The Impact of Student-Versus Teacher Led Error Correction in the EFL Classroom: Validity and Reliability Considerations

Aric Denfield, Nihon University Sakuragaoka High School, Japan

The Asian Conference on Education 2023
Official Conference Proceedings

Abstract
Corrective Feedback (CF), defined by Lightbown and Spada (1999) as, 'Any indication to the learners that their use of the target language was incorrect', can be classified as being either teacher- or student-led. Empirical evidence suggests that student-led correction is more effective (Lyster and Ranta, 1997); however, it has been found that teacher-led correction is the most commonly used (Pawlack, 2014). The objective of these interventions is to establish the comparable efficacy and perceived effectiveness of the two forms of error correction with students in a Japanese senior high school and to ascertain their views on appropriate error correction (EC) methods. To do this, a series of tests were designed to gauge students' emerging grammatical accuracy in both oral and written communication. In order to guage students' perceptions of the efficacy of the error correction methods, a short survey was administered at the end of the testing stage. Before beginning the large-scale main trial, an external pilot study was conducted to validate the feasibility of the planned research. The current paper notes the considerations involved in the study, as well as its limitations. It then moves on to detail the modifications that were made to the instruments, the testing procedures and other data collection instruments which increased the validity and reliability of the proposed quasi-experimental study. At the conclusion of the pilot, it was found that the full study could proceed.

Keywords: Error Correction, Pilot, Validity, Reliability, Instrument, Procedure

Introduction

The importance of error correction (E.C.) was described by Lee (2017: 582), who asserts that it can, 'play a critical role in eliminating [learner] errors.' Lyster and Saito state that E.C. has 'significant and durable effects on target language development' (2010: 266). The significance of grammatical accuracy was highlighted by Edge (1997), who writes that successful communication can depend on it.

The value of EC can also be located in student expectations. Fukuda (2004) found that students wanted more error correction than their teachers believed was necessary. Dornyei and Ryan (cited in Kartchava, 2016: 19) describe learner beliefs as, 'significant learner characteristics to take into account when explaining learning outcomes'.

In his work on E.C., Hendrickson (1978) posed the question of who should be responsible for error correction. Support for student-led approaches has come from a number of researchers (see, for example, Harmer, 1991; Bartram and Walton, 1991 and Edge, 1997). However, this conclusion is not universally accepted (see for example, Miao, Badger and Zhen, 2006; Conor and Asenavage, 1994; and Paulus, 1999).

In order to explore the impact of the different error correction methods, a study among senior high school students shall be conducted. Prior to the full-scale main trial, a pilot study was undertaken. This paper presents the results of the pilot and it discusses the importance of pilot studies in the research process in general.

Significance of the Study

Pilot studies allow researchers to assess the validity and reliability of the instrumentation and to make any necessary alterations to the procedures. Doody and Doody (2015) assert that a good pilot study will ensure methodological rigour and can lead to higher quality research and scientifically valid work. Malmqvist et al state that they are a crucial part of the research process.

In spite of their importance, however, Fraser et al (2018) note that there is a lack of published studies on the conduct of pilot studies. The authors state that it would be beneficial if more attention were given to them. This paper, therefore, will help to remedy this deficit.

Structure of the Study

First, there shall be an overview of the proposed study, which will include a literature review, the research questions, and details of the site, sample and intervention. This will be followed by a brief discussion of pilot studies and their role in the research process. The paper will move on to describe the pilot conducted prior to the proposed full-scale main trial. This section shall include details on the sample, the instrumentation and procedures, the results of the item analysis and some initial findings on the impact of the different correction methods. To conclude, the conditions under which the full-scale main trial can proceed shall be presented.

Literature Review

Error correction is, according to Ellis (2006), a form of negative feedback. It was defined by Lightbown and Spada (2017: 216) as, 'Any indication to the learners that their use of the target language was incorrect'. Russell (2009), writes that although E.C. remains a contentious issue in second language learning, it is now generally accepted to play an important role in improving learner outcomes (see also, for example, Li, 2010 and Russell and Spada, 2006).

In his review on the theory and practice of error correction, Hendrickson (1978) questioned the dominant assumption among teachers that it was their responsibility to correct learner errors, writing that peer correction might be more effective in developing the grammatical accuracy of learners. However, it should be noted that the author posited this in relation to written work. With regards to spoken errors, the he felt that the impact of peer correction would be limited to lexical errors.

