatment tasks via computer in a single 90-minute session. Participants were randomly assigned to one of three groups: implicit feedback, explicit feedback, and comparison. Participants completed each of the following tasks in the order presented: pretest ODCT, pretest written discourse completion task (WDCT), an advising simulation (completed three times), posttest ODCT, posttest WDCT, and a retrospective comparison task where participants explained differences between their pretest and posttest WDCT responses. Finally, participants completed a short background questionnaire on their experiences with language learning and academic advising. The procedure is summarized in Figure 1.
Figure 1

Procedure of Current Study

Oral and Written Discourse Completion Tasks

An ODCT and WDCT were used to examine the effectiveness of the feedback conditions on participants’ production of refusals. Both the ODCT and WDCT consisted of 10 refusal prompts, 10 request/course nomination prompts, and 10 fillers. Items were presented randomly on each test and for each participant. This study only addresses results for refusal items (see Appendix A).

The DCTs in this study made use of visual prompts to elicit the desired speech act and to avoid priming participants linguistically. The visual prompt used to elicit refusals was a frowning face (Figure 2). Prior to starting the pretest DCT, participants completed practice items to ensure that they understood how to complete the task. On the ODCT, participants responded to aural prompts. After the prompt was played, a microphone image appeared on screen indicating that the participant should record a response. Participants were given seven seconds before the program moved to the next item.

On the WDCT, participants typed responses to text prompts. WDCT items were untimed. Responses on the WDCT pretest were then used to deliver semi-adaptive response options and feedback in the simulation. Additionally, by including a WDCT, it was possible to allow participants to compare their pretest and posttest responses on the computer immediately after completing the posttests.

Figure 2

ODCT Refusal Item Prompt

You are speaking with an academic advisor

This is how you feel about taking a music class semester.

Participant hears:
“O.K., so I usually recommend students to take a music class.”
Advising Simulation

The advising simulation created for this study was based on a corpus of 51 advising session roleplays conducted as part of a previous study. The first of these was conducted between an academic advisor and an NS student at the host institution, and the remaining 50 were conducted between the researcher and international students. The simulation involved a student negotiating a course schedule with their advisor. During the simulation, the advisor (a) queries course performance, (b) suggests courses for the following semester, (c) elicits course nominations, and (d) confirms the schedule (Appendix B). Success in the simulation was based on whether participants were able to create their self-designated schedule. To do this, participants needed to refuse two suggestions made by the advisor. This was done by having participants select their desired courses prior to starting the simulation (Figure 3) and programming the advisor to recommend two courses that did not match the students’ self-identified scheduling goals (Figure 4).

Figure 3

Screenshot of Course Selection Screen (Selections Enhanced)
Advising Simulation

The target refusal presented to participants in the simulation consisted of a request for an alternative and the bi-clausal form *Would I be able to + VP*. The nontarget refusal was adapted to reflect participant responses on the pretest WDCT. This was done by using regular expressions to search for key words and phrases associated with the more direct responses produced by participants on a pilot version of this study. For instance, participants saw one of *I don’t want to take x, I don’t like x*, or *I will not take x*, depending on which of these was most frequent on their pretest WDCT. Additionally, *No* would be included in the nontarget response option if participants made use of it on the WDCT pretest. Further, if participants made use of certain lexical intensifiers (e.g., *hate*), they would see one of the following: *It sounds bad, I hate that*, or *I think it is not good*. **Example 1** shows the refusal response options available to participants in the simulation. The target response option is shown in (1a), and possible permutations for the nontarget response option are shown in (1b). Elements included in parentheses were only presented if they were detected on that participant’s pretest WDCT responses. Finally, participants would fail the task if they accepted the undesired course (1c).
Example 1

Refusal Response Options

1a. Target response option

Would I be able to take another besides <undesired course>? Are there more options to replace that course?

1b. Nontarget response option

(No) + I don’t want to take / I don’t like / I will not take <undesired course> + (It sounds bad. / I hate that. / I think that is not good.)

