

Output and English as a Second Language Pragmatic Development: The Effectiveness of Output-focused Video-based Instruction

Justin Jernigan

School of Education, English for Academic Purposes Program, Georgia Gwinnett College

1000 University Center Lane, Lawrenceville, GA 30043, USA

Tel: 1-678-360-8379 E-mail: jjerniga@ggc.edu

Received: December 22, 2011

Accepted: January 16, 2011

Published: April 1, 2012

doi:10.5539/elt.v5n4p2

URL: <http://dx.doi.org/10.5539/elt.v5n4p2>

Abstract

Swain's Output Hypothesis proposes a facilitative effect for output on the acquisition of second language morphosyntax. In the context of classroom instruction, a number of studies and reviews suggest that explicit instruction in pragmatic elements promotes development. Other studies have offered less conclusive evidence of the effectiveness of various instructional options. The present classroom-based study examined the effectiveness of output-focused, video-based instruction for the development of second language intercultural pragmatic development among adult ESL learners representing different first language backgrounds in an intensive English program setting. The results suggest that the instructional treatment had a significant effect on a written pragmatic acceptability judgment task (PAJT). Results on a written discourse completion task (DCT) were mixed; no significant effects were identified, but a relatively large effect size was calculated for the output group.

Keywords: Pragmatics, Intercultural communication, Second language learning, Output hypothesis

1. Introduction

The acquisition of pragmatic competence in an additional language entails developing mastery over linguistic, social, and cultural elements. The complex nature of second language (L2) pragmatic development presents learners of English as a second language (ESL) and their classroom instructors with significant challenges. Given the challenges, greater understanding of the process of L2 pragmatic development and those factors (including instruction) that may contribute to this development is a worthwhile goal, and should help to illuminate further the place of pragmatic development in the overall process of second language acquisition (SLA). In particular, the present study examined the effectiveness of an output-focused instructional treatment that makes use of video vignettes for the development of the intercultural pragmatic development of adult ESL learners.

2. Background

2.1 Situating Intercultural Pragmatic Development

The place of pragmatic development within the broader context of second language learning has become an area of focus for many second language researchers over the past decade or more (e.g., Bardovi-Harlig, 1999; Belz & Kinginger, 2003; Kasper, 2001). In the context of classroom instruction, several studies and reviews suggest that explicit instruction in pragmatic elements promotes development (e.g., Jeong & Kaya, 2006; Roever, 2009). Other studies have offered less conclusive evidence of the effectiveness of various instructional options (e.g., Tateyama, 2001). In order to acquire pragmatic competence, learners must develop in terms of both linguistic proficiency and sociocultural awareness, attaining a useful understanding of how language functions in social and cultural contexts (Kasper & Roever, 2005). In addition, it can be a challenge for learners to understand what is pragmatically acceptable in different cultural contexts such as the classroom (Barron, 2005; Davies and Tyler, 2005). In order to address those factors at work in the L2 pragmatic development of ESL students, the focus of the present classroom-based study was the effectiveness of a particular type of output-focused classroom instruction on L2 intercultural pragmatic development.

Kasper (2001) reported that developing pragmatics has been described as being either (1) a component that interacts with a grammar component within a larger model of communicative competence, (2) a form of information processing that resembles the learning of grammar, (3) an emerging competence within a sociocultural framework, or (4) a competence that develops with cultural understanding in recurring situated events within a language socialization framework (p. 502). For the present study, a model of L2 pragmatic competence acquisition was

adopted that assumes the first description offered by Kasper.

Increased understanding of L2 intercultural pragmatic development among ESL learners should also contribute to the broader effort to describe and account for the acquisition of additional languages more generally. In fact, the stated goal of much developing L2 pragmatics research has been to contribute results that are of interest to general SLA theory (Bardovi-Harlig, 1999; Kasper & Roever, 2005). A number of studies have found beneficial effects for interaction (e.g., Belz & Kinginger, 2003; Davies, 2004; Ohta, 2001, 2005; Yoshimi, 2001) or instruction (e.g., Koike & Pearson, 2005; Liddicoat & Crozet, 2001; Martínez-Flor & Fukuya, 2005; Silva, 2003) on pragmatic development. Roever (2009) concludes that, although both sociopragmatic and pragmalinguistic proficiency appear to be amenable to instruction, pragmatics is rarely a featured part of instructional approaches.

2.2 Two Key Concepts in Pragmatic Development

The pragmatic acquisition model assumed in the present study builds on the basic integrated psycholinguistic model of SLA proposed by Gass (1988) and assumed in Izumi (2003), among others. Two key concepts for understanding pragmatic development in the context of the present study are explicated below.

2.2.1 The Role of Output in Pragmatic Development

The model as understood in the present study is realized within the framework of Swain's (1985, 1995, 2005) Output Hypothesis; thus there is a strong role proposed here for output in the acquisition process. If output functions as Swain (1995) has proposed, it serves three basic purposes in the L2 learning process: (1) helping learners notice their inability to express their thoughts in L2 and triggering their seeking input; (2) allowing learners to test hypotheses about L2; and (3) allowing learners to reflect on the forms and functions of L2. The place of spoken production, in particular, has been investigated in numerous studies and edited volumes (e.g., Boxer & Cohen, 2004; Kormos, 2006). Further, in their work on interaction, output, and communicative learning, Swain and Suzuki (2010) propose that meaningful output modified in response to feedback in the context of communicative interaction facilitates language learning. If learners' acquisition of L2 morphosyntax is helped by output, as studies suggest, it is reasonable to investigate the degree to which pragmatic development may similarly benefit from output-focused instruction.