Edge advocates a student-led approach to error correction. He writes that the advantage of self-correction is that it is easier to remember because, 'someone has put something right in his or her own head' (1997: 24). According to the author, the advantages of peer correction are: it involves learners in listening to and thinking about language; students become less dependent on teachers; and finally, students will be better able to assist each other during pair and group work.

Bartram and Walton (1991) also call for a student-led approach to error correction. The authors state it has four advantages, which can be located within more general discussions on the value of active learning (AL). The authors state that learners will feel more involved; they will learn to be more independent; there will be a greater feeling of cooperation; and finally, it will reduce the amount of time that the teacher spends talking.

The value of active learning is supported by Michael (2006), who asserts that available research supports the effectiveness of a student-centered active pedagogy. The author ascribes this to the positive effect of learners explaining their reasoning to themselves, their peers or to their teachers. In terms of how this should be applied, Michael writes that teachers should, 'reform [their] teaching, employing those particular approaches to fostering active learning that match the needs of [their] students, [their] particular courses, and [their] own teaching styles' (2006: 165).

Thinking of the impact of active learning in EFL and ESL, Caine (2020) states that it has been 'standard fare' for many years. The author notes that in the context of Japan, the Ministry of Culture, Sports, Science and Technology (MEXT) has been explicitly using the term in relation to educational reform at both high school and university level since 2014.

With regards to error correction methods, empirical evidence suggests that a more active approach leads to better results. Lyster and Ranta (1997) looked at the uptake rates of different error correction methods. The methods associated with a student-led approach (elicitation, clarification request and repetition) led to higher rates of uptake than teacher-led approaches: recasts and explicit correction. In response to their research question, 'What combinations of corrective feedback and learner uptake constitute the negotiation of form?' (1997:56), Lyster and Ranta concluded it was student-generated repair that was the most effective. Further, the authors assert that, 'elicitation and metalinguistic feedback proved to

be particularly powerful ways of encouraging repairs that involve more than a student's repetition of the teacher's utterance' (1997: 56).

The value of student-led correction is not, however, universally accepted. Miao, Badger and Zhen (2006) conducted a comparative study on the impact of peer and teacher correction on students' writing. The authors found that, while peer correction did have a positive role to play, teacher correction was more likely to be taken up by students and it would also lead to greater improvement. Similar results were obtained by Paulus (1999) and Conor and Asenavage (1994). Conor and Asenavage found that only 5% of peer feedback resulted in change. It is, of course, important to point out that these findings pertain to the impact of feedback for written as opposed to oral errors.

As noted in the introduction, students' beliefs are an important determinant of the success of a given approach and a number of studies support the adoption of a student-centered approach to E.C. In her study of Japanese as a foreign language class in Sydney, Yoshida, (2008) found that, in general, learners preferred to have the opportunity to think about their own errors before being given the correct form by recast. A study by Katayama (2007), which looked at learners' perceptions of oral error correction, found that the most favoured method was for teachers to indicate that a mistake had been made which would enable the student to self-correct.

A more nuanced picture of error correction emerges from a study conducted by Zembytska et al (2022). The authors assert that the choice of error correction method and corrector will depend on the proficiency level of the students. When investigating students' opinions on which E.C. method would be most likely to have the strongest preventive effect, subjects were required to indicate their proficiency level. The results, which are presented in the table opposite, suggest that more proficient students have a preference for teacher-led techniques, while student-led techniques are favoured by less proficient learners.

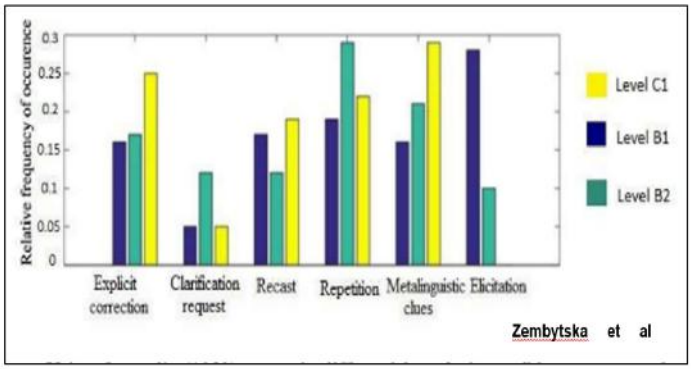


Figure 1: An analysis of students' preferences regarding the choice of corrector

Proposed Study – Outline

Objective

The objective of this study is to test whether a more active approach to error correction will lead to better short and medium-term outcomes in terms of Japanese high school students' grammatical accuracy, specifically with regards to their use of modals. Additionally, students' attitudes to teacher and student-led error correction will also be investigated.