1c. Fail response (accept course)

I think it’s good.

Although participants only failed the task when they did not refuse the suggestion of the advisor, participants in the experimental groups were shown feedback when they selected the nontarget response. For participants in the implicit feedback group, feedback was presented as a thought bubble (Figure 5). Participants in the explicit feedback group saw the same feedback as participants in the implicit group and were additionally given feedback on the forms that contributed to the advisor’s evaluation of the utterance as impolite (Figure 6). A thought bubble and support character were used to reflect Bardovi-Harlig and Hartford’s (1996) observation that advisors generally did not comment on issues of (im)politeness in the advising session. Further, these methods of feedback avoid identifying the selection as simply incorrect, as they indicate the advisor’s interpretation of the response selected by the participant.

Figure 5

Screenshot of Implicit Feedback
Participants completed the simulation three times. This gave participants six opportunities to refuse in the simulation. Completion of all three attempts took approximately 11 minutes on average.

**Retrospective Comparison Task**

After completing the posttests, participants were asked to compare differences in their language use on the pretest and posttest WDCTs (Figure 7). The purpose of this task was to understand how participants would describe changes in their own refusal patterns.

**Participants**

Ninety international students enrolled in sections of English support classes at a major Midwestern university in the United States of America were randomly assigned to three groups of thirty participants. At the time of the study, most participants (86 out of 90) had spent less than one year living in an English-speaking country. Differences in TOEFL scores were analyzed statistically. Normality was assumed after confirming that z-scores of skewness and kurtosis fell within the range of ± 1.96, as described by Kim (2013). An ANOVA analysis \( F(2,80) = 0.401, p = .61, \eta^2 = .01 \) revealed that these differences were not significant. Participant information and descriptive statistics are summarized in Table 1.
Figure 7

Hypothetical Comparison Task Item

Advisor said:

O.K., so I usually recommend students to take finite math

First Response

No, I don’t want to take it. I hate math.

Second Response

Do I have any other options?

Your answers from time 1 and time 2 are different. Explain why

Table 1

Participant Demographic Information

<table>
<thead>
<tr>
<th></th>
<th>Explicit</th>
<th>Implicit</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Mandarin = 24</td>
<td>Mandarin = 23</td>
<td>Mandarin = 24</td>
</tr>
<tr>
<td></td>
<td>Hindi = 1</td>
<td>Hindi = 2</td>
<td>Hindi = 2</td>
</tr>
<tr>
<td></td>
<td>Korean = 1</td>
<td>Thai = 2</td>
<td>Korean = 3</td>
</tr>
<tr>
<td></td>
<td>Other = 4</td>
<td>Korean = 1</td>
<td>Other = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other = 2</td>
<td></td>
</tr>
<tr>
<td>Mean and SD</td>
<td>$M = 87.8$</td>
<td>$M = 89.4$</td>
<td>$M = 87.7$</td>
</tr>
<tr>
<td></td>
<td>$SD = 6.76$</td>
<td>$SD = 10.8$</td>
<td>$SD = 9.38$</td>
</tr>
</tbody>
</table>

Analysis

To answer RQ1 and RQ2, responses to refusal prompts were first coded at the level of speech act. All responses identified as refusals were subsequently coded at the level of semantic formulas, following a coding scheme adapted from Beebe et al. (1990) (Appendix C). A separate rater coded a random selection of 20 percent of the entire data set, and inter-rater reliability was found to be 89% (Cohen’s $\kappa = .87$). Differences in coding were resolved through discussion.

Uptake was examined by comparing the use of pragmatic features from the simulation on the posttests that
were absent at the time of the pretest (Sydorenko & Tuason, 2016). Although no participants showed uptake of the bi-clausal form, participants were found to diverge in whether they requested or asserted alternatives. Example 2 presents this difference with learner spellings retained.

