Journal of Second Language Writing 64 (2024) 101098

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/jslw

Is corrective feedback during telecollaboration beneficial? The effects of peer and teacher corrections on L2 writing proficiency

Lieselotte Sippel [a,*,1], Ines A. Martin [b,2]

[a] Yale University, Department of Germanic Languages and Literatures, 320 York Street, New Haven, CT 06511, USA
[b] United States Naval Academy, Languages and Cultures Department, 105 Maryland Ave, Annapolis, MD 21402, USA

ARTICLE INFO

Keywords:
Corrective feedback
L2 writing
Peer feedback
Teacher feedback
Telecollaboration

ABSTRACT

This study investigated the effectiveness of peer and teacher feedback during telecollaboration, specifically, a semester-long email exchange. Participants included 38 students from three sections of a second-semester German course at a US university. They were paired with 40 students of English at a high school in Germany. Of the three second-semester German sections, one group was assigned to a peer feedback condition ($N = 15$), one to a teacher feedback condition ($N = 10$), and one to a telecollaboration-only condition ($N = 13$). Over the course of the semester, each learner sent one German and one English email per week to their email partner. The peer feedback group received corrective feedback on their German emails from their email partners whereas the teacher feedback group received corrections from their instructor. The telecollaboration-only group did not receive corrections. Results from the DIALANG writing test for German, administered at the beginning and the end of the semester, showed that the peer feedback group improved the most on measures of L2 writing, the teacher feedback group made some improvement, and the telecollaboration-only group did not improve. Additional data from tailor-made posttests mirrored these results. Pedagogical implications of these findings are discussed.

1. Introduction

Telecollaboration, that is, virtual exchanges such as email exchanges or video-conferences between learners from different cultural and linguistic backgrounds, has become increasingly popular in second language (L2) classrooms in the past two decades, and perhaps even more so during the COVID-19 pandemic. There are many reasons to implement telecollaboration into L2 classrooms. Most importantly, virtual exchanges can promote learners' intercultural competence (O'Dowd & Dooly, 2020) and language development (Schenker, 2016).

Although some telecollaboration research focuses on language development (Satar & Özdener, 2008; Schenker, 2016, 2017), this research does not typically incorporate corrective feedback (CF) into the exchanges in a systematic manner. Given that CF plays a crucial role in classroom L2 learning and has been shown to be effective both in the oral and written domain (Li, 2010; Kang & Han, 2015), the present study examined the role of CF during telecollaboration. Specifically, the study investigated whether students who

* Corresponding author.
E-mail address: liese.sippel@gmail.com (L. Sippel).
[1] Orcid ID: https://orcid.org/0000–0003-3558–6075
[2] Orcid ID: https://orcid.org/0000–0002-5590–5662

https://doi.org/10.1016/j.jslw.2024.101098
Received 1 May 2023; Received in revised form 22 February 2024; Accepted 2 March 2024
Available online 16 March 2024

received written feedback during a semester-long email exchange improved their writing proficiency more than students who participated in the same email exchange but received no feedback. Moreover, the study examined whether CF was more beneficial when given by an email partner ("peer feedback") or by the course instructor ("teacher feedback").

2. Review of the literature

In the past decades, the field of second language acquisition (SLA) has seen a spike in research studies dealing with telecollaboration. Overall, these studies overwhelmingly show that telecollaboration is an effective tool to promote both intercultural competence (Chen & Yang, 2014; Liaw, 2006; O'Dowd & Dooly, 2020; Thomé-Williams, 2016) and second language acquisition (Angelova & Zhao, 2016; Kabata & Edasawa, 2011; Satar & Özdener, 2008; Schenker, 2016, 2017; Yanguas, 2012). For example, Schenker (2016) reported that a semester-long email exchange improved learners' syntactic complexity, and Angelova and Zhao (2016) found that a written virtual exchange helped learners improve their grammar. In the oral domain, Satar and Özdener (2008) reported that text-chats and voice-chats improved learners' speaking proficiency.

Since telecollaboration is such a powerful tool for cultural and language learning, some researchers have argued that it should become a regular component of L2 classrooms (Çiftçi & Savaş, 2018; O'Dowd, 2016). However, researchers have also pointed out that linguistic gains are "by no means automatic and […] exchanges need to be set up with care as well as with an awareness of best practices" (Godwin-Jones, 2019, p. 8). One limitation of telecollaboration as typically implemented in language classrooms is the lack of CF (Tai et al., 2015; Ware & O'Dowd, 2008). A notable exception is a study by Ware and O'Dowd (2008). In this study, the researchers investigated how and to what extent learners of English and Spanish provided feedback to their exchange partners during online communication, specifically during weekly asynchronous discussions. Learners were assigned to one of two experimental conditions: in the first condition (referred to as "e-tutoring"), students were asked correct linguistic forms they perceived as incorrect; in the second condition (referred to as "e-partnering"), students were not asked to correct errors but could do so on their own initiative. Findings indicated that although students in both conditions expressed a desire for receiving feedback from their exchange partners, such feedback only occurred when explicitly required (that is, in the e-tutoring condition). While this study did not investigate whether students in the e-tutoring condition improved their language skills as a result of the feedback (and if they improved more than students in the e-partnering condition), the study showed that teaching students how to respond to their exchange partners' errors led to more instances of feedback during telecollaboration.

SLA research in areas outside of telecollaboration has reported similar findings. For instance, Fujii et al. (2016) investigated whether training learners to provide CF to each other resulted in more instances of feedback during oral classroom interactions. While the experimental group was trained to correct their peers' errors, the Telecollaboration Only Group was not instructed to do so. Results showed that learners in the experimental group provided more feedback and made better use of the feedback they received than learners in the Telecollaboration Only Group.

In addition to examining whether peer feedback training leads to more instances of feedback, SLA research (but not telecollaboration research) has systematically investigated whether peer feedback can benefit second language development. A growing number of research studies provides evidence that peer feedback is beneficial for language development both in the written and oral domain, as long as instructors ask students to correct their peers' errors and teach them how to do so (Lundstrom & Baker, 2009; Martin & Sippel, 2021; Sato & Lyster, 2012; Toth, 2008). For example, in the oral domain, Sato and Lyster (2012) trained learners to correct their peers' grammatical mistakes during oral peer interaction activities. They found that only students who received such training improved in terms of grammatical accuracy. In the written domain, Lundstrom and Baker (2009) examined whether peer review facilitated learners' writing. They found that it was beneficial, especially for the students who provided peer feedback rather than received it. Based on this research, one would expect that peer feedback during telecollaboration also has the potential to facilitate L2 development. It is also noteworthy that learners wish to receive CF on their errors in classroom settings, both from teachers and from peers (Jean & Simard, 2011; Kaivanpanah et al., 2015; Kartchava, 2016; Sato, 2013; Schulz, 1996, 2001; Sippel, 2020). Thus, one would expect that learners wish to receive feedback during virtual exchanges as well (see also Ware & O'Dowd, 2008).