Research Questions

The four research questions that the paper shall answer are:

1. Does error correction have a significant impact on students' grammatical accuracy?
2. Is there a significant difference between the effects of student and teacher led correction?
3. Do students want to have their errors corrected?
4. Which form of error correction do they prefer?

Site and Sample

The site where the study will be conducted is a private senior high school in Tokyo. The school is part of an escalator system. According to NIER (undated), in the escalator system 'a school corporation' will provide education from pre-school all the way through to university.

In terms of the sample, participants will be grade 2 students. They are between 16 and 17 years old, and their level – using the CEFR as a frame of reference – tends to be around A2 or B1. The sample size will be 90, divided into six groups. Groups of students will be randomly assigned to the different treatment methods: two each for the student and teacher-led correction groups and two groups acting as the control. Intact classes will be used. Burden (2011: 80) writes that, although the results might be less generalizable they provide 'authentic learning environments'. The author also asserts that fully randomized experimental designs often suffer from a lack of 'ecological validity' because of the inauthentic environments in which such studies are conducted.

Intervention

The language focus will be on students' use – both oral and written – of the following modals: must / must not, have to / don't have to, and can / cannot. Allowed to and not allowed to have also been included. While these are not modals, they are in the students' textbook in the unit covering this grammar point.

The techniques that shall be used are student and teacher-led correction. Techniques commonly associated with teacher-led correction are recasts, explicit correction and metalinguistic feedback. Those associated with student-led correction are repetition, clarification requests and elicitation. Error correction for oral mistakes will be either immediate or delayed depending on the activity with which students are engaged. Pawlack (2014) writes that the timing of the corrective move will depend on whether the activity is fluency orientated or accuracy based.

The intervention is planned to last for four weeks. A brief outline of the lessons is as follows.

Week 1 ~ must / must not
Input: Reading and listening activities from the students' textbook
(*Time Zones, 3rd Ed. Level 3*)
Output: Writing rules for a school club using must and must not.
Presenting rules to other groups of students.
Discussion of the rules for the different clubs
Conclusion: Error correction, either peer or teacher led

Week 2 ~ can / cannot and allowed to / not allowed to
Input: Listening 'Life in an American school'
Output: Discuss the differences between a Japanese school and an American school
In groups, write rules for a school
Vote on which school they would want to attend
Conclusion: Error correction, either peer or teacher led

Week 3 ~ have to / don't have to
Input: Jigsaw reading 'Rules at home in different countries'
Output: Write a survey and question partner
Create graphs from data and describe the results
Conclusion: Error correction, either peer or teacher led

Week 4 ~ practice and immediate post test
Output: games using modals learnt during the intervention
Immediate post test

Instrumentation

The intervention materials consist of oral production, reading, listening and writing activities. The materials must elicit the target structure, and they have to be appropriate for the students' proficiency level and of interest to them. The materials cover all four language domains, in order to cater to both visual and auditory learners. The majority of the tasks are two-way oral interactional tasks.

The instruments for data collection are:

1. A level check to assess the homogeneity of the students
2. A record sheet to monitor the number of corrective moves made and the frequency with which this led to uptake
3. An oral picture description task
4. A gap-fill task
5. A timed grammaticality test
6. An exit survey

Methodology

The paper uses a quantitative approach. Creswell (2014) defines this as, 'an approach for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures.' (Creswell and Creswell, 2018: 41).

The analyses that shall be conducted are: determining the measures of central tendency to show the degree of variability in the data; a repeated measures ANOVA to determine whether there are statistically significant differences among the means of multiple groups; an independent sample t-test to assess if there is a statistically significant difference between groups; and finally, descriptive statistics to summarize the key findings of the survey.

Error Correction Pilot Study

Introduction

The Dictionary of Epidemiology (2018) describes pilot studies as, 'a small-scale test of the methods and procedures to be used on a larger scale'. A good pilot study, according to Doody and Doody (2015) will ensure methodological rigour and it can lead to higher quality research that can be published. Cohen et al (2015) assert, moreover, that where researchers are using 'a home-grown test' conducting a pilot is 'unavoidable' as it will provide the researcher with essential information on item difficulty and discriminability. In spite of the importance of pilot studies, Fraser et al (2018) write that there is a lack of published works.