**Example 2**

*Use of Question Versus Declarative Syntax in Semantic Formula of Alternative*

a. E11: *May I have other classes to replace it?* (Request for alternative)
b. C9: *I will take something else* (Assertion of alternative)

Participants who did not use alternatives on either pretest were eligible to show uptake of the request and assertion perspectives. Although assertions of alternatives did not match the target request in the simulation, the author still considered that as the initial stages of learning. Participants were, therefore, awarded 1 point for requests for alternatives and 0.5 points for assertions. Participants who asserted alternatives on the pretests, however, were only eligible to show uptake of the request perspective. For these participants, requests for alternatives scored 0.5 points.

To answer RQ3, responses to the retrospective comparison task were first coded according to content with one code assigned to each response (Appendix D). Inter-rater reliability was calculated for a subset of 20 percent of the data and found to be 91% (Cohen’s $\kappa = .87$). Differences in coding were resolved through discussion. Responses to this task were then used to triangulate instances of pragmatic development by examining them in relation to pre-posttest language use and feedback provided in the simulation.

**Results**

**Process Data**

This section provides an overview of in-game activity for the subset of learners ($n = 59$) who were eligible to show uptake in the study. Smith and González-Lloret (2020) articulate the importance of examining process data, noting that simply because a feature is included in a simulation does not mean that learners will use it. Crucially, all participants in the treatment groups selected the non-target option at least once, ensuring these participants were exposed to at least one instance of feedback. Groups were also found to select the nontarget selection at similar rates (explicit, 21%; implicit, 20%; comparison, 20%). The target option was the most frequently selected option by all groups (explicit, 57%; implicit, 62%; comparison, 50%). Notably, the comparison group selected this response less frequently than the treatment groups. The comparison group also accepted undesired courses at a higher rate than participants in the feedback groups (explicit, 22%; implicit, 18%; comparison, 30%).

**Uptake**

RQs 1 and 2 examined differences in uptake on the oral and written DCTs. Of the 90 participants in the study, 37 did not request or assert an alternative on either the oral or written pretests: explicit feedback group ($n = 13$), implicit feedback group ($n = 14$), and comparison ($n = 10$). Of these participants, 12 of 13 in the explicit group, 10 of 14 in the implicit group, and 3 of 10 in the comparison group began using alternatives on one of the two posttests (Table 2).
**Table 2**

*Use of Alternative by Participant Group*

<table>
<thead>
<tr>
<th></th>
<th>Explicit (n = 13)</th>
<th></th>
<th>Implicit (n = 14)</th>
<th></th>
<th>Comparison (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part. No</td>
<td>Post ODCT</td>
<td>Post WDCT</td>
<td>Part. No</td>
<td>Post ODCT</td>
<td>Post WDCT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>A</td>
<td>R</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E9</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>E11</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E14</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E16</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>E19</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>E21</td>
<td>7</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E25</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E26</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E28</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E29</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>130</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>6</td>
<td>40</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Gray cells indicate no uptake of alternative on either posttest; R = requests for alternative; A = assertions.*

Additionally, 22 participants (Explicit, n = 7; Implicit, n = 9; Comparison, n = 6) asserted, but did not request alternatives on the pretests and were therefore eligible to show uptake of the request perspective (Table 3). Total uptake scores and measures of central tendency are presented in Table 4.
Table 3

Uptake of Request Perspective Among Participants Who Asserted Alternatives on Pretests

<table>
<thead>
<tr>
<th>Explicit (n = 7)</th>
<th>Implicit (n = 9)</th>
<th>Comparison (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part. No</td>
<td>Post ODCT</td>
</tr>
<tr>
<td>E1</td>
<td>E4</td>
<td>0</td>
</tr>
<tr>
<td>E10</td>
<td>E20</td>
<td>1</td>
</tr>
<tr>
<td>E22</td>
<td>E24</td>
<td>2</td>
</tr>
<tr>
<td>E24</td>
<td>E27</td>
<td>2</td>
</tr>
<tr>
<td>E27</td>
<td>I10</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>I13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I19</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>24</td>
</tr>
</tbody>
</table>

Note. Gray indicates no uptake on either posttest.