2.2.2 Speech Acts

The targeted outcomes of the present study were the perception and production of speech acts. The ability to use speech acts in real communication is at the core of illocutionary ability (Bachman, 1990). Speech acts may be defined as communicative actions, realized by means of spoken or written language and drawing on the social and cultural context, that reflect the intended meaning or desired effect of a speaker (Brown & Levinson, 1987; Searle, 1979). Five categories of speech acts were identified by Searle (1979): *representatives* (assertions), *directives* (attempts to get someone to do something), *commissives* (obligations one places on oneself), *expressives* (expressions of feelings or reactions), and *declarations* (statements that change the actual state of affairs). More recent approaches to speech acts that recognize their situatedness in interactional sequences (e.g., Bardovi-Harlig & Salsbury, 2004; Félix-Brasdefer, 2006) are acknowledged. Further, as Bara (2010) has pointed out, the analysis of speech acts in isolation is somewhat reductionist in the sense that it ignores the ways in which speech acts are situated within dynamic conversational contexts (p. 63). In the context of the present study, however, Searle's categories were useful to ensure a range of speech act types.

3. Research Questions

In order to investigate the possible contributions of output in the context of classroom-based instruction to the development of L2 pragmatic competence, a study was designed to examine the effectiveness of video-based pragmatic instructional units that included output-focused activity for improving pragmatic awareness and pragmatically appropriate production (Bou-Franch & Garcés-Conejos, 2003; Mir, 2001). The study addressed two principal questions.

3.1 Research Question One

The first research question addressed in the present study was: Do the effects of output-focused video-based instruction differ from the effects of comprehension-focused instruction with respect to learner perceptions of pragmatic appropriateness?

3.2 Research Question Two

The second research question the study examined was: Do the effects of output-focused video-based instruction differ from the effects of comprehension-focused instruction with respect to learners' ability to express speech acts of given types in written form?

4. Methodology

4.1 Study Design

The study design followed that of Izumi (2002) to some extent, but with several important differences. Izumi's study examined the effects of visual input enhancement and output on the acquisition of relativization by 47 participants in 4 treatment groups and 14 in a control group and featured computer-assisted reconstruction and reading tasks. Treatment group participants demonstrated comprehension of a text by writing a recall summary in their first language (L1). The two output groups also reconstructed the text following exposure to the input, whereas the two non-output groups answered extension questions. In the present study, video-based materials were used, and a pragmatic acceptability judgment task (PAJT) served as one measure of pragmatic awareness. Additionally, in response to Bardovi-Harlig and Hartford (2005), who have called for a focus on more authentic data when assessing the production of speech acts, the study included a written discourse completion task (DCT), which offered a fuller understanding of learner pragmatic development than the pragmatic awareness data alone (Kasper & Roever, 2005; Martínez-Flor & Fukuya, 2005; Yuan, 2001).

4.2 Participants

The study drew from the population of adult ESL learners at an intensive English program (IEP) affiliated with a U.S. university. Four different intact ESL classes in two separate rounds of data collection and testing participated in the study, and the final number of participants (*n*) was 34. Although random assignment of students to treatment groups is preferable with respect to promoting the validity of results, the use of intact classes was consistent with the classroom focus of the present study. There were 18 male and 16 female learners, with English intermediate and advanced proficiency levels, based on TOEFL scores and placement at the IEP. The participants' self-reported TOEFL scores ranged from a 56 Internet-based score (approximately equivalent to a 483 paper-based score) to a 240 computer-based score (approximately equivalent to a 587 paper-based score; conversions based on Educational Testing Service, 2005). The 34 learners represented 10 different first language (L1) backgrounds: Arabic, Brazilian Portuguese, French, German, Italian, Korean, Mandarin Chinese, Spanish, Turkish, and Vietnamese.

4.3 Instructional Treatment Materials

The instructional treatment featured a series of 12 video vignettes produced by the researcher with assistance from other ESL instructors. Video-based materials were utilized because of the relationship between successful pragmatic performance and the use of visual cues such as gestures and facial expressions. Following Mir (2001) and Nedashkivska (2004), the loosely scripted videos reflected actual language use as far as possible as the primarily US graduate students and teachers being filmed carried out illocutionary moves. Four types of speech acts, representing two categories, were selected for inclusion in the materials: request refusals, compliment responses, advice giving, and invitations. The first two speech act types are *expressives*, which require speakers to produce surface forms representing deep motives, whereas the latter two are *directives*, which require the socially acceptable use of politeness techniques. In all, there were three vignettes for each speech act, for a total of 12 (see Appendix 1 for two sample transcripts).