The limitations of pilot studies must also, however, be borne in mind. In (2017) states that pilot studies are not suitable for testing hypotheses and the data generated in a pilot study should be treated with caution. As Dzwigol (2020) states, the information obtained from a pilot study is necessarily incomplete.

Pilot Study – Implementation

The pilot study was conducted over a period of roughly 8 months. The stages of the pilot were: planning and preparation, the intervention and data collection, data analysis, the refinement and modification of instruments and procedures, and finally report writing.

The intervention was conducted in line with the lesson outline given above. It ran for four weeks. In that time, students had four lessons, each lasting 50 minutes. The objective of the lessons was to enable students to talk and write about rules in different situations and to use appropriate language to make rules for different contexts. The error correction methods used were: elicitation, repetition and clarification requests for the student-led group, and explicit correction, recasts and meta-linguistic explanation for the teacher-led group.

In terms of the sample, a non-probability convenience sample was used. The participants were 30 grade 3 students, divided into three classes. Group 1 was the control group, group 2 the student-led group and group 3 the teacher-led group. Although the sample consisted of 3rd grade students, their level of English is sufficiently similar for them to be representative. Furthermore, they are familiar with the type of communicative activities that will be used in the intervention. Intact classes were used in order to replicate the conditions of the full-scale main trial.

Instructional Materials

The intervention materials consisted of oral production, reading, listening and writing activities. The parameters set out for judging their suitability were that the materials elicited the target structure, that they were appropriate for the students' proficiency level and that they were of interest to the students. On piloting the materials, it could be seen that these objectives were met and the materials can, therefore, be used in the main trial. Examples of the materials are presented in Appendix A.

Data Collection Instruments

Ellis (2006) writes that it is important to have a range of tests so that both the implicit and explicit knowledge of learners can be accessed. Four home-grown tests were, therefore, devised: a level check to assess the degree of homogeneity among the students, a gap fill-test, a timed grammaticality judgement test and a modals speaking task. The tests were distributed using Loilonote. The students are familiar with this application and this did not cause any problems. Additionally, measures were taken to prevent students from copying each other's work and these appear to have been generally effective.

Level Check

Students were given 15 minutes to complete the test. In the full-scale main trial, however, it is anticipated that the test will take between 8 and 10 minutes. The pilot test was a little longer so that items could be easily removed if they were found to be either too easy or too difficult.

Part 1	
1. I often (park play to the go to) football.	87%
2. I have a (allows that job work me to) with animals.	50%
3. Bill (easy-going is is the who one)	23%
4. What's (in smallest the country world the)?	46%
5. You (home at rest stay should and) tomorrow	37%

Table 1: A sample of the data from the item analysis conducted on the level check test

The results presented opposite are for part 1 of the test. The difficulty of each item was calculated using the formula $\frac{A}{N} \times 100$. 'A' refers to the number of students who answered the question correctly, while 'B' is the total number of students who attempted the item. Following Cohen et al (2015), items falling below 33% or above 67% were deemed as being either too easy or too difficult and so were discarded.

	Class -	XXXXX			
Test No.	Grammar	Vocabula	T. Marks	Total	Rounded
1	3	0	12	3	25
2	1	0	12	1	8
3	4	2	12	6	50
4	5	2	12	7	58
5	4	0	12	4	33
6	4	2	12	6	50
7	7	3	12	10	83
8	8	2	12	10	83
9	3	1	12	4	33
10	7	2	12	9	75

Table 2: Data on the division of students into high and low scoring groups

The level check allowed for the division of students into high and low scoring groups – a prerequisite to establish item discriminability – using the formula $\frac{A-B}{1/2(N)}$. The table presents data for one of the groups. As can be seen, the number of students in the high and low scoring groups is roughly equal.

Gap Fill Test

Feedback from the pilot highlighted a number of issues with the test, which would negatively impact on its validity and reliability. First, the instructions were only in English, which led to the test taking longer than it should have. Next, it was not immediately apparent what the answers should be and for some of the items more than one answer was grammatically possible. Finally, some of the students also found the title of the test to be confusing.