Table 4

Uptake Scores and Measures of Central Tendency by Group

<table>
<thead>
<tr>
<th></th>
<th>Explicit (n = 20)</th>
<th>Implicit (n = 23)</th>
<th>Comparison (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ODCT</td>
<td>WDCT</td>
<td>ODCT</td>
</tr>
<tr>
<td>Total uptake score</td>
<td>43</td>
<td>55</td>
<td>26</td>
</tr>
<tr>
<td>Mean</td>
<td>2.15</td>
<td>2.75</td>
<td>1.26</td>
</tr>
<tr>
<td>SD</td>
<td>2.25</td>
<td>3.15</td>
<td>2.24</td>
</tr>
<tr>
<td>Median</td>
<td>1.5</td>
<td>2.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Lower quartile (Q1)</td>
<td>0.625</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Upper quartile (Q3)</td>
<td>3.25</td>
<td>4.875</td>
<td>1.75</td>
</tr>
</tbody>
</table>

Research Question 1

RQ1 examined differences in uptake scores on the posttest ODCT between the participants eligible to demonstrate uptake (n = 59). Due to a violation of normality (z-scores for skewness and kurtosis exceeding ± 1.96, as described in Kim, 2013), a Kruskal-Wallis test was used to investigate group differences. The test ($H(2) = 11.3, p = .003$) indicated a significant difference between groups. Pairwise comparisons using Dunn’s tests with Holm corrections revealed that only the difference between the explicit and comparison groups was significant ($p = .003$, $r = .56$). The strength of the effect was found to approach Plonsky and Oswald’s (2014) recommended guideline ($r = .6$) for strong effects.

Research Question 2

Median uptake values were higher for the feedback groups on the WDCT than on the ODCT but were unchanged for the comparison group. Once again, a Kruskal-Wallis test was used due to violations of
normality (z-scores for skewness and kurtosis greater than ±1.96). The test \(H(2) = 10.0, p = .007\) indicated a significant difference between groups. Post hoc Dunn’s tests with Holm corrections showed the differences between the explicit feedback group and comparison group \((p = .006, r = .49)\) and the implicit feedback and comparison group \((p = .03, r = .44)\) to be significant and of moderate strength. The difference between the two feedback conditions was found to be non-significant \((p = .42, r = .15)\).

**Research Question 3**

**Retrospective Comparison Task**

Responses on the comparison task were used to understand how participants would describe changes in their language use following treatment. Three responses on this task were identified as potentially indicative of learning from the simulation: politeness, simulation, and evaluation. Responses coded as *politeness* involved cases where participants judged one response as more (im)polite than the other, *simulation* included those where the participant referenced the simulation as the source of the change, and *evaluation* involved participants describing one response as better or more appropriate than another (see Appendix D for a list of codes and counts).

The most frequent response for these participants was that their answers were the same or approximately the same (explicit, 34%; implicit, 32%; comparison, 60%). For participants in the explicit feedback group, politeness was a close second (33% of responses), whereas the second most frequent response type for participants in the implicit and comparison groups was that language naturally varies (25% and 13%, respectively). Politeness was the third most frequent code used by the implicit group (17%) but was not mentioned by any of the comparison group participants.

A small percentage of participants in each group were also found to evaluate one response as better than another (explicit, 5%; implicit, 7%; comparison, 3%). Further, only participants in the feedback groups directly attributed the change to the simulation (explicit, 2%; implicit, 2%). Crucially, no responses from the comparison group suggested uptake of the target pragmatic features from the simulation.

**Learning Logs**

This section examines changes in participant E21’s refusals from the pre- to posttests in relation to activity in the simulation and their responses on the comparison task. These findings are then contrasted with those of participant C28 from the comparison group to illustrate the nature of learning from the simulation.