While there has been increased interest in peer feedback in recent years, the majority of CF studies has examined teacher feedback, both in the oral and the written domain. Overall, there is ample evidence for the effectiveness of oral and written teacher feedback, as shown by several meta-analyses (e.g., Li, 2010, for oral feedback; Kang & Han, 2015, for written feedback). Relevant to the present study, several studies on written CF have shown that it is more effective than no feedback. For example, Bitchener (2008) found that written feedback focusing on English articles (that is, focused feedback) led to more improvement in grammatical accuracy than no feedback. In another study, Frear and Chiu (2015) compared the effectiveness of written CF on weak verbs (focused feedback) to written CF on all structures (unfocused feedback) and found that both led to greater accuracy than no feedback. In a meta-analysis, Kang and Han (2015) also reported a substantive effect of written CF on L2 written accuracy.

In addition, some research has compared the effectiveness of teacher feedback to that of peer feedback. This research has produced mixed findings. In the oral domain, Sippel and Jackson (2015) investigated the effectiveness of peer and teacher feedback during classroom interactions. Results showed that in the short term, both peer and teacher feedback facilitated the acquisition of grammatical structures but in the long term, peer feedback was slightly more effective than teacher feedback. In another study, Martin and Sippel (2021) compared the effectiveness of providing peer feedback, receiving peer feedback, receiving teacher feedback, and receiving no feedback on pronunciation. Results showed that while all experimental conditions were more effective than no feedback, providing peer feedback helped learners improve their own pronunciation the most, followed by receiving teacher feedback, followed by receiving peer feedback. A follow-up study further revealed that peer feedback providers showed better retention of learning gains over time than teacher feedback receivers (Martin & Sippel, 2023). In the written domain, Yang et al. (2006) compared peer and

teacher feedback on learners' essays. Findings indicated that teacher feedback was more likely to be adopted and led to greater improvements in writing. Taken together, these studies suggest that the relative effectiveness of peer and teacher feedback may be mediated by a number of factors, such as the nature of the linguistic target or the medium of the feedback (oral or written).

From a theoretical perspective, the interaction approach (Gass & Mackey, 2015; Long, 1996) supports both the use of telecollaboration and CF in L2 classrooms. This approach stresses not only the importance of interaction between learners and other speakers of the L2 (e.g., telecollaboration), but also the crucial role of negative evidence (i.e., CF). Feedback draws learners' attention to potential problems with their interlanguage, encourages them to modify their output, and facilitates their noticing of non-target-like forms in their own language productions in the future. According to Schmidt's (1990, 2001) noticing hypothesis, this noticing of non-target-like forms is essential for second language development. Therefore, learners should be encouraged to notice mismatches between their own language productions and target language forms (Mitchell et al., 2013).

3. Present study

In light of the lack of previous studies that systematically investigate the effectiveness of CF during telecollaboration using a pretest-posttest design, the present study examined written CF during a semester-long email exchange between learners of German at a US university and learners of English at a German high school. Specifically, the study compared the effectiveness of peer feedback, teacher feedback, and no feedback on learners' emails. While we use the term "peer feedback" to refer to feedback provided by email partners, we recognize that email partners are not the same as "peers" in an L2 classroom: unlike "peers" in an L2 classroom, the "peers" in the present study lived in an environment and went to a school where the language that their partners were learning was spoken, and in many cases, it was their L1 (first language).

Learning gains were assessed by means of a pre- and post-writing-test as well as individualized tailor-made posttests. The study addressed the following research questions:

RQ1: To what extent does corrective feedback during telecollaboration improve learners' writing proficiency when compared to no feedback?

RQ2: Is corrective feedback during an email exchange more effective when provided by peers (that is, email partners) or teachers?

4. Methods

4.1. Participants

Thirty-eight participants (16 females, 22 males) from a private US university took part in the study. Their average age was 21 years (range = 18–34; $SD = 3.68$). The majority of the participants (58%) identified as monolingual L1 speakers of English, 21% identified as bilinguals (L1 English and another L1), and the remaining 21% were L2 speakers of English. At the time of the study, all participants were enrolled in one of three sections of a second-semester German course at the university. All three sections of this course used the same curriculum, following the communicative approach. The main objectives of the course were the improvement of learners' speaking, listening, reading, and writing abilities in a cultural context. Some participants (37%) had chosen to study German as an elective, but most (63%) were taking the class to fulfill the university's foreign language requirement.

In addition to the 38 US participants, 40 students from a German high school took part in the study. They were 16–18 years old, and they were either juniors or seniors at the high school. All of them were enrolled in an English class at the time of the study. Unlike the US participants, who received credit for participating in the email exchange, the German high school students had volunteered to participate in the exchange. The present study focuses only on the learning gains of the US participants.

At the beginning of the semester, each section of the second-semester German course at the US university was assigned to either the PeerF Group ($N = 15$), the TeacherF Group ($N = 10$), or the Telecollaboration Only Group ($N = 13$). A one-way ANOVA comparing participants' writing proficiency on the writing section of the DIALANG test (Lancaster University, n.d.) showed no significant difference between the three groups ($F_{2,35} = 1.19$, $p = .315$), confirming that the groups were of similar writing proficiency at the onset of the study.

Three different instructors taught the three sections. The PeerF Group and the Telecollaboration Only Group were taught by graduate student instructors who were L2 speakers of German, and the TeacherF Group was taught by a language lector who was an L1 speaker of German. The researchers supervised the project but did not participate in the study as instructors.

4.2. Instructional materials and procedure

At the beginning of the semester, each US participant was assigned an email partner from the partner high school in Germany. Before partners were assigned, students from both schools wrote down their academic and personal interests on a piece of paper. This information was used to match students with a partner who shares similar interests. The email exchange took place weekly over the course of a 13-week-semester. It started in week 2 of the semester and was on hold during school breaks (three weeks total). Therefore, the email exchange took place during nine weeks of the semester. Every week, the US participants initiated the exchange on Monday by sending an email to their partners, written in German. The researchers provided the topics for the German emails (the list of topics is provided in Appendix A). By Wednesday, the German participants responded to the German email. They were instructed to write the first half of their email in German. In this first half of the email, they commented on the topic of the Monday email. In the second half of their email, the German participants were instructed to write about a different topic, provided by their English teachers, in English. By

Friday, the US participants responded to the email from Wednesday in English.

Students in the Telecollaboration Only Group were not given any feedback on their emails. Students in the PeerF and TeacherF Group were given CF on all of their emails written in German (except during the first week of the exchange) by either their partner in Germany (PeerF Group) or their instructor (TeacherF Group), and both students in the PeerF and TeacherF Group gave feedback to their email partners on their emails written in English. Because the present study focuses on the US participants, we will focus here on the feedback given to US participants.

During the first week of the project, the German high school teachers had instructed the high school students whose email partners were in the PeerF Group to correct their email partners' errors. To do so, the high school students copied and pasted the email from Monday into a Word Document and then used the "Comment" function to provide direct written feedback on any errors they noticed. Direct feedback was chosen because it was assumed that this type of feedback would be easiest for the German high school students to provide, since they were L1 speakers of the language who did not have previous teaching experience. Comprehensive feedback (i.e., feedback on all types of errors) was chosen after the researchers found during the first week of the email exchange that the German learners did not make sufficient mistakes in their emails to justify a more focused approach to feedback. It is likely that learners made fewer mistakes because the exchange was written rather than oral, giving them enough time to write their sentences.

Example 1 contains a sentence from an email written by a US participant about their trip to Munich.