In the second version, instructions are in both English and Japanese. Translations were generated by DeepL Translate and were then checked by a native Japanese speaker. Additionally, in the second version, pictures were included in the left-hand column in order to more clearly show what the answer should be and to make the test more visually appealing and less intimidating. Finally, the title was changed to further reduce any confusion on the nature of the task. The tasks are presented in Appendix B.

Both discriminability and difficulty were calculated. The maximum index of discriminability is 1.00. Cohen et al (2015) assert that any items whose index of discriminability is less than 0.67 should be reviewed as the item is not sufficiently discriminating. Whether that item should still be included, however, is for the researcher to decide. A sample of the results for the item analyses is presented opposite. The questions that are marked in red did not test the target language and so results for these items did not have to be calculated.

Questions	Group 1	Group 2	Group 3	Totals	All St. Total	Disc.	Difficulty
1							
2	H = 2 M = 1 L = 0	H = 4 M = 0 L = 0	H = 2 M = 0 L = 0	H = 8 M = 0 L = 0	9	0.76%	0.30%
3	H = 2 M = 0 L = 1	H = 1 M = 0 L = 0	H = 0 M = 1 L = 0	H = 3 M = 1 L = 1	5	0.19%	0.16%
4	H = 2 M = 1 L = 0	H = 3 M = 0 L = 1	H = 3 M = 1 L = 0	H = 8 M = 2 L = 1	11	0.67%	0.37%
5	H = 4 M = 2 L = 2	H = 3 M = 1 L = 1	H = 3 M = 2 L = 0	H = 10 M = 5 L = 3	18	0.67%	0.60%
6							
7							
8	H = 2 M = 1 L = 0	H = 3 M = 1 L = 1	H = 1 M = 1 L = 0	H = 6 M = 3 L = 1	10	0.48%	0.33%

Table 3: A sample of the data obtained from the item analysis for the gap-fill test

Looking at the results, question 3 had to be discarded as both discriminability and difficulty are quite far outside of the acceptable range. The results for question 2 indicated an acceptable level of discriminability; however, it had a difficulty level of 0.30%. Because, however, this is just outside of the recommended range and because it tests language items that will be included in the intervention, this item was retained.

Timed Grammaticality Judgment Test

The same piloting process was completed for the grammaticality judgment test as for the gap-fill test. The layout of the test was changed following feedback from the students, to include instructions in Japanese as students were initially unclear of what was required of them. Furthermore, the test was shortened quite considerably, as it took far longer to complete than had been anticipated. The final version of the test contains 13 items, 10 of which test for knowledge of the target language. A sample of the test is presented opposite.

Q.	Sentence	Right	Wrong	Don't Know	Correction
1	When students start at a new school, they can be join a club activity.				
2	Students are allowed to use not their mobile phones in class.				
3	Because he trains so hard, he can plays soccer very well.				
4	The longer river in the world is, of course, The Nile.				
5	People are allowed to can walk their dogs in the park.				
6	Many high school students can't getting a part-time job.				
7	To be honest, I hardly never help with the housework.				
8	Passengers have to buy a ticket before they get on the train.				
9	High school students are not allowed to dying their hair.				

Table 4: A sample of the layout and questions in the timed grammaticality judgement test

The table shows a sample of the results of the item analysis. When calculating the results, for each item students could score a total of 2 points. When a student could correctly identify if a mistake had been made, he/she would score 1 point. If the student was further able to provide the correction, the student was awarded 2 points, if the correction was appropriate. Where 'don't know' was marked, this was treated as being incorrect and was given 0.

B	C	D	E	F	G	H	I
Q. No.	Group 1	Group 2	Group 3	Group Totals	Total Correct Answers	Disc.	Difficulty
1	H= 5 M= 0 L= 0	H= 6 M= 1 L=1	H=0 M= 1 L= 1	H= 11 M= 2 L= 2	15	0.43%	0.25%
2	H= 9 M= 4 L= 1	H= 5 M= 3 L= 2	H = 4 M = 1 L= 0	H= 18 M= 8 L= 3	29	0.71%	0.48%
3	H= 3 M=2 L= 0	H= 2 M= 1 L= 1	H= 3 M= 1 L= 1	H= 8 M=4 L=2	14	0.29%	0.23%
4	H= M= L=	H= M= L=	H= M= L=	H= M= L=			
5	H= 6 M= 3 L= 0	H= 4 M= 2 L= 2	H= 6 M= 4 L= 2	H=16 M=9 L=4	29	0.57	0.48
6	H= 3 M= 3 L= 1	H= 6 M= 1 L= 2	H=4 M= 2 L=1	H=13 M=6 L=4	23	0.43	0.38%
7	H= M= L=	H= M= L=	H= M= L=	H= M= L=			
8	H= 4 M= 1 L= 2	H= 4 M= 2 L= 0	H= 3 M= 1 L= 0	H=11 M=4 L=2	17	0.43	0.28
9	H= 6 M= 1 L=1	H= 4 M= 2 L=1	H= 4 M= 0 L= 0	H = 14 M=3 L=2	19	0.57	0.32