Participant E21 did not make use of alternatives on any of the pretest items. When refusing a math class from the advisor, E21 did so by stating *no* and expressing negative willingness with *I don’t*. On the pretest WDCT, E21 refused the same prompt by stating *no* and reporting a negative feeling.

During the first play-through of the simulation, E21 accepted the advisor’s recommendation of an undesired course and failed the task. On the second play-through, E21 selected the individualized, nontarget response option and was exposed to feedback targeting the semantic formulas previously used on the pretests. Following this, E21 selected the target semantic formula for all remaining refusals.

On the posttest ODCT, E21 refused by giving a reason and requesting an alternative; on the WDCT, E21 again refused by requesting an alternative. Example 3 shows changes in E21’s language use across each of the pre- and posttests as well as the nontarget prompt E21 selected in the simulation and subsequent feedback they received.
Example 3
Participant E21 (Pretest-Posttest Data)
Prompt: O.K., so I usually recommend students to take finite math
Pretest ODCT: No [No]
I don't [Neg. Willingness]
I'm not good at math. [Reason]
It will* *cut
Pretest WDCT: No. [No]
It is so hard for me [Reason]
and I am not interested in Math. [Neg. Feeling]
Selection: “No, I don't want to take MATH-M 118. That is not good.”
Feedback: That was rude.
These phrases are very direct: No, I don’t want, not good
Posttest ODCT: I think it's so hard for me. [Reason]
Can I have other classes? [Alternative]
Posttest WDCT: Can I take any other options for me? [Alternative]

On the comparison task, E21 indicated that the difference between their responses on the pretest and posttest WDCT related to politeness. They noted that the first response was rude, and the second response was more polite (Example 4).

Example 4
Participant E21 (Comparison Task)
Comparison Prompt: Your answers from time 1 and time 2 are different. Explain why.
Comparison: first is too rude. second is more polite.

In contrast, participants in the comparison group showed minimal change following the use of the simulation, gave no responses indicative of uptake, and even identified the nontarget response as the most appropriate. Participant C28, for instance, showed an increased use of I don’t want from pretests to posttests. As shown in Example 5, C28 initially refused the advisor’s suggestion of Math 301 (item one) by first saying sorry and giving a reason on the pretest WDCT. On the posttest WDCT, C28 refused this prompt by stating I don’t want. On the comparison task, participant C28 evaluated I don’t want as a more appropriate response.

Example 5
Participant C28 (Pretest-Posttest Data)
Prompt: Usually I have students take Math 301 next semester.
Pretest WDCT: Sorry. [Regret]
I'm not good at math. [Reason]
Selection: “No. I don’t want to take MATH-M 118. That is not good.”
Posttest WDCT: Sorry [Regret]
but I don't want to take a math class. [Neg. Willingness]
Comparison Prompt: Your answers from time 1 and time 2 are different. Explain why?
Comparison: Because I think the second way might be better for this question.

Discussion
The findings of the study illustrate the potential for feedback to drive pragmatic learning in a computer simulation. This was evidenced by the fact that the explicit feedback group showed significantly greater
uptake than the comparison group on the ODCT ($p = .003$) and by the fact that both feedback groups showed significantly greater uptake than the comparison group on the posttest WDCT (explicit-comparison, $p = .006$; implicit-comparison, $p = .02$). Responses on the retrospective comparison task supported this finding, as participants in the treatment groups, particularly those in the explicit feedback group, attributed differences in their language use to (im)politeness or learning from the simulation more frequently than participants in the comparison group.

These results are consistent with those of Tang and Taguchi (2021), as both studies demonstrate how feedback in a simulated environment can contribute to pragmatic learning, even when instruction external to the simulation is not provided. Moreover, this study found that participants in the explicit feedback condition were able to implement what they had learned from the simulation orally and under time pressure. This is particularly encouraging given the extremely short instruction provided to learners (i.e., 11 minutes).