(1) Wir haben *ins Hofbräuhaus und *die Restaurant Dicke Sophie gegessen.
[We ate in the Hofbräuhaus and the restaurant Dicke Sophie].

The German high school student highlighted the ungrammatical parts of the sentence "ins" [into the] and "die" [the, feminine, nominative or accusative case] using the comment function and then provided the correct words "im" [in the] in the first comment and "dem" [the, neuter, dative case] in the second comment. In the TeacherF Group, participants received the same type of feedback on their emails. However, the feedback was provided by their instructor rather than their email partner. Once the students in either group had received their feedback, they were asked to revise their initial email, highlight the parts that they had corrected, and upload the final version of their email to the learning management system used by their university. Since students in the Telecollaboration Only Group did not receive feedback on their emails, they were not required to revise their emails.

4.3. Pre- and post-assessments

To assess to what extent the feedback provided by email partners (PeerF Group) and teachers (TeacherF Group) had impacted the students' writing proficiency and general proficiency in German, students in all three groups completed two tests.

The first test was the DIALANG test (Lancaster University, n.d.). Research has confirmed the validity of this test (Alderson, 2005). The test is based on the proficiency scales of the Common European Framework of Reference (CEFR), with six potential outcomes: A1, A2, B1, B2, C1, or C2. The DIALANG test is available online and is free of charge. It consists of five parts: reading, writing, listening, vocabulary, and grammar. The participants of the present study only completed the writing part of the test, once at the beginning of the semester (pretest) and once at the end of the semester (posttest). Because one student from the PeerF Group did not complete the DIALANG posttest, 37 participants are included in the results presented below.

The second test was a tailor-made posttest. Although tailor-made tests are less common, there is ample precedence for the benefits of using them in SLA research (e.g., Arellano-Soto & Parks, 2021; McDonough & Sunitham, 2009; Payant & Kim, 2017; Swain & Watanabe, 2013; Zeng & Takatsuka, 2009). The main advantage of these tests is that they are individualized rather than standardized. For instance, in the present study, tailor-made posttests were designed based on the individual errors participants had made in their emails and the personalized feedback they had received on these exact errors. Thus, unlike the DIALANG writing test, which allowed us to assess students' general writing proficiency in German before and after the email exchange, the tailor-made posttests allowed us to assess whether learners benefited from specific feedback they had received during the exchange. Overall, a combination of a standardized testing instrument (such as the DIALANG writing test) and an individualized testing instrument (such as the tailor-made posttests) allow for more precision and greater depth in assessing learning gains between and within groups, thus, making findings more reliable and generalizable.

Each tailor-made posttest contained 25 items and consisted of four sections. The first section was a fill-in-the blank translation task testing a variety of linguistic targets such as vocabulary items, definite articles, or possessive adjectives (see example 2).

(2) _____ (*when*) machen die Kinder ihre Hausaufgaben?
[When do the kids do their homework?].

The second section was another fill-in-the blank task testing grammatical gender, case, and number (see example 3).

(3) Frau Wagner mag am liebsten italienisch__ Essen.
[Frau Wagner prefers Italian food.].

The third section was an error correction task which required participants to find the error in a given sentence and correct it (see example 4). This task was used for a wide variety of linguistic targets, including word order, subject-verb agreement, and vocabulary.

(4) Claire findet, dass Deutsch ist eine schöne Sprache.
[Claire thinks German is a beautiful language.].

The fourth section was a multiple-choice task (see example 5). This task was also used for a variety of linguistic targets, including prepositions, verb conjugation, and vocabulary.

(5) Sumita, Lydia und Yamina Okonkwo gehen jeden Tag um 8 Uhr _____.
a) nach Schule b) nach ihre Schule c) zu Schule d) zur Schule.
[Sumita, Lydia, and Yamina Okonkwo go to school every day at 8 o'clock.].

The researchers designed 38 tailor-made posttests, one for each participant. (One example of a tailor-made posttest is provided in Appendix B.) Before designing the tests, the researchers identified what errors participants had made in their emails over the course of the semester, whether they had received feedback on these errors, and if they had, whether they had corrected these errors in their revised emails. The researchers then designed 25 test items for each participant. The items in the tailor-made posttests only targeted those linguistic items that students had used incorrectly in their emails, received feedback on from their partner or instructor, and then successfully corrected in their revised emails. Because students made very different types of mistakes, each test was different. For example, some tests only contained very few items in the second section (fill-in-the blank task) whereas others contained many. This was because some students made many mistakes related to grammatical gender, case, and number, whereas other made fewer mistakes in this area.

In addition to the DIALANG test and the tailor-made posttests, students completed an exit survey which asked students to answer four open-ended questions, specifically, what they liked about the email exchange, what they learned from the exchange, how it could be improved, and, in the PeerF Group and the TeacherF Group, what impact they believed the corrections had on their learning.

4.4. Scoring and data analysis

To analyze the data from the DIALANG writing test, learners' results from the pretest and the posttest were converted from the CEFR scale to a numerical value, following Schenker (2021) and Tschirner (2016). Conversions were coded as CEFR levels A1 = 1, A2 = 2, B1 = 3. No learner reached the B2, C1, or C2 level, so no conversions above B1 were carried out. The numerical conversion facilitated descriptive statistics and statistical analysis, but we acknowledge that CEFR scores do not represent linear changes in learners' progress toward higher levels of L2 writing proficiency (see Schenker, 2018).

The tailor-made posttests were coded by the first author. For reliability measures, the second author coded 26% of the data set. Inter-rater reliability was 99.8%. The remaining.2% of discrepancies were resolved in discussing the items in question. Each test consisted of 25 items. Some items, such as articles, were binary in coding and received a score of "1" if they were correct and a score of "0" if they were incorrect. Some items, such as vocabulary items, received a score of "1" if they were correct, a score of ".5″ if they were correct but contained a minor spelling mistake that did not render the word incomprehensible (e.g., "*L**ei**blingsfach" instead of "L**ie**blingsfach" '*favorite subject*'), and "0" if the answer was incorrect or no answer was given. Points were added up and a score of correct percentage of answers was computed by dividing the correct points through the total points (e.g., if a learner had 15 points out of 25, their total score was 60%).

For the qualitative analysis of the exit survey, grounded theory methodology (Corbin & Strauss, 2008) was employed to reveal trends in the data. The data were coded separately by both authors and discrepancies were discussed and resolved. The individual student responses were then organized according to the trends identified and are being provided in the Discussion to contextualize the findings.