An important note at this point concerns the variability of the vignettes with respect to sociopragmatic elements such as difference in power (P), social distance (D), and degree of imposition (R), which may affect task difficulty (Taguchi, 2007). In terms of social power and distance, of the 12 vignettes, five featured teacher-student interaction (higher-to-lower P at relatively high D); five featured interaction between peers (approximately equal P at relatively low D); one involved a patron interacting with an auto mechanic (a slight differential in P at relatively high D); and one featured a father and daughter (higher-to-lower P at relatively low D). In terms of degree of imposition, the three request refusal vignettes involved a fairly high R and two of the invitation vignettes could be considered to have a moderate R.

4.4 Test Instruments

The testing materials included an instrument to gauge pragmatic perception called a pragmatic acceptability judgment task (PAJT; Garcia, 2004) and a written discourse completion task (DCT; Beebe, Takahashi, & Uliss-Weltz, 1990; Liao & Bresnahan, 1996; Lorenzo-Dus, 2001). The pragmatic acceptability judgment task (PAJT) is a closed (i.e., multiple choice), instrument that gives participants the opportunity to identify pragmatically acceptable responses in 12 different dialogues (three for each of the targeted speech acts). Garcia (2004) utilized a PAJT, but focused on the identification of speaker intention. In the present study, the target was both identification of speaker intention and the acceptability of alternative pragmalinguistic forms as learners were required to choose from among possible responses. A sample item follows:

Student: Hi, Dr. Smith. You wanted to see me about my paper?

Professor: Yes. Your paper is well written. I was wondering if you would be willing to share what you have written with the class.

Student: a) Oh, I'd like to, but I don't want to do that.

b) Oh, I'd like to, but I won't because I don't feel prepared.

c) Oh, I'd like to, but I really don't feel prepared.

d) Oh, I'd like to, but since I don't feel prepared, I can't.

The next test of development in pragmatic competence was performance on a twelve-item pretest and posttest written DCT (see Appendix 2 for sample DCT items used in the study). The DCT provides a scenario to which a participant or informant must respond. Boxer (2002) noted that DCTs may take one of two forms: open-ended questionnaires simply ask for the subject to supply the relevant speech act in response to the stimulus (the first line); closed questionnaires elicit a speech act from the subject, but then follow the blank line with a reply from the first speaker, so that the response must take into account not only the initial statement, but also the reply (p. 15). For the present study, the DCTs were open-ended adaptations of instruments used in previous L2 pragmatics research (e.g., Beebe, Takahashi, & Uliss-Weltz, 1990; Sasaki, 1998; Yamashita, 1996) in order to promote the comparability of the results to previous findings. An oral DCT was also included in the design, but because some learners' responses were not clearly recorded, the oral DCT is not included in this discussion.

4.5 Procedure

The learners were assigned as a class to one of two treatment groups: those receiving opportunities for output (the + Output group) and those who did not (the – Output group). The learners were tested in their intact ESL classes: one intermediate-level class and one advanced-level class for each of the two rounds of data collection (for a total of four classes). The assignment of each class to one of the two Output groups was random in the first round of data collection, with the advanced class taking the + Output assignment, and the intermediate group assigned to – Output status for round one. In the second round, the assignments were reversed, so that the intermediate class was assigned to the + Output treatment, with the advanced class assigned to – Output. The two sessions of data collection took place approximately 14 weeks apart, and the procedure was identical in each session. Thus, after the fifth and final day of data collection and posttests for the first round, about 14 weeks elapsed before the first day of pretests for the second round. The number of participants in the first round of data collection and treatment was 17 (7 advanced-level learners in the + Output group, 10 intermediate-level learners in the – Output group). In the second round, another 17 learners participated (10 intermediate-level learners in the + Output group, 7 advanced-level learners in the – Output group).

All participants were told that they would complete pretests, view video vignettes, and complete a learning activity after each vignette. The learners in the + Output group were asked to reconstruct the vignettes after viewing them and having access to the transcripts for a limited period. They were told they would sometimes be reconstructing the vignettes in writing, and at other times, orally. The – Output group participants were told they would respond to several comprehension questions after viewing each video vignette and having access to the transcript. Finally, all participants were informed they would be completing posttests that corresponded to the pretests at the end of the week. They were not otherwise informed of the pragmatic focus of the study or the activities completed by the other group.

The basic treatment and testing schedule, summarized in Table 1, was modeled after that of Izumi (2002). On Day 2, Day 3, and Day 4 of the study, both groups (+ Output and – Output) first viewed the video vignettes. Next, learners in both groups read transcripts of the video vignettes for approximately one minute. Limited exposure time to the written transcripts was given so that participants could not simply copy the transcripts and rely solely on their notes.

Next, following Izumi (2002), the transcripts of the vignettes were taken away from all participants, so that both the + Output and – Output groups had an identical amount of exposure to the pragmatic materials. At that point, the + Output group received opportunity for output in the form of a text reconstruction task. For half of the reconstruction tasks, participants worked to reconstruct the dialogue in writing as accurately and completely as possible. The learners were instructed to take notes during the viewing of the videos and time with the transcript, having been advised that they would not have sufficient time to copy the words verbatim. After the transcripts had been taken away, they were told to write out the exact text of the dialog on the video/transcript as accurately as possible. The oral reconstruction tasks were similar to the written reconstructions, except that the participants reconstructed the vignette dialogues by speaking into an audio recorder.