Differences between the findings of the current study and those of Sykes’ (2009) can potentially be explained by the more explicit nature of the feedback given in this study. In Sykes’ (2009) study, feedback was operationalized as behavioral feedback that required participants to infer that they had selected an (im)polite response based on whether the NPC became angry with them in the simulation, but the implicit feedback in the current study directly indicated that the advisor had interpreted the learner’s selection as rude. The implicit behavioral feedback in Croquelandia, therefore, largely replicated the challenges of learning pragmatics through real-world language use, as it required the learner to work out the mappings of language and (im)politeness based on the actions of an interlocutor. This finding aligns with Carroll’s (2001) claim that “[t]he more inferencing the learner must make on the basis of (non-verbal) perceivable events or information in long-term memory, the less likely she is to identify the corrective intention” (p. 390). This analysis is also consistent with the relatively greater gains Sykes found for apologies, as Sykes (2013) noted that the feedback and hints for apologies were more explicit than the feedback provided on requests.

**Pedagogical Implications**

Because of the short nature of this simulation and the ability for students to complete it on their own, it is easy to imagine it being used as a type of mobile, self-access teaching/learning tool for international students in study-abroad contexts. Ideally, multiple modules related to the study-abroad experience would also be created to provide learners with opportunities to simulate interactions across a variety of contexts.

When deciding to implement any technology-based instruction, however, educators must be mindful of the degree and scope of learning that is likely to occur. Learners in the current study showed change at the level of semantic formulas but did not show uptake of the bi-clausal form. Although this could be due to the short instructional period of this study, this may represent a limitation of learning from this type of text-based simulation, as learners may need explicit metapragmatic instruction to acquire more complex forms, such as in the similation + explicit group of Sydorenko et al.’s (2020) study. Changes at the level of semantic formula alone, however, may lead to a learner’s language being evaluated as more appropriate by an interlocutor. For instance, consider the difference between a refusal from participant E21 on the pretest (No, I hate math) and their response to the same prompt on the posttest (Can I take any other options for me).

It is also important to keep in mind that the development that was observed in this study occurred under controlled conditions where learners were required to repeat the simulation three times in succession. As previous research indicates, learners may not explore the various options in simulations when left to complete them on their own (Sykes, 2013). Teachers should consider ways to get their students to experiment in the simulation. One approach for doing this would be to provide learners with additional objectives to complete in the simulation, such as try to be polite and try to be rude (see Richards, in press). Not only would these exercises give students reasons to explore the various options in the simulation, but they would also direct students’ attention toward pragmatic features and provide students with opportunities to experiment with impoliteness. Notably, impoliteness instruction is often omitted from instructed language courses (see Mugford, 2008). In addition to being an important skill in its own right, it is reasonable to assume that knowing how to be impolite is a necessary component of learning how to be
polite.

**Limitations and Future Directions**

The current study is subject to limitations that can be addressed by future research. First, this study did not use delayed posttests to assess the duration of learning gains. Notably, Tang and Taguchi (2021) found that while both productive and receptive gains were evident on an immediate posttest task, only receptive knowledge gains were observable on the delayed posttest. The current study’s finding that only the explicit feedback group outperformed the comparison group on the ODCT, while both groups outperformed the comparison group on the WDCT, further suggests a need to investigate the durability of learning in relation to task type, as differences between these groups could widen or diminish over time.

Additionally, approximately one-third of the participants were excluded from the analysis because they used the target semantic formula on the pretest and were, therefore, ineligible to demonstrate uptake. While this is a relatively large number, it is important to stress that the decision to examine uptake was made because it is consistent with a view of pragmatic development as an expansion of the pragmatic repertoire and because it demonstrates a stronger causal link between instruction and learning than gain scores. Still, this resulted in a loss of study power and reduced the ability to identify meaningful differences between the treatment conditions. Future studies could work to reduce this loss by adapting the target options in the simulation so that each participant is presented with a pragmatic feature that they did not use during the pretest. In this way, the research instrument could be designed to maximize the potential for learners to demonstrate uptake.