Table 5: A sample of the data obtained from the item analysis for the timed grammaticality judgement test

Modals Speaking Tasks

The objective of this task was to have students use the target language to make rules in different situations. Prompts were given at the bottom of the task that students could follow. On piloting the task it became apparent that it suffered from a number of defects. First, the instructions were initially only in English, and this caused students some difficulty as they were not familiar with this type of exercise. Second, many students found the prompts in the box to be confusing and multiple answers were possible. Finally, while students were told orally that they needed to record 8 sentences, many stopped after only 1 or 2. In the second version of the task, instructions are given in both English and Japanese, and symbols were added to the prompts in the box, more clearly directing the students to produce the target language. The tasks are presented in Appendix C.

In spite of these modifications, a number of problems, related to the procedures, persisted. First, some of the audio files were useable others were not. Students used the audio recording function on their iPads to record themselves. IC recorders would have been preferable; however, these are not available at the school. Moreover, it was also clear that some of the

students had copied each other. This could have been overcome by having students record themselves individually; however, this was not practical because of the disruption that would have been caused. In light of these problems, the modals speaking task can not be included in the full-scale main trial.

Survey

The survey was initially pre-piloted, with both Japanese and non-Japanese colleagues, to check that the questions were clear and the translations accurate and to ask for suggestions on whether there were any items that should have been included. Cohen et al (2015) write that structured surveys can prevent respondents from adding further important information. By piloting the questionnaire researchers can see if there are any significant gaps in the instrument.

The instrument was then tested with one of the grade 3 groups to make sure that students understood how to complete a Likert-style survey and to confirm that the items were easily understood. The time taken to complete the survey also had to be confirmed. Porter et al (2004), warn of the danger of survey fatigue. The authors note that research on respondent burden – defined as the time and effort required to complete a survey – has generally found that longer surveys will often result in a lower response rate. A number of modifications were made to the instrument at this point. The surveys are presented in Appendix D.

Having obtained positive feedback from the pilot, the survey was tested again, this time using 2 groups (n=20). Using Excel, the value for Cronbach alpha was calculated, producing a value of 0.72 which indicates that it is a reliable and valid instrument and so it can be used in the full-scale main trial.

Statistical Analyses

As noted above, pilot studies cannot be used to test hypotheses as the data is, necessarily, incomplete. The sample is not sufficiently large and modifications that are made to the data collection instruments will affect results that are subsequently collected. The data presented should, therefore, be regarded as preliminary.

Statistical Analysis Gap Fill Test

An ANOVA was conducted to see if there was a statistically significant difference within the groups in terms of their test scores before and after the intervention. 3 tests were conducted: the pre-test, an immediate post-test and a delayed post-test, which was conducted one month after the intervention. As the analysis was conducted using Excel, it's the columns bar that shows the data for the repeated measures.

Looking first at the control group, the calculated f of 0.762 is less than the critical f of 3.55, which suggests that there was not a significant difference in the grammatical accuracy of these students. Moving on to the results of the teacher-led and student-led groups, in both there was a significant difference between their pre- and post-test scores, which suggests that error correction is effective in developing the grammatical accuracy of learners.

An independent samples t-test on the results of the immediate post-tests was then conducted. Between the control and the teacher-led group the results were statistically significant, as can

be seen from the p value. However, the t-test comparing the teacher-led and student -led groups failed to produce a statistically significant result.

The data are presented in Appendix E.

Statistical Analysis Timed Grammaticality Judgement Test

Looking at the values for the calculated f and the critical f the results of the ANOVA indicate that there was a significant difference in the results of all of the groups. As can be seen, however, the results for the teacher and student-led groups suggest that for these groups the intervention had a far greater impact.