The – Output group did not receive the opportunity for output, but rather answered comprehension questions about the video vignettes—analogueous to the “extension questions” answered by the – Output groups in Izumi's (2002) study. A possible concern is that incorporating comprehension questions might mean introducing a new variable, but

a pure control group with no activity was ruled out because of the classroom-based nature of this study. Thus, the non-output group was assigned a typical L2 instructional activity.

4.6 Scoring and Rating Procedures

The PAJT in the present study differed from Garcia's (2004) instrument in that the four choices were actual pragmatic forms (examples of speech acts). The PAJT items were organized around the four speech acts: request refusals and compliment responses (expressive speech acts) and advice giving and invitations (directives). The participants were instructed to choose the one form that was most appropriate, given the situation. The participants' PAJT responses were evaluated by the researcher and one assistant (also an ESL instructor). Participants' scores on the PAJT pretest and posttest instruments were obtained by calculating correct responses for all items.

The scoring system developed for scoring the DCT responses is analogous to the Next Generation/internet-Based TOEFL 0-4 scale offered by Educational Testing Services (ETS, 2005), with the exception that the 0-level was eliminated in order to eliminate a middle or "default rating," which might introduce a possible item discrimination problem in the response options (Bachman, 2004; Creswell, 2005). Kasper and Roever (2005) endorsed rating scales for use in pragmalinguistic assessment, noting that Hudson, Detmer, and Brown (1995), among others, used rating scales profitably to allow external raters to evaluate the pragmatic performance of learners in a language-testing context. Also, Cohen (2004) noted that self-assessment rating scales have been used to good effect by researchers such as Brown (2001). Additionally, Yamashita (1996) made use of scoring scales, some of which were adapted from Hudson, Detmer, and Brown (1995) and others, to evaluate the interlanguage pragmatics of learners of Japanese as a foreign language.

The four levels used for rating the written DCTs in the present study are presented in Table 2, along with a general description of each level and indicators included to help flesh out the characteristics of each level. The maximum score for any given 12-item DCT was 48, or an average item score of 4.0.

4.7 DCT Rater Selection and Training

To score the DCT instrument, it was essential to recruit outside raters. Weir (2005) reported that raters who make use of multi-level criteria are limited in the number of levels that they can practically and consistently distinguish. In order to use global scales effectively, raters must become familiar with the criteria used to assess responses and be able to apply their instincts as native speakers to the rating process. Cohen (2004) also discussed the importance of recognizing individual rater characteristics. In order to encourage greater reliability of rating scores, the 28 raters recruited shared three characteristics; all were: (a) of similar age to the majority of the study participants (i.e., traditional university age of 18-25 years old); (b) native speakers of American English; and (c) laypersons with respect to teaching ESL students. These characteristics were selected to encourage ratings that reflected pragmalinguistic norms of English as used in the university context. Utilizing non-experts as raters helped avoid the problem of the raters' having too much sympathy for nonnative-like production because of extensive experience (McNamara, 1996).

The raters were given training in two 30-minute sessions over two weeks. They were exposed to examples of written DCT responses and practice in rating those responses in pairs and as a large group. In this way, the raters were able to become familiar with the scale, to rate several sample responses from the pilot study data and discuss any discrepancies with fellow raters and the researcher.

5. Results

5.1 Research Question One: PAJT Results

The first research question asked whether the effects of output-focused video-based instruction differed from the effects of comprehension-focused instruction with respect to learner perceptions of pragmatic appropriateness. The results of the Pragmatic Acceptability Judgment Task (PAJT) revealed that student participation in the outcome-focused, video-based instructional units had a significant effect on a written PAJT. These results confirm that the effects of this output-focused instruction differed from the effects of typical comprehension-focused instruction on the perception of pragmatic appropriateness.

5.1.1 Descriptive Statistics for the PAJT Results

The descriptive statistics for the PAJT results are described in Table 3. A repeated measures ANOVA revealed a significant effect for Time, $F(1, 32) = 5.284, p < .05$, but no significant Output $\times$ Time interaction and no effect for Output group. Table 4 summarizes these results.

Interestingly, both groups had identical means on the pretest but the + Output group posttest mean was greater than that of the - Output group by a margin of 0.647. A follow-up paired-samples *t*-test revealed that the source of the

Time effect was the pretest-posttest change among the + Output group, $t(16) = 2.452$, $p < .05$, two-tailed. As indicated in Table 5, the small p -value observed for the + Output group pretest-posttest (.026) contrasted with the large p -value for the – Output group (.351).

The effect sizes calculated for the + Output group on the PAJT pretest-posttest, expressed in terms of Cohen's d , was 1.226, much greater than the .8 level normally considered to represent a large effect size. The effect size for the – Output group was not nearly as large ($d = .480$). In fact, the – Output effect size was below the .5 level normally associated with a medium effect size (Cohen, 1988). Effect size calculations for the overall PAJT results confirmed the finding of the follow-up t -test that the + Output group outperformed the – Output group on the PAJT task.