Finally, this study relied on data collected from a single site, which Vitta et al. (2022) identify as a reoccurring, yet problematic design component of instructed L2 research, as individual sites (i.e., universities) cannot be assumed to fully represent a target population (i.e., university students). Future studies in this area will need to carefully evaluate how to conduct multisite investigations that are more representative of a generalized learner population, while also providing instruction that accurately captures the linguistic environments and pragmatic norms across those sites.

**Conclusions**

This study contributes to the small but growing body of research on computer simulations for pragmatics instruction by demonstrating that feedback, particularly explicit negative feedback, can drive pragmatic development. The feedback in this study was designed to concisely convey the relationship between language use in the simulation and the advisor’s internal, negative affective response that resulted from this language use. Future research should continue to systematically explore the unique potential that simulations hold for delivering feedback to language learners and how different implementations of feedback lead to differential learning outcomes. In this way, research on feedback methods unique to gaming/simulated contexts may inform future work on CF in language teaching more generally.

**Acknowledgements**

I would like to express my sincere thanks to my dissertation committee (Kathleen Bardovi-Harlig, Sandra Kübler, César Félix-Brasdefer, and Yucel Yilmaz) for their invaluable insights and for keeping me focused throughout the dissertation process. I would also like to express my appreciation to the wonderful academic advisors at Indiana University who allowed me to observe advising sessions and who answered my questions regarding advising undergraduate students. Thanks also go to Ryan Lidster for supplying the voice of the advisor on the ODCT items.

**Notes**

1. Item prompts with example screenshots from the ODCT and WDCT are available at the IRIS database:
https://www.iris-database.org/details/65IHe-FI1dC

To ensure that the simulation closely reflected advising interactions at the host institution, 12 university division academic advising sessions were observed at the host institution. Advisors were then interviewed on their advising practices, and a general advising scenario for a prospective business student was created.

References


Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching, 48*(1), 1–50. https://doi.org/10.1017/S0261444814000263


Appendix A. ODCT/WDCT Refusal Prompts

Refusals Prompts

1. Usually, I have students take Math 301 next semester.
2. So, how does intro to chemistry sound next semester?
3. O.K., so I usually recommend students to take finite math
4. Usually, I recommend students to take micro-economics.
5. O.K., so for your humanities requirement how does American history sound?
6. Usually, I have students take an activities class next semester.
7. How does an art class sound next semester?
8. O.K., so I usually recommend students to take a music class.
9. Usually, I recommend students to take a computer class.
10. O.K., so for your humanities requirement how does a language class sound?
Appendix B. Advising Simulation Flowchart

Legend
- Green Text/Arrow ⇒ Winning Path
- Red Text/Arrow ⇒ Failing Path
- ◼ = Key Decision Point
- *Taking any RED path results in fail state