Given the unreliable and arbitrary nature of null hypothesis significance testing (Plonsky, 2015), we present bootstrapped descriptive statistics for data analysis below, we focus on an interpretation of standardized effect sizes (Cohen's *d*) and Confidence Intervals (CI), and we provide data-rich graphs (following reporting standards set by Larson-Hall & Plonsky, 2015). Cohen's *d* was calculated using Cambridge University Press and Assessments' Effect Size Calculator (n.d.) for between-group comparisons and Mizumoto's R-based application (n.d.) for within-group comparisons. SLA field-specific benchmarks were used for interpretation of these effect sizes: between-group comparison: Cohen's $d = 0.40$ (small), 0.70 (medium), 1.0 (large) and within-group comparison: Cohen's $d = 0.60$ (small), 1.00 (medium), 1.4 (large) (Plonsky & Oswald, 2014). As recommended by LaFlair et al. (2015), bootstrapping was performed by resampling 10,000 times, using a bias-corrected and accelerated 95% CI, and resampling (with replacement) from the entire data set rather than from within each group separately. By resampling, bootstrapping allows researchers to simulate a much larger data set, simulating participant numbers in the thousands, and thus making findings more reliable. Since LaFlair et al. (2015) argue in favor of using bootstrapping in conjunction with parametric statistics, particularly when samples are small—as is often the case in L2 research—, we also provide results of traditional null hypothesis significance testing. These are, however, interpreted alongside CIs of means and of the difference between group means. For example, CIs can be used as a measure of whether the difference between a pair of mean scores is statistically significant: if the mean of one group falls outside the CI for the other group's mean, the difference between these two means is statistically significant. Moreover, for a CI on the comparison of a difference between two means, if the CI does not cover zero, it can be assumed that the difference is statistically significant (see Larson-Hall & Plonsky, 2015; Plonsky, 2015).

5. Results

5.1. DIALANG writing test

As can be seen in Table 1, 57% of the learners in the PeerF Group improved by either one (43%) or even two (14%) levels on the CEFR (from A1 to A2, from A2 to B1, or from A1 to B1), while 30% of learners in the TeacherF Group improved by one level (from A1 to A2 or from A2 to B1), and 23% of learners in the Telecollaboration Only Group improved by either one (15%) or two (8%) levels (from A1 to A2, from A2 to B1, or from A1 to B1). No learner in the PeerF or the TeacherF Group received a lower CEFR score at the time of the posttest, but 14% of the learners in the Telecollaboration Only Group did. The proportions of these findings are displayed in Fig. 1. Descriptively speaking, it is apparent that almost twice as many learners in the PeerF Group showed an improvement of at least one level on the CEFR as compared to learners in the TeacherF Group, and both groups showed more robust gains than the learners in the Telecollaboration Only Group.

Table 2 provides descriptive statistics and results of statistical analyses to illustrate to what extent the groups' posttest scores (i.e., mean CEFR levels converted into a numeric value) were greater than their pretest scores. The pre-to-post comparison in the PeerF Group reaches statistical significance (the 95% CI [0.25, 2.47] does not cross zero) with a medium effect size of Cohen's $d = 1.36$. The same comparison is not statistically significant in the TeacherF Group (the 95% CI [−0.16, 1.40] crosses zero with a small effect size of Cohen's $d = 0.62$) or in the Telecollaboration Only Group (the 95% CI [−0.50, 0.95] crosses zero with a negligible effect size of Cohen's $d = 0.23$).

5.2. Tailor-made posttests

Descriptive statistics for the tailor-made posttests are displayed in Table 3. As can be seen in the table, the mean score (i.e., mean percentage correct) for the PeerF Group (57.6) does not fall within the bootstrapped 95% CI for the TeacherF Group [28.6, 51.4] or the Telecollaboration Only Group [25.2, 46.2], thus indicating that the differences between the means of the PeerF Group and the other two groups are statistically significant. The mean score of the TeacherF Group (40.8), however, falls within the boostrapped 95% CI for the Telecollaboration Only Group [25.2, 46.2], thus indicating that there is no statistical difference between the means of these two groups. This is further reflected in the effect sizes for each level of comparison between the groups. The effect size for the comparison of the mean percentage correct in the PeerF Group and the TeacherF Group is Cohen's $d = 0.94$ with the CI for the effect size not crossing zero [0.07, 1.75], suggesting that this comparison is statistically significant with a medium-to-large effect size. The effect size for the comparison of the mean percentage correct in the PeerF Group and the Telecollaboration Only Group is Cohen's $d = 1.20$ with the CI for the effect size not crossing zero [0.36, 1.96], suggesting that this comparison is statistically significant with a large effect size. The effect size for the comparison of the mean percentage correct in the TeacherF Group and the Telecollaboration Only Group is Cohen's $d = 0.33$ with the CI for the effect size crossing zero [− 0.51, 1.15], suggesting that this comparison is not statistically significant with a negligible-to-small effect size. Fig. 2 provides a visual representation of the data spread for each group.

Parametric statistics run in conjunction with the bootstrapped descriptive statistics described above (as recommended by LaFlair et al., 2015) show the same pattern of statistical significance. That is, a one-way ANOVA suggests that there is a statistically significant difference between the groups' mean scores ($F_{2,35} = 5.61$, $p = .008$). Post hoc independent comparisons (*t*-tests) suggest that this result stems from the PeerF Group significantly outperforming the TeacherF Group ($t_{23} = 2.32$, $p = .030$) and the Telecollaboration Only Group ($t_{26} = 3.15$, $p = .004$). The comparison between the TeacherF and the Telecollaboration Only Group does not reach statistical significance ($t_{21} = .78$, $p = .446$).

6. Discussion

The present study was designed to explore whether written corrective feedback in telecollaboration improves learners' writing proficiency when compared to no feedback (RQ1). It further examined whether CF in telecollaboration is more effective when provided by email partners in the target country or by the learners' course instructor (RQ2). Two testing instruments were employed to answer these questions: each with a different objective, but each informing both research questions. The DIALANG writing test provided information about the development of the three groups—PeerF, TeacherF, Telecollaboration Only Group—over time, that is, over the length of the telecollaborative email exchange spanning one academic semester. The individualized tailor-made posttests, on the other hand, allowed for a comparison between the three groups at the end of the semester and provided data on how learners in the feedback groups benefited from the feedback they received.

Results from the DIALANG pre- and posttest revealed that 57% of learners in the PeerF Group improved their writing proficiency by

Table 1
Change in CEFR Writing Levels Over Time in Each Group.

	Change in CEFR writing level in percentage per group (# of learners)			
	-1 level	No change	+ 1 level	+ 2 levels
PeerF (n = 14)	0% (0)	43% (6)	43% (6)	14% (2)
TeacherF (n = 10)	0% (0)	70% (7)	30% (3)	0% (0)
Telecollaboration Only (n = 13)	15% (2)	62% (8)	15% (2)	8% (1)

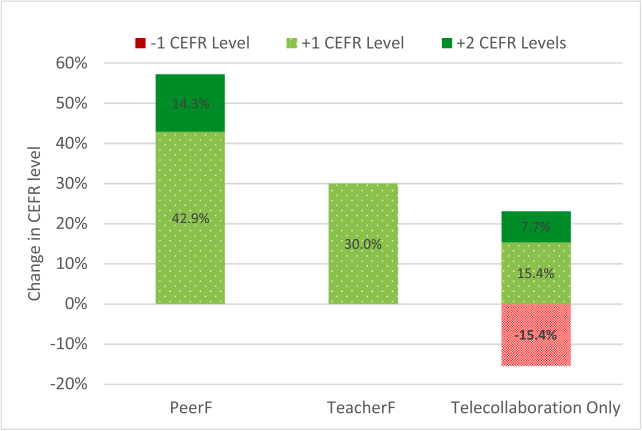

Fig. 1. Proportional Display of Change in CEFR Writing Levels in Each Group.

Table 2
Statistics for the DIALANG Writing Test.