5.1.2 Reliability Estimates of the PAJT Instruments

Given that correct or incorrect responses on the PAJT were decided by the researcher and one assistant, reliability estimates for the pretest and posttest instruments were calculated. For the PAJT pretest, the reliability coefficient of the + Output group's responses was calculated at $\alpha = .820$. For the – Output group, the reliability coefficient was $\alpha = .656$, just below the .7 threshold normally considered to be acceptable in social science research (Vogt, 2005). On the posttest instrument, the reliability coefficient for the + Output responses was $\alpha = .791$. Reliability was calculated at .869 for the – Output group's responses. Three of the four measures were well above the 0.7 threshold, with one measure just below, but approaching, the threshold, suggesting that the instruments were generally reliable.

5.2 Research Question Two: DCT Results

The second research question asked whether the effects of output-focused video-based instruction differed from the effects of comprehension-focused instruction with respect to learners' ability to express speech acts of given types in written form. The study results were mixed in this regard, only partially confirming the existence of a difference between the effects of the output-focused instruction and comprehension-focused instruction on expressing speech acts on the Discourse Completion Task (DCT).

5.2.1 Descriptive Statistics for the DCT Results

The raters' assessments of the participants' DCT responses on a 4-point scale were analyzed. The descriptive statistics for the DCT scores appear in Table 6. The mean change for the + Output group was positive and clearly greater than the small decrease observed for the – Output group (an increase of 3.294 as compared with a mean decrease of 0.353). However, repeated measures ANOVA, as indicated in Table 7, revealed no statistically significant effects or interactions for the DCT pretest-posttest data; thus, no follow-up paired-samples t -test was carried out.

Effect sizes (Cohen's d) calculated for the + Output group was .965, and for the – Output group, $d = .177$. Surprisingly, the effect size calculations revealed that the effect of the treatment on the + Output group was relatively large ($d = .965$), despite the lack of statistical significance identified by repeated measures ANOVA. The effect size calculated for the – Output group was quite small ($d = .177$), as had been expected.

5.2.2 Reliability of the DCT Rating Rubric

The internal consistency and reliability of the pragmatic rating rubric used in the present study to assess the responses of the participants on the DCT pretest and posttest measures was obviously an important area of concern in reviewing the study results. The results of the calculations of the coefficient alpha for inter-rater reliability and internal consistency indicated acceptable levels for the DCT pretest ($\alpha = .885$) and the DCT posttest ($\alpha = .768$). The reliability levels calculated for these results were above the 0.7 threshold considered acceptable in social science research (Vogt, 2005).

6. Discussion

The findings of the study partially support the application of Swain's (1985, 1995, 2005) Output Hypothesis to the acquisition of second language intercultural pragmatic ability. Significant effects for the output-focused instruction were clear for pragmatic recognition (the PAJT), and some effect may be argued for (based on effect size calculations) for pragmatic production, at least for some speech act types. Also, the rating rubric used to assess the pragmalinguistic responses was largely, though not completely, supported.

6.1 Effects of Output-focused Instruction on Pragmatic Development

The positive effects of output-focused instruction on the ESL learners' pragmatic perception in the present study were possibly a result of increased depth of processing of speech acts (pragmalinguistic forms). The output-focused treatment required the + Output participants to reconstruct the texts of the video vignettes with attention to both formal accuracy and pragmatic appropriateness. This combination of grammatical and pragmatic requirements may be thought of as analogous to the *integrative processing* to which Izumi (2003) and Swain and Suzuki (2010) have

attributed successful learning.

The fact that no significant results for the video-based output treatment were identified by statistical tests for the DCT task results despite observable mean increases from pretest to posttest and large effect sizes calculated for the + Output instructional treatment may have more than one possible cause. First, it is possible that the learners who focused on input (the – Output group) were able to produce more acceptable forms on the DCT task precisely because of their focus on input. The + Output group, in this interpretation, may have needed more time than the five days of treatment provided for their production-focused activities to yield results in a production-oriented test such as the written DCT, which calls for a level of precision of expression. If such was the case, it might argue for more perception-focused activities early in learners' development, with more output-focused tasks integrated as learners advance. An additional possible limitation of the study that might have had an effect on results was the fact that the size of each instructional treatment group ($n = 17$ per group) was not large. Fraenkel and Wallen (2003) indicated that experimental or quasi-experimental studies in educational settings should include at least 15 participants per treatment group, so the sample size was not unacceptably small. However, the number of scores may have been too low for the analysis of variance (ANOVA) to identify the significant effects and interactions suggested by the trends in the raw data and effect sizes (see Mackey & Gass, 2005).

6.2 Implications for Pragmatics Instruction and Research

If output is shown over time and in repeated studies to benefit L2 intercultural pragmatic perception more than production, as one might conclude from the results of the present study, classroom instruction should be informed by the distinction. Instructors might be encouraged to focus on production activities later in learners' development, employing more perception-focused approaches to pragmatic instruction at earlier stages of development. As Swain and Suzuki (2010) propose, the facilitation of attention by modified output in the context of interaction should benefit overall L2 development, which is consistent with Roever's (2009) observation that research supports the benefit of instruction for pragmatic development, particularly where instruction involves opportunities for communication or problem solving. The use of video-based instructional units, such as those employed in the present study, is suggested as a useful component of such communicative instruction.