An independent samples t-test on the results of the immediate post-tests was undertaken to look at the differences between the different groups. Between the control and the teacher-led group the results were statistically significant, as can be seen from the p value. However, once again the t-test comparing the teacher-led and student-led groups failed to produce a statistically significant result.

The data are presented in Appendix F.

Conclusion

Pilot studies are an essential part of the research process. They allow researchers to test the feasibility of a proposed study and to test the instrumentation and procedures that will be used. This paper has looked at the implementation of a pilot study prior to a full-scale main trial, which will investigate the efficacy of student- versus teacher-led correction, as well as students' attitudes to error correction.

Regarding the instructional materials, benchmarks were defined against which their suitability could be measured. The materials successfully elicited the target structures and they were both appropriate for the students' proficiency level and of interest. It was, therefore, concluded that they could be used.

Moving on to the data collection instruments, in terms of the gap fill and timed grammaticality judgement tasks, as a result of the feedback obtained from the students, a number of important modifications were made, which greatly increased their clarity. Furthermore, the results of the item analyses show that with the modifications that were made both instruments can be used in the full-scale main trial. The procedures that were used were suitable and did not have to be altered.

In terms of the modals speaking task, in spite of the changes that were made, the data that was obtained was often neither valid nor reliable. Copying continued to be a problem and a number of the audio files could not be transcribed because of the poor sound quality. The tasks will not, therefore, be used in the full-scale main trial.

With regards to the survey, feedback from the students led to a number of the items being changed. The result of the Cronbach analysis confirms its internal consistency and shows that it can be used in the full- scale trial.

Looking at the data that was obtained, while this is only preliminary, a number of tentative conclusions can be drawn. Error correction does appear to have a positive impact on students' grammatical accuracy with regards to their use of modals expressing permission. This impact could be seen in both immediate and delayed post-tests. The data was not clear, however, as to whether student or teacher-led correction was more effective. The results of the survey indicated, though, that a small majority favoured student-led techniques.

Appendixes

Appendix A– Intervention Instructional Materials

Cooking Club

You are the leaders of the cooking club. You want the students to have fun, to make a lot of delicious food and, most importantly to be safe. Kitchens can be a dangerous place and rules are important. So what rules do you think are important?

Task

In your group, use **must** and **must not** to write 5 rules for cooking club members.
(You can use these ideas to help)

1. _____
2. _____
3. _____
4. _____
5. _____

Rules at Home ~ Reading (A)

Sarah is sixteen and she lives with her mother and father in a small home in Brighton in England.

Part 1
Read the text about the different rules that Sarah has and then answer the questions.

Some of my friends have very strict parents. But my parents are pretty easy-going. Of course, there are some rules though.

I have to clean my room every day. I have to do the vacuuming and put all my clothes into the washing machine. I also have to take the trash out. I don't have to clean the other rooms. My parents do that.

During the week I have to go to bed by 11:30 p.m. It's important to sleep a lot. At the weekend, I don't have to go to bed early, because on Saturday and Sunday I don't have to go to school, so I can stay in bed until late.

One strange rule that we have is that I have to cook once a week. Luckily, I don't have to do the washing up. I hate washing up!

Sarah		Dave	
1. How often does she have to clean her room?		1. How often does he have to clean her room?	
2. What time does she have to go to bed?		2. What time does he have to go to bed?	
3. Does she have to go to school at the weekend?		3. Does he have to go to school at the weekend?	
4. What is a strange rule that she has?		4. What is a strange rule that he has?	

Part 2
Now, talk to your partner and answer the questions for Dave.

Glossary
Strict = 厳しい
Easy-going = 気楽な
Washing up = 洗い物

Life at Home – Survey

Every home has different rules.
Write questions with have to to find out the rules at your partner's house.
Remember, we can start questions with 'do' or a question word, who, what, when etc.

Questions	Student 1	Student 2
1. How much homework do you have to do every day?		
2. Do you have to take your shoes off at home?		
3.		
4.		
5.		
6.		
7. What rule is the hardest to do?		