	Mean CEFR Level Score (SD)		BCa 95% Confidence Interval		Pre-Post Comparison		
	Pre	Post	Pre	Post	*t*	*p*	*d* [95% CI]
PeerF (n = 14)	1.43 (.51)	2.14 (.54)	[1.29, 1.57]	[1.93, 2.36]	-3.68	.003 *	1.36## [0.25, 2.47]*
TeacherF (n = 10)	1.5 (.53)	1.8 (.42)	[1.30, 1.70]	[1.60, 2.00]	-1.96	.081	0.62# [– 0.16, 1.40]
Telecollaboration Only (n = 13)	1.69 (.48)	1.85 (.80)	[1.46, 1.85]	[1.54, 2.15]	-0.69	.502	0.23– [– 0.50, 0.95]

Note. Boostrap results are based on 10,000 bootstrap samples. * $p < .009$; – = negligible effect size; # = small effect size; ## = medium effect size

Table 3
Descriptive Statistics for Tailor-Made Posttests.

	Mean Percentage Correct	BCa 95% Confidence Interval	Standard Deviation
PeerF (n = 15)	57.6	[48.4, 67.2]	18.1
TeacherF (n = 10)	40.8	[28.6, 51.4]	17.3
Telecollaboration Only (n = 13)	34.5	[25.2, 46.2]	20.8

Note. Boostrap results are based on 10,000 bootstrap samples.

at least one level on the CEFR scale (e.g., from A1 to A2), followed by 30% in the TeacherF Group, and 23% in the Telecollaboration Only Group. That is, descriptively speaking, both groups that received feedback showed more improvement than the group that did not receive feedback. Notably, gains in the PeerF Group were almost twice as strong as gains in the Telecollaboration Only Group. When looking at statistical significance for pre- to posttest comparisons, only the PeerF Group improved significantly in their writing proficiency over the course of the semester. This comparison had a medium effect size. These trends were further reflected in the results from the tailor-made posttests, comparing the magnitude of improvement between learners in each group at the end of the email exchange. The PeerF Group significantly outperformed both the TeacherF and the Telecollaboration Only Group with a medium-to-large effect (PeerF > TeacherF) and a large effect (PeerF > Telecollaboration Only) for the end-of-semester comparison. The TeacherF Group did not significantly outperform the Telecollaboration Only Group and this comparison only reached a negligible-to-small effect size.

These findings clearly inform the research questions of the present study. First, taken together, they suggest that receiving written CF in a telecollaborative setting is more effective in improving learners' writing proficiency than not receiving feedback. This can be seen in the data showing larger benefits among the two groups that received written CF as compared to the group that did not (i.e., improvement in DIALANG: PeerF 57% & TeacherF 30% > Telecollaboration Only 23%; mean percentage correct in tailor-made posttests: PeerF 58% & TeacherF 41% > Telecollaboration Only 35%). This finding is in line with previous research showing that written CF is effective in improving learners' L2 proficiency (Kang & Han, 2015). It is particularly important, however, in light of the fact that telecollaboration has recently received increasing attention, but so far, studies have focused predominantly on benefits in intercultural competence or language acquisition in general (Angelova & Zhao, 2016; Satar & Özdener, 2008; Schenker, 2016; Thomé-Williams, 2016), rather than on the role and potential of CF. Knowing that learners generally express a desire to receive feedback (Jean & Simard, 2011), and that instructing them to provide feedback increases instances of such feedback (Ware & O'Dowd, 2008), it is important to see that receiving feedback also leads to gains in L2 writing proficiency.

Regarding the question whether feedback in telecollaboration is more effective when provided by the email partner in the target

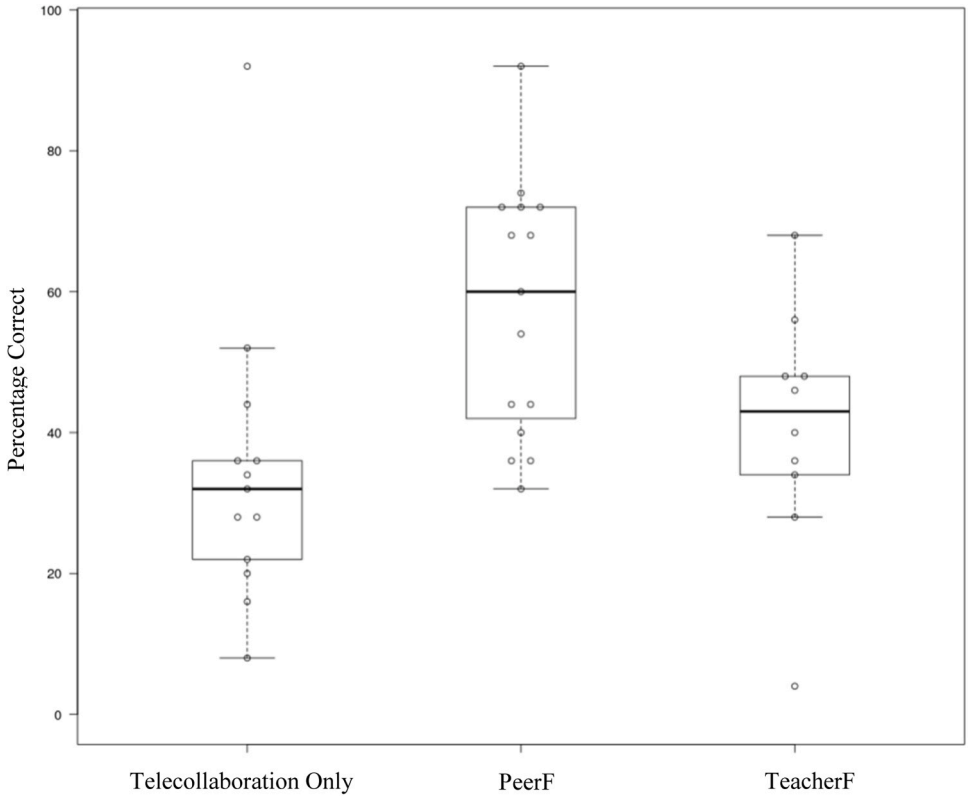

Fig. 2. Visual Representation of Data Spread on the Tailor-Made Posttest for Each Group.

country or by the instructor, results suggested that feedback from the partners in Germany was most beneficial. This was revealed by the findings that (1) in the pre- to post comparison (DIALANG test), only the PeerF Group showed significant improvement in writing proficiency from the beginning to the end of the semester, and (2) at the end of the email exchange (tailor-made posttests), the PeerF Group significantly outperformed the other two groups on measures of grammatical and lexical accuracy in L2 writing. These findings are in line with previous research that has also found peer feedback to be beneficial in improving learners' writing proficiency (Lundstrom & Baker, 2009), as well as research that has shown larger benefits stemming from peer feedback when compared to teacher feedback (Martin & Sippel, 2021, 2023; Sippel & Jackson, 2015; but see Yang et al., 2006). However, previous research looked at peer feedback among L2 learners in the same country, thus, all peers were learners of the same L2 rather than peers in a telecollaborative setting where peers reside in the country of the targeted L2. Nevertheless, there are several reasons that could explain why learners in the PeerF Group still outperformed learners in the TeacherF Group in the telecollaborative setting. These particularly revealed themselves when taking into account the learners' responses to the exit survey. That is, in the learners' responses, it became apparent how important the authenticity of the exchange with their email partners in Germany was to them, specifically, how influential it was that their partners were of a similar age and living in the target country. The survey showed an overwhelming number of responses such as:

(1) It was very interesting to be able to connect to people similar to my age in Germany. – P302
(2) I liked having exposure to the way someone my age would construct phrases, as opposed to just a textbook example. – P313
(3) I enjoyed practicing my German skills with a native speaker. – P309
(4) You get to know an actual person and have a peek of their life. – P112
(5) I liked having a source for how actual Germans speak/write. We are exposed to a lot of German, but it's often hard to tell how things are really said/which translations don't really work/are too literal. – P301

Considering these responses, it comes as no surprise that feedback provided by these peers in Germany was more meaningful/valuable to the learners—and as such, more effective—than the feedback provided by the instructor. This might especially be the case when considering that the feedback during the email exchange was the only feedback the learners received from their partners, whereas the feedback the learners in the TeacherF Group received was just additional feedback amongst a wealth of feedback they had already received from the same source: that is, their instructor also provided feedback on learners' homework and other assignments in the course. This line of reasoning is further reflected in learners' responses to the survey question "What did you learn from the email exchange?" Notably, while all groups received the same question, 8 of 15 learners in the PeerF Group referenced the fact that their

German proficiency improved (e.g., "*I gained a better understanding of word order and colloquial writing." – P108* or "*I learned how to express myself in written German better." – P113*), whereas only 3 of 10 learners in the TeacherF Group provided comments along the same line. Instead, responses from learners in the TeacherF Group focused on cultural gains more so than gains in language or writing proficiency. What is more, a learner in the Telecollaboration Only Group who did not receive feedback even suggested this as a possible improvement of the email exchange ("*Perhaps we could have offered short corrections to one another? Maybe this isn't in the spirit of the exchange, but I didn't know if I sounded competent or whether I was all over the place." – P301*), which further demonstrates how valuable the email partner in the target country was to learners as a source for authentic feedback.[3]

Pedagogically speaking, the finding that CF—whether stemming from an email partner or an instructor—is an effective tool to promote writing proficiency in telecollaboration is important. It shows that although telecollaboration alone can also benefit learners' language skills (e.g., Satar & Özdener, 2008; Schenker, 2016, 2017), gains are greater when CF is incorporated into the exchange. It is then even more encouraging to see that feedback from the email partners in the target country was particularly effective because this lessens the burden on instructors to provide feedback themselves. Our findings show that peers in a telecollaborative setting can be a great source of feedback: on the one hand because learners relate to them differently and seem to strongly value their partners' corrections resulting in a lasting effect on their L2 proficiency, on the other hand because teachers are generally grateful if their learners receive additional help without causing more work for themselves. As such, we call for future telecollaboration projects which include an element of CF as part of the exchange to maximize benefits for the learners.

Finally, we would like to acknowledge some limitations of the present study and make suggestions for the design of future studies. First, the three groups in the present study were taught by three different instructors, two graduate students who were L2 speakers of German and one lector who was an L1 speaker of German. It is possible that students whose teachers were L2 speakers (i.e., students in the PeerF and Telecollaboration Only Group) appreciated L1 interaction with their email partners more than students whose teacher was an L1 speaker (i.e., students in the TeacherF Group). To avoid possible teacher effects in future telecollaboration studies, researchers could randomly assign individual students to a condition (i.e., Peer CF, Teacher CF, or Telecollaboration Only) instead of assigning entire classes to a condition, which was the case in the present study. Second, the sample size in the present study was relatively small, ranging from 10–15 participants per group. However, this limitation was attenuated by the use of robust statistics for the analyses. That is, by resampling 10,000 times, bootstrapping allowed us to simulate a much larger data set, simulating participant numbers in the thousands, and thus making the findings of the present study more reliable. Third, the present study was designed to only investigate the effectiveness of written feedback on L2 writing proficiency. Some telecollobarative exchanges only include asynchronous, written modes of communication, simply because different target languages sometimes come with a difference in time zones and learners cannot be expected to communicate synchronously. Nevertheless, Gen-Z learners in college classrooms are so used to synchronous communication tools (e.g., Zoom/Skype) as well as asynchronous tools that allow for oral communication (e.g., Whatsapp, Flipgrid) that they increasingly wish for other modes of communication than written correspondence. We saw this in our study in survey responses such as *I would like to skype/zoom my partner next time, it would be more beneficial – P308*. Therefore, future studies could investigate the effectiveness of written or oral CF on other L2 competencies as well, such as speaking proficiency or vocabulary development. Finally, the present study did not assess the retention of gains in writing proficiency over an extended period of time. That is, the DIALANG pre- and posttest provided a somewhat delayed assessment because it spanned an entire semester and some corrections occurred at the very beginning of the semester; nevertheless, future studies could attempt to include an assessment several weeks or even months after the end of the telecollaborative exchange, to assess long-term retention of learning benefits.

7. Conclusion

The goal of the present study was to examine whether corrective feedback can promote writing proficiency during telecollaboration, specifically, during a semester-long email exchange. Moreover, the study compared the effectiveness of feedback from email partners to feedback from an instructor. Results showed that feedback played a crucial role: students who received feedback improved their writing proficiency over the course of the semester, and, interestingly, feedback from email partners was even more effective than feedback from an instructor. Importantly, students who received no feedback did not improve their writing proficiency, even though they also participated in the email exchange. This finding underscores Godwin-Jones' (2019) observation that learning gains resulting from telecollaboration are "by no means automatic" and that "exchanges need to be set up […] with an awareness of best practices" (p. 8). Corrective feedback is one of those best practices. To maximize learning gains during telecollaboration, we call for more telecollaboration research that systematically incorporates feedback into virtual exchanges.

CRediT authorship contribution statement

Lieselotte Sippel: Data curation, Methodology, Project administration, Writing – original draft, Writing – review & editing, Conceptualization. **Ines A. Martin:** Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

[3] Anonymous reviewers pointed out that the differences between the PeerF and TeacherF Group could also stem from differences in the number of corrections students received and revisions that were made in each group. To address this question, we have run statistical analyses investigating these factors but found no significant differences on various measures of error correction between the PeerF Group and the TeacherF Group.

Data availability

Data will be made available on request.

Appendix A

Topics for German Emails

Email 1:

Kennenlernen & meine Heimatstadt: Introduce yourself. What's your name? How old are you? What's your major / what subject are you interested in? Why do you study German and how long have you been learning German? What do you like to do in your free time? etc. Also, tell your partner what your home town is and describe it to them. At the end of your email, please ask your partner some questions about them and their home town.

Email 2:

Meine Lieblingsreise: Write about your favorite trip/vacation. Where did you travel (*reisen*) to (country, state, city…)? How did you get there (did you drive, fly, take a train,…)? When did you travel there? Who did you go with (family, friends, alone,…)? What was the weather like (did it rain, did the sun shine, did it snow,…)? What did you do while you were there? Did you try (*probieren*) any local food/drinks? If so, what did you try and how did you like it? Did you meet new people while you were there? Did you go shopping/did you buy anything while you were there? Did you see any local sights? Which one did you find most interesting and why? At the end of your email, ask your partner about their favorite trip/vacation.