It is likewise important that researchers take into account the differences between the ways learners process oral and written language when designing pragmatic studies that involve multiple modes of exposure to input (Leeser, 2004). It may be that learners who focus on the input they are receiving (e.g., the – Output group in the present study) are better able to assimilate some pragmalinguistic and sociopragmatic features needed for pragmatically successful production. It is admittedly difficult to isolate particular aspects of L2 pragmatic development in authentic classroom settings, but isolating production from interaction, as was attempted in the present study, is a suggested step. Discourse or conversation analytic approaches (e.g., Félix-Brasdefer, 2006), particularly those that take a longitudinal perspective, could also yield more understanding of specific pragmalinguistic or sociopragmatic features affected by instruction. Finally, research into the development of L2 pragmatics in multiple target languages (e.g., Bardovi-Harlig, Félix-Brasdefer, & Omar, 2006; Kasper and Rose, 2002) must be extended in order to identify universal principles of development.

7. Conclusion

In order to promote understanding of the L2 pragmatic development of ESL learners and address the question of the role of output in this developmental process, this classroom-based study examined the effectiveness of an output-focused instructional treatment featuring video vignettes in an intensive English program setting. The results of the study suggest that the instructional treatment had a significant effect on learners' performance on the perception-oriented PAJT. Results on the written DCT, which emphasized learners' ability to express acceptable pragmalinguistic forms, were less clear. Although no significant effects were identified, a relatively large effect size was observed for the group receiving the output instruction. As such, the results offer a measure of support for Swain's (1985, 1995, 2005) Output Hypothesis in an area of second language development outside of morphosyntactic learning. The findings of the study further suggest that the effects of output may be linked to the type of task that is used to measure pragmatic perception and production—a finding that could lead to research into how learners balance the need to focus on formal features of pragmatic speech acts with the need to work under the communicative pressure that certain speech acts entail. Finally, the use of video-based pragmatic instruction is recommended not only in research settings, but in the ESL classroom and other second language learning settings in order to teach more successfully the complexities of pragmatics.

References

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Bachman, L. (2004). *Statistical analyses for language assessment*. New York: Cambridge University Press.
- Bara, B. (2010). *Cognitive pragmatics: The mental processes of communication*. (J. Douthwaite, Trans.). Cambridge, MA: MIT Press.
- Bardovi-Harlig, K. (1999). Exploring the interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics. *Language Learning*, 49, 677-713. <http://dx.doi.org/10.1111/0023-8333.00105>
- Bardovi-Harlig, K., Félix-Brasdefer, C., & Omar, A. (Eds.). (2006). *Pragmatics and language learning*, Vol. 11. Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Bardovi-Harlig, K., & Hartford, B. S. (Eds.). (2005). *Interlanguage pragmatics: Exploring institutional talk* (pp. 1-6). Mahwah, NJ: Lawrence Erlbaum.
- Bardovi-Harlig, K., & Salsbury, T. (2004). The organization of turns in the disagreements of L2 learners: A longitudinal perspective. In D. Boxer & A. D. Cohen (Eds.), *Studying speaking to inform second language learning* (pp. 199-227). Buffalo, NY: Multilingual Matters.
- Barron, A. (2005). Variational pragmatics in the foreign language classroom. *System*, 33, 519-536. <http://dx.doi.org/10.1016/j.system.2005.06.009>
- Beebe, L., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. Scarcella, E. Andersen, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55-74). New York: Newbury House.
- Belz, J. A., & Kinginger, C. (2003). Discourse options and the development of pragmatic competence by classroom learners of German: The case of address forms. *Language Learning*, 53, 591-647. <http://dx.doi.org/10.1046/j.1467-9922.2003.00238.x>
- Bou-Franch, P., & Garcés-Conejos, P. (2003). Teaching linguistic politeness: A methodological proposal. *International Review of Applied Linguistics*, 41, 1-22. <http://dx.doi.org/10.1515/iral.2003.001>
- Boxer, D. (2002). *Applying sociolinguistics: Domains and face-to-face interaction*. Philadelphia: John Benjamins.
- Boxer, D., & Cohen, A. D. (Eds.). (2004). *Studying speaking to inform second language learning*. Buffalo, NY: Multilingual Matters.
- Brown, J. D. (2001). Pragmatic tests: Different purposes, different tests. In K. R. Rose and G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 301-325). New York: Cambridge University Press.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York: Cambridge University Press.
- Cohen, A. D. (2004). Assessing speech acts in a second language. In D. Boxer & A. D. Cohen (Eds.), *Studying speaking to inform second language learning* (pp. 302-327). Buffalo, NY: Multilingual Matters.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Davies, C. E. (2004). Developing awareness of crosscultural pragmatics: The case of American/ German sociable interaction. *Multilingua*, 23, 207-231. <http://dx.doi.org/10.1515/mult.2004.010>
- Davies, C. E., & Tyler, A. (2005). Discourse strategies in the context of crosscultural institutional talk: Uncovering interlanguage pragmatics in the university classroom. In K. Bardovi-Harlig & B. Hartford (Eds.), *Interlanguage pragmatics: Exploring institutional talk* (pp. 133-156). Mahwah, NJ: Lawrence Erlbaum.
- Educational Testing Service. (2005). *TOEFL iBT score comparison tables*. Princeton, NJ: ETS.
- Félix-Brasdefer, J. C. (2006). Teaching the negotiation of multi-turn speech acts: Using conversation-analytic tools to teach pragmatics in the FL classroom. In K. Bardovi-Harlig, J. C. Félix-Brasdefer, & A. Omar (Eds.), *Pragmatics and language learning*, Vol. 11. (pp. 165-198). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.
- Fraenkel, J., & Wallen, N. (2003). *How to design and evaluate research in education* (5th ed.). New York: McGraw-Hill.
- Garcia, P. (2004). Differences in speech act recognition: A pragmatic awareness study. *Language Awareness*, 13, 96-115. <http://dx.doi.org/10.1080/09658410408667089>

- Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, 9, 198–217. <http://dx.doi.org/10.1093/applin/9.2.198>
- Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Technical report #7). Honolulu: University of Hawaii at Manoa, Second Language Teaching and Curriculum Center.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541–577. <http://dx.doi.org/10.1017/S0272263102004023>
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, 24, 168–196. <http://dx.doi.org/10.1093/applin/24.2.168>
- Jeong, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris and L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam: Benjamins.
- Kasper, G. (2001). Four perspectives on L2 pragmatic development. *Applied Linguistics*, 22, 502–530. <http://dx.doi.org/10.1093/applin/22.4.502>
- Kasper, G., & Roever, C. (2005). Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 317–334). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.
- Koike, D. A., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, 33, 481–501. <http://dx.doi.org/10.1016/j.system.2005.06.008>
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.
- Leeser, M. J. (2004). The effects of topic familiarity, mode, and pausing on second language learners' comprehension and focus on form. *Studies in Second Language Acquisition*, 26, 587–615.
- Liao, C., & Bresnahan, (1996). A contrastive pragmatic study on American English and Mandarin refusals. *Language Sciences*, 18, 703–727. [http://dx.doi.org/10.1016/S0388-0001\(96\)00043-5](http://dx.doi.org/10.1016/S0388-0001(96)00043-5)
- Liddicoat, A. J., & Crozet, C. (2001). Acquiring French interactional norms through instruction. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 125–144). Cambridge: Cambridge University Press.
- Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students: A contrastive study. *Journal of Pragmatics*, 33, 107–127. [http://dx.doi.org/10.1016/S0378-2166\(99\)00127-7](http://dx.doi.org/10.1016/S0378-2166(99)00127-7)
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Martínez-Flor, A., & Fukuya, Y. J. (2005). The effects of instruction on learners' production of appropriate and accurate suggestions. *System*, 33, 463–480. <http://dx.doi.org/10.1016/j.system.2005.06.007>
- McNamara, T. (1996). *Measuring second language performance*. New York: Addison Wesley Longman.
- Mir, M. (2001). Un modelo didáctico para la enseñanza de la pragmática. *Hispania*, 84, 542–549. <http://dx.doi.org/10.2307/3657820>
- Nedashkivska, A. (2004). Positive negativity: Attaining pragmatic competence in Ukrainian. *Canadian Slavonic Papers*, 46(1–2), 37–57.
- Ohta, A. S. (2001). A longitudinal study of the development of expression of alignment in Japanese as a foreign language. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in Language Teaching* (pp. 103–120). New York: Cambridge University Press.
- Ohta, A. S. (2005). Interlanguage pragmatics in the zone of proximal development. *System*, 33, 503–517. <http://dx.doi.org/10.1016/j.system.2005.06.001>
- Roever, C. (2009). Teaching and testing pragmatics. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 560–577). Malden, MA: Wiley-Blackwell. <http://dx.doi.org/10.1002/9781444315783.ch29>
- Sasaki, M. (1998). Investigating EFL students' production of speech acts: A comparison of production questionnaires and role plays. *Journal of Pragmatics*, 30, 457–484. [http://dx.doi.org/10.1016/S0378-2166\(98\)00013-7](http://dx.doi.org/10.1016/S0378-2166(98)00013-7)
- Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. New York: Cambridge University Press.

- Silva, A. D. (2003). The effects of instruction on pragmatic development: Teaching polite refusals in English. *Second language studies*, 22, 55-106.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-484). Mahwah, NJ: Lawrence Erlbaum.
- Swain, M., & Suzuki, W. (2010). Interaction, output, and communicative language learning. In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics* (pp. 557-570). Malden, MA: Wiley-Blackwell.
- Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, 28, 113-135. <http://dx.doi.org/10.1093/applin/aml051>
- Tateyama, Y. (2001). Explicit and implicit teaching of pragmatic routines: Japanese *sumimasen*. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 200-222). New York: Cambridge University Press.
- Vogt, W. (2005). *Dictionary of statistics and methodology* (3rd ed.). Thousand Oaks, CA: Sage.
- Weir, C. (2005). *Language testing and validation*. New York: Palgrave Macmillan.
- Yamashita, S. O. (1996). *Six measures of JSL pragmatics* (Technical Report #14). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- Yoshimi, D. R. (2001). Explicit instruction and JFL learners' use of interactional discourse markers. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 223-244). New York: Cambridge University Press.
- Yuan, Y. (2001). An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral DCTs, field notes, and natural conversations. *Journal of Pragmatics*, 33, 271-292. [http://dx.doi.org/10.1016/S0378-2166\(00\)00031-X](http://dx.doi.org/10.1016/S0378-2166(00)00031-X)