Greet

Discuss Current Courses

Refuse course 1

Advisor Recommends Undesired Course

Accept course

Refuse course 2

Advisor Recommends Undesired Course

Accept course

Advisor Elicits Course Nomination

Refuse course 1

Do not nominate course 1

Nominate course

Advisor Elicits Course Nomination

Refuse course 2

Do not nominate course 2

Nominate course

Advisor Confirms Schedule

Win / Fail Screen
### Appendix C. Semantic Formula use by Group

<table>
<thead>
<tr>
<th>Code Example</th>
<th>Explicit (n = 20) (%)</th>
<th>Implicit (n = 23) (%)</th>
<th>Comparison (n = 16) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Reason</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>That's uh difficult. (P13)</td>
<td>99 (31)</td>
<td>62 (20)</td>
<td>60 (19)</td>
</tr>
<tr>
<td><strong>Neg. Feeling</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don't like history. (I1)</td>
<td>99 (31)</td>
<td>111 (36)</td>
<td>98 (31)</td>
</tr>
<tr>
<td><strong>Neg. Willingness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will not take science class. (C15)</td>
<td>55 (17)</td>
<td>55 (18)</td>
<td>40 (13)</td>
</tr>
<tr>
<td><strong>No</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. (I30)</td>
<td>31 (10)</td>
<td>44 (14)</td>
<td>24 (8)</td>
</tr>
<tr>
<td><strong>Alternative (Request)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there another options? (E25)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>44 (14)</td>
</tr>
<tr>
<td><strong>Alternative (Assertion)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will take something else. (C9)</td>
<td>3 (1)</td>
<td>6 (2)</td>
<td>6 (2)</td>
</tr>
<tr>
<td><strong>Regret</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorry. (C21)</td>
<td>14 (4)</td>
<td>26 (9)</td>
<td>36 (11)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 (5)</td>
<td>1 (0)</td>
<td>8 (3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>317</td>
<td>305</td>
<td>316</td>
</tr>
</tbody>
</table>

Note: Other combines semantic formulas totaling less than 5%
## Appendix D. Comparison Task Coding Categories and Counts

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Examples</th>
<th>Explicit (n = 20)</th>
<th>Implicit (n = 23)</th>
<th>Comparison (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>Participant identifies pretest and posttest response as same or similar</td>
<td>they mean same thing (E6)</td>
<td>60 (34)</td>
<td>67 (32)</td>
<td>89 (60)</td>
</tr>
<tr>
<td>Variable</td>
<td>Participant comments that the difference is due to natural variability of language or that they changed their mind</td>
<td>I don't remember the first response. So I can't give the exactly same answer. (C3)</td>
<td>36 (20)</td>
<td>52 (25)</td>
<td>19 (13)</td>
</tr>
<tr>
<td>Politeness</td>
<td>Participant indicates that one response is more (im)polite than the other.</td>
<td>the second answer is much polite to advisor (16)</td>
<td>59 (33)</td>
<td>36 (17)</td>
<td>0</td>
</tr>
<tr>
<td>Explanation</td>
<td>Participant explains a difference between the two responses, but does not mention politeness</td>
<td>that's easier to pronounce (18)</td>
<td>6 (3)</td>
<td>22 (10)</td>
<td>13 (9)</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Participant indicates that one response is either better or worse than another</td>
<td>Second seemed better (112)</td>
<td>9 (5)</td>
<td>14 (7)</td>
<td>4 (3)</td>
</tr>
<tr>
<td>Shorter</td>
<td>Participant indicates that they wanted to write less</td>
<td>Maybe second one I am a little bit board and I didn't think it too much. (C8)</td>
<td>3 (2)</td>
<td>15 (7)</td>
<td>8 (5)</td>
</tr>
<tr>
<td>Simulation</td>
<td>Participant attributes the difference to the treatment or indicates that they aimed to use or not use a specific form presented in the game</td>
<td>Because I saw the sentence &quot;is there any other options&quot;, in the previous part. (C1)</td>
<td>3 (2)</td>
<td>4 (2)</td>
<td>0</td>
</tr>
<tr>
<td>Mistake</td>
<td>Participant indicates they made a mistake or misunderstood some part of the prompt</td>
<td>I type it wrong (E24)</td>
<td>1 (1)</td>
<td>1 (&lt;1)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Uncertain</td>
<td>Participant indicates they do not know why they made their decision</td>
<td>I don't know. (C3)</td>
<td>0</td>
<td>0</td>
<td>13 (9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>177</td>
<td>211</td>
<td>148</td>
</tr>
</tbody>
</table>

*Note. Codes are from subset of 59 participants eligible to show uptake.*
About the Author

Paul Richards is an associate professor in the Department of Humanities at Miyazaki Municipal University where he teaches English language courses and sociolinguistics. His primary research interests relate to L2 pragmatics, CALL, and the role of linguistics instruction in language teacher training.

E-mail: richards@miyazaki-mu.ac.jp