Email 3:

Mein erstes Semester an meiner Uni: Write about your first semester at your university. When did you come to your university? Where did you live during your first semester at your university? How did you like your room/apartment? What was the weather like during your first semester at your university? Which classes did you take during your first semester? Did you join any clubs or teams? Did you find new friends / meet other interesting people? What was a typical day like at your university? (When did you get up, what did you do in the morning, afternoon, evening, etc.?) What did you do on the weekends? Did your friends or family come to visit you at your university? Did you travel during Thanksgiving Break? Where did you go?

Email 4:

Essen: What do you typically eat and drink for breakfast, lunch, and dinner? How do you like the food in the dining halls at your university? Do you like to cook and if so, what do you cook? Which restaurant(s) do you like to go to? What do you order when you go there? At the end of your email, ask your partner similar questions about food.

Email 5:

Eine typische Woche in meiner Schulzeit: Write about a typical week in high school.

During the week: When did you get up in the morning? How did you get to school (bus, bike, car)? What classes did you take? What classes were the most interesting/fun/difficult? Did you eat lunch in school and what did you typically eat? What did you do after lunch? Did you have extracurricular events in the afternoons (sports, music, etc.)? What did you do in the evenings? When did you do your homework and how long did it take to do your homework? How much free time did you have during a typical week? How much did you sleep in a typical night?

On weekends: When did you get up on weekends? What did you do during the day on weekends? What did you do in the evenings?

Ask your partner about a typical week in high school for them.

Email 6:

Eine Geschichte/ein Märchen aus meinem Land: Tell your partner a story or fairy tale that is popular in your home country and/or that you liked as a child. Write the story/fairy tale in simple past tense. Please use your own words to tell the story (it is not allowed to copy the story from the internet etc.) At the end of your email, ask your partner to tell you a popular story/fairy tale from Germany.

Email 7:

Spring Break: Tell your partner about your spring break (you can say "Spring Break" or "Frühlingsferien"). What did you do? Did you go anywhere? Did you have to work/study? At the end of your email, ask your partner when their next break is and what they are planning to do.

Email 8:

Haustiere: Write about Haustiere. Do you currently have a pet/pets and/or did you have a pet/pets growing up? If so, tell your partner about your pet(s) (how old are they, what do they look like, what are they like, what do you like to do/play with them etc.). Attach a picture if you like. If you never had a pet, what is the reason you didn't have one, and would you like to have a pet in the future? If so, what pet would you like to have and why? (If you never want to have pets, you can write about your favorite animal: What is your favorite animal and why?).

Ask your partner what pets are popular in Germany, if they have or had pets etc.

Email 9:

Filme/Serien: What is/are your favorite movie(s) and/or your favorite TV show(s)? Tell your partner what they are about and why you like them. Ask your partner what they like to watch. Perhaps they can recommend a German TV show/movie?

Appendix B

Sample Tailor-Made Posttest

Fill in the blanks with appropriate German words. You can find the English translations in parentheses.

1. _____ (*where*) gehen wir heute Abend?
2. Sofie und Nesrin haben in München angehalten, _____ (*when*) sie nach Österreich gefahren sind.
3. Frau Schulz steht _____ (*usually*) um sechs Uhr auf.
4. Yusuf findet Mathematik sehr _____ (*difficult*).
5. Veronika und Kobe Okonkwo haben gestern einen interessanten Film _____ (form of *to see/sehen*).
6. Antonia hat zuerst ihre Hausaufgaben gemacht und _____ (*afterwards*) hat sie mit ihren Freundinnen gespielt.
7. Veronika Okonkwo hat die Kinder gefragt, was sie heute in _____ (*the*) Schule gemacht haben.
8. _____ (*when*) machen die Kinder ihre Hausaufgaben?

Fill in the appropriate ending. If no ending is needed write X or – .

9. Als die Wagners letztes Jahr Urlaub in Italien gemacht haben, haben sie gut__ Wetter gehabt.
10. Frau Wagner mag am liebsten italienisch__ Essen.
11. Was ist dein___ Lieblingsmärchen?
12. Gestern hat Richard in sein__ Lieblingsrestaurant zu Abend gegessen.

There is one mistake in each of the sentences below. Please correct the mistakes.

13. Letztes Jahr haben Phan und Daniel nach München gefahren.
14. Sofie studiert in Dresden, und einmal ist ihre Mutter nach Dresden gegangen.
15. Antonia studiert jeden Tag nach der Schule.
16. Heute ist Noah um 7 Uhr aufgewachsen.
17. Frau Körner führte als Kind jeden Tag mit dem Fahrrad.
18. Am Wochenende Maria hat mit Michael im Restaurant gegessen.
19. Montags arbeitet Herr Moser acht Stunde.
20. Claire findet, dass Deutsch ist eine schöne Sprache.
21. Herr und Frau Ruf verliebten vor vielen Jahren.

Please choose the correct answer.

22. Sumita, Lydia und Yamina Okonkwo gehen jeden Tag um 8 Uhr _____.
    a) nach Schule b) nach ihre Schule c) zu Schule d) zur Schule
23. Im Urlaub _____ ich jeden Tag 1-2 Stunden am Nachmittag.
    a) schlief b) schliefte c) schläf d) schläfte
24. Herr Moser _____ jeden Tag im Park spazieren.
    a) macht b) spaziert c) geht d) läuft
25. Als Kind _____ Frau Gruber gern Kuchen.
    a) backte b) bäckte c) buk d) bukte.

References

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment*. Continuum,.

Angelova, M., & Zhao, Y. (2016). Using an online collaborative project between American and Chinese students to develop ESL teaching skills, cross-cultural awareness and language skills. *Computer Assisted Language Learning, 29*(1), 167–185. https://doi.org/10.1080/09588221.2014.907320

Arellano-Soto, G., & Parks, S. (2021). A Video-Conferencing English-Spanish eTandem Exchange: Negotiated Interaction and Acquisition. *Calico Journal, 38*(2), 222–244. https://doi.org/10.1558/cj.38927

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Cambridge University Press & Assessment – Centre for Evaluation & Monitoring. (n.d.). *Effect size calculator*. ⟨https://www.cem.org/effect-size-calculator⟩.

Chen, J. J., & Yang, S. C. (2014). Fostering foreign language learning through technology-enhanced intercultural projects. *Language Learning & Technology, 18*(1), 57–75. https://doi.org/10125/44354.

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd edition.,). Sage,. https://doi.org/10.4135/9781452230153

Çiftçi, E. Y., & Savaş, P. (2018). The role of telecollaboration in language and intercultural learning: A synthesis of studies published between 2010 and 2015. *ReCALL, 30*(3), 278–298. https://doi.org/10.1017/S0958344017000313

Frear, D., & Chiu, Y. H. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System, 53*, 24–34. https://doi.org/10.1016/j.system.2015.06.006

Fujii, A., Ziegler, N., & Mackey, A. (2016). Peer interaction and metacognitive instruction in the EFL classroom. In M. Sato, & S. Ballinger (Eds.), *Peer Interaction and Second Language Learning: Pedagogical Potential and Research Agenda* (pp. 63–89). John Benjamins.

Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition* (2nd ed..,, pp. 180–206). Routledge.