Table 1. Daily Schedule for Instructional Treatment and Data Collection

Group	Day 1	Day 2 - Day 4	Day 5
+ Output	• Informed consent • PAJT pretest • DCT pretest	• Video viewing • Text reconstruction (written, oral)	• Video viewing • Text reconstruction (written, oral) • Posttest PAJT • Posttest DCT
- Output	• Informed consent • PAJT pretest • DCT pretest	• Video viewing • Comprehension questions	• Video viewing • Comprehension questions • Posttest PAJT • Posttest DCT

Table 2. Rating Scale for DCT Responses

Score	Description	Indicators
4	response is completely acceptable pragmatically given the context, not noticeably affected by any errors	- approaches native-like usage - minor grammatical errors do not interfere with pragmatic effectiveness - totally appropriate to the situation
3	response is generally appropriate given the context, but contains one or more pragmalinguistic flaws that affect the intended meaning	- near native-like usage - minor grammatical errors may distract from pragmatic effectiveness - may be too brief or too long - somewhat appropriate to the situation
2	response is generally unacceptable pragmatically in this context, though perhaps not in all contexts	- generally non-native-like usage - noticeable errors distract from effectiveness - too brief or too long - generally inappropriate to the situation
1	response is unacceptable pragmatically given the context	- clearly non-native-like usage - numerous errors distract from pragmatic effectiveness throughout - clearly inappropriate to the situation

Table 3. PAJT Descriptive Statistics

Group and Time	<i>M</i>	<i>SD</i>	<i>N</i>
+ Output pretest	7.353	2.473	17
- Output pretest	7.353	2.206	17
Total pretest	7.353	2.308	34
+ Output posttest	8.588	1.544	17
- Output posttest	7.941	2.249	17
Total posttest	8.265	1.928	34

Note: Maximum score on each administration was 12.

Table 4. PAJT ANOVA Table

Source	<i>SS</i>	<i>Df</i>	<i>MS</i>	<i>F</i>
Output	1.779	1	1.779	.272
Time	14.132	1	14.132	5.284*
Output × Time	1.779	1	1.779	.665
Error	85.588	32	2.675	

Note: * $p < .05$

Table 5. PAJT *t*-test Table

Source	<i>t</i>	<i>df</i>	<i>P</i>
+ Output	2.452	16	.026
- Output	.960	16	.351

Table 6. DCT Descriptive Statistics

Group and Time	<i>M</i>	<i>SD</i>	<i>N</i>
+ Output pretest	33.000	5.117	17
- Output pretest	35.471	5.533	17
Total pretest	34.235	5.395	34
+ Output posttest	36.294	5.489	17
- Output posttest	35.118	4.745	17
Total posttest	35.706	5.087	34

Note: Maximum score on each administration was 48 (12 items, 4 points on each).

Table 7. DCT ANOVA Table

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Output	7.118	1	7.118	.187
Time	36.765	1	36.765	2.210
Output × Time	56.529	1	56.529	3.397
Error	532.456	32	16.639	

Appendix 1

Sample pragmatic vignette transcripts

Classroom request refusal

K: Okay everyone, would you please open your books to page 37? Let's do the questions on page 37, please.

D: Dr. Kott?

K: Deborah, yes.

D: Yes, I was wondering, could we review the old questions before we start the new questions?

K: Oh, well, that would be a very good idea, but I think I gave everyone two days to finish that assignment, and you may have finished, but the others perhaps are still working through it, so maybe it would be better to wait until tomorrow for that.

D: I had just, I forgot you gave us an extra day.

K: Oh, well, please, Deborah, please, could you begin on page 37 for us then?

D: Sure.

Car advice

R: Hi there. Can I help you?

L: Hi, um, yeah, my brakes have been making this really loud, funny noise lately. And I was wondering if you could look at them for me.

R: Sure, let's see if we can see anything... Yep, looks like your brake pads are worn down there. You really should change those out as soon as possible.

L: How much is that going to cost?

R: Probably about forty dollars for each pair.

L: You mean I have to change both sides?

R: You really should go ahead and take care of both sides now, so down the road you don't have uneven wear.

L: Gosh. Okay, go ahead and change both pairs then.

R: Alright, sounds good. We'll take care of it for you, alright?

L: Okay. Thank you.

R: You're welcome.

Appendix 2

Sample discourse completion task (DCT) items

(Adapted from Beebe, et al., 1990; Liao & Bresnahan, 1996; Lorenzo-Dus, 2001)

1. You are the owner of a bookstore. One of your best workers asks to speak to you in private.

Worker: As you know, I've been here a little over a year now, and I think you've been pleased with my work. I really enjoy working here, but honestly, I really need an increase in pay.

You: _____

Worker: Then I guess I'll have to look for another job.

2. You are talking to your younger brother about a decision that he has to make.

Brother: You know I've always respected your advice. Yesterday, my boss asked me if I wanted to take a new position in the company. I would make more money, but I would also have to move far away. I'm just not sure what to do. What's your advice?

You: _____

3. You just got your hair cut in a different style. When you see your friend later that day, he says: 'Hey, your new haircut looks great!'

You: _____

4. You want to invite your close friend to attend a party at your house this weekend. What would you say to invite your friend to your party?

You: _____

Note: Complete DCT instruments included 12 items each.