Godwin-Jones, R. (2019). Telecollaboration as an approach to developing intercultural communication competence. *Language Learning and Technology, 23*(3), 8–28. ⟨http://hdl.handle.net/10125/44691⟩.

Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary, but boring? *Foreign Language Annals, 44*(3), 467–494. https://doi.org/10.1111/j.1944-9720.2011.01143.x
Kabata, K., & Edasawa, Y. (2011). Tandem language learning through a cross-cultural keypal project. *Language Learning & Technology, 15*(1), 104–121. https://doi.org/10125/44239.
Kaivanpanah, S., Alavi, S. M., & Sepehrinia, S. (2015). Preferences for interactional feedback: differences between learners and teachers. *The Language Learning Journal, 43*(1), 74–93. https://doi.org/10.1080/09571736.2012.705571
Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal, 99*(1), 1–18. https://doi.org/10.1111/modl.12189
Kartchava, E. (2016). Learners' beliefs about corrective feedback in the language classroom: perspectives from two international contexts. *TESL Canada Journal, 33*(2), 19–45. https://doi.org/10.18806/tesl.v33i2.1235
LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, *t* tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (pp. 23–45). Routledge.
Lancaster University. (n.d.). *DIALANG*. ⟨https://dialangweb.lancaster.ac.uk/⟩.
Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*(S1), 127–159. https://doi.org/10.1111/lang.12115
Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365. https://doi.org/10.1111/j.1467-9922.2010.00561.x
Liaw, M.-L. (2006). E-learning and the development of intercultural competence. *Language Learning & Technology, 10*(3), 49–64. https://doi.org/10125/44074.
Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bathia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). Academic Press.
Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002
Martin, I. A., & Sippel, L. (2021). Is giving better than receiving? The effects of peer and teacher feedback on L2 pronunciation skills. *Journal of Second Language Pronunciation, 7*(1), 62–88. https://doi.org/10.1075/jslp.20001.mar
Martin, I. A., & Sippel, L. (2023). Long-term effects of peer and teacher feedback on L2 pronunciation (Advance online publication) *Journal of Second Language Pronunciation*. https://doi.org/10.1075/jslp.22041.mar.
McDonough, K., & Sunitham, W. (2009). Collaborative dialogue between Thai EFL learners during self-access computer activities. *TESOL Quarterly, 43*(2), 231–254. https://doi.org/10.1002/j.1545-7249.2009.tb00166.x
Mitchell, R., Myles, F., & Marsden, E. (2013). *Second Language Learning Theories*. Routledge,.
Mizumoto, A. (n.d.). *Comparing paired samples*. ⟨https://langtest.jp/shiny/paired/⟩.
O'Dowd, R. (2016). Emerging trends and new directions in telecollaborative learning. *Calico Journal, 33*(3), 291–310. https://doi.org/10.1558/cj.v33i3.30747
O'Dowd, R., & Dooly, M. (2020). Intercultural communicative competence development through telecollaboration and virtual exchange. In J. Jackson (Ed.), *The Routledge Handbook of Language and Intercultural Communication* (pp. 361–375). Routledge.
Payant, C., & Kim, Y. (2017). Impact of task modality on collaborative dialogue among plurilingual learners: A classroom-based study. *International Journal of Bilingual Education and Bilingualism, 22*(5), 614–627. https://doi.org/10.1080/13670050.2017.1292999
Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (pp. 23–45). Routledge.
Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079
Satar, H. M., & Özdener, N. (2008). The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal, 92*(4), 595–613. https://doi.org/10.1111/j.1540-4781.2008.00789.x
Sato, M. (2013). Beliefs about peer interaction and peer corrective feedback: Efficacy of classroom intervention. *The Modern Language Journal, 97*(3), 611–633. https://doi.org/10.1111/j.1540-4781.2013.12035.x
Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development: Monitoring, practice, and proceduralization. *Studies in Second Language Acquisition, 34*(4), 591–626. https://doi.org/10.1017/S0272263112000356
Schenker, T. (2016). Syntactic complexity in a cross-cultural e-mail exchange. *System, 63*, 40–50. https://doi.org/10.1016/j.system.2016.08.012
Schenker, T. (2017). Synchronous telecollaboration for novice language learners – Effects on speaking skills and language learning interests. *Alsic Apprentissage des Langues et Systèmes d'Information et Dèleïött Communication, 20*(2), 1–18. https://doi.org/10.4000/alsic.3068
Schenker, T. (2018). Making short-term study abroad count – Effects on German language skills. *Foreign Language Annals, 51*(2), 411–429. https://doi.org/10.1111/flan.12339
Schenker, T. (2021). Online journaling and language learning in intensive summer study abroad programs (Advance online publication) *Language Teaching Research*. https://doi.org/10.1177/13621688211036673.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–32). Cambridge University Press.
Sippel, L. (2020). German learners' beliefs about peer interaction and peer feedback. *Die Unterrichtspraxis/Teaching German, 53*(2), 175–190. https://doi.org/10.1111/tger.12135
Sippel, L., & Jackson, C. N. (2015). Teacher vs. peer oral corrective feedback in the German language classroom. *Foreign Language Annals, 48*(4), 688–705. https://doi.org/10.1111/flan.12164
Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals, 29*(3), 343–364. https://doi.org/10.1111/j.1944-9720.1996.tb01247.x
Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal, 85*(2), 244–258. https://doi.org/10.1111/0026-7902.00107
Swain, M., & Watanabe, Y. (2013). Languaging: Collaborative dialogue as a source of second language learning. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3218–3225). Blackwell.
Tai, H. C., Lin, W. C., & Yang, S. C. (2015). Exploring the effects of peer review and teachers' corrective feedback on EFL students' online writing performance. *Journal of Educational Computing Research, 53*(2), 284–309. https://doi.org/10.1177/0735633115597490
Thomé-Williams, A. C. (2016). Developing intercultural communicative competence in Portuguese through Skype and Facebook. *Intercultural Communication. Studies, 15*(1), 213–233.
Toth, P. D. (2008). Teacher- and learner-led discourse in task-based grammar instruction: Providing procedural assistance for L2 morphosyntactic development. *Language Learning, 58*(2), 237–283. https://doi.org/10.1111/j.1467-9922.2008.00441.x
Tschirner, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals, 49*(2), 201–223. https://doi.org/10.1111/flan.12198
Ware, P., & O'Dowd, R. (2008). Peer feedback on language form in telecollaboration. *Language Learning & Technology, 12*(1), 43–63. https://doi.org/10125/44130.
Yang, M., Badger, R., & Zhen, Yu (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179–200. https://doi.org/10.1016/j.jslw.2006.09.004
Yanguas, I. (2012). Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO, 29*(3), 507–531. https://doi.org/10.11139/cj.29.3.507-531
Zeng, G., & Takatsuka, S. (2009). Text-based peer–peer collaborative dialogue in a computer-mediated learning environment in the EFL context. *System, 37*(3), 434–446. https://doi.org/10.1016/j.system.2009.01.003

Lieselotte Sippel is a Senior Lector II and Associate Research Scholar in the Department of Germanic Languages and Literatures at Yale University. Her research interests include corrective feedback, peer interaction, and telecollaboration.

Ines A. Martin is an Associate Professor in the Languages and Cultures Department at the United States Naval Academy. Her research interests include pronunciation instruction, online language learning, and peer feedback.