Appendix B – Modals Gap-fill Task

Gap Fill - Modals

Instructions

- **Look at the sentences below.**
- **Write the missing word or words in the spaces.**
- **Try to answer all of the questions.**
- **This is a review exercise. Don't worry about the grades!**

1. How many times have you _____ to Disneyland?
2. In Japan, people under 18 are _____ to buy alcohol.
3. The doctor says I _____ stay at home and take some medicine.
4. People are _____ to play ball games in the park.
5. You _____ ride your bicycle here.
6. Eating a lot of fruit and vegetables is very _____
7. I started to _____ tennis three years ago.
8. You _____ wear a tie if you don' t want to.
9. You are _____ to use your hands in soccer if you are not the goalkeeper.
10. You are _____ to touch the animals.
11. In soccer, you are _____ to use your feet and your head.

What are the missing words?

- Write the missing word or words in the spaces.
  空欄に足りない単語を記入してください
- Try to answer all of the questions.
  すべての質問に答えよう
- You have 4 minutes to complete the questions
  4 分間で質問に答えてください。

1.		How many times have you _____ t Disneyland?
2.		In Japan, people under 18 are _____ to bu alcohol.
3.		You _____ ride your bicycle here.
4.		You _____ wear a tie if you don' t want to.
5.		Eating a lot of fruit and vegetables is very _____

Appendix C – Modals Speaking Task

Talking about School Rules

- Make sentences about the pictures below.
- Record your speech on Loilonote.
- Use the words and ideas from the boxes to help.
- Try to speak for about one and half minutes.

Picture 1

Schools often have many rules. What rules do students have to follow so that they will not get into trouble?

- **Use the words and ideas from the boxes to help.**

School uniform
Homework
Part-time job
Fighting
On time
School club
Mobile phone / cell phone
Bento

Talking about School Rules

- Make sentences about the pictures below. (下の写真について文章を作りなさい。)
- Record the sentences on Loilonote. (ロイロノートに文章を記録する)
- Use the words and ideas from the boxes to help. (箱の中にある言葉やアイデアを役立てよう。)
- You need to make 8 sentences. (8つの文章を作る必要がある)

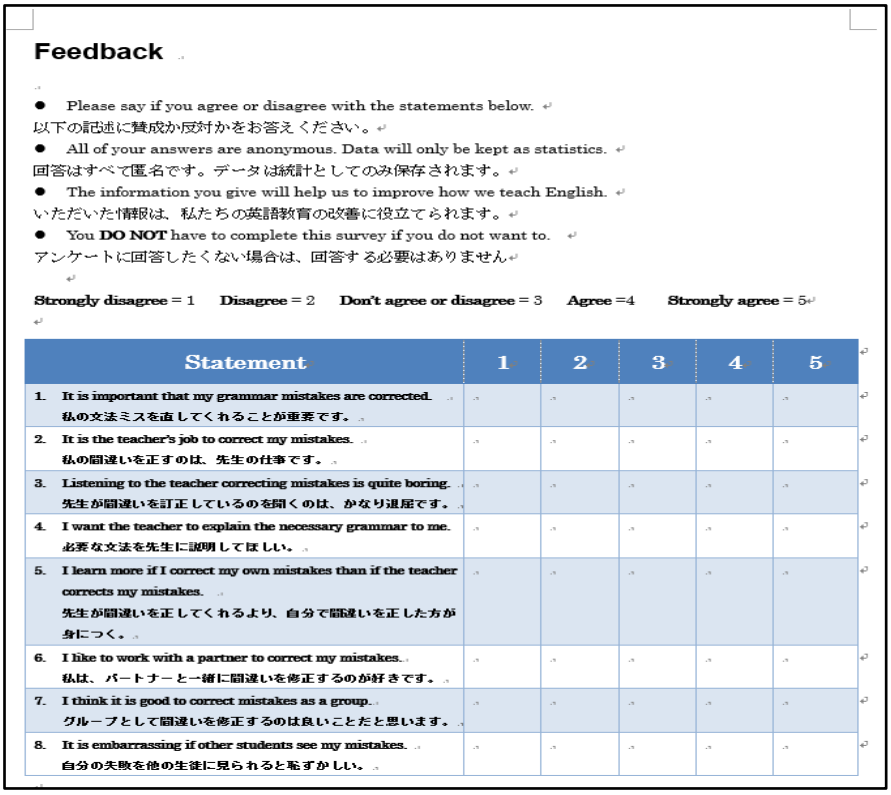
Picture 1

Schools often have many rules. What rules do students have to follow so that they will not get into trouble?

- **Use the words and ideas from the boxes to help.**

School uniform ✓
Homework ✓
Part-time job ✗
Fighting ✗
On time ✓
School club
Mobile phone / cell phone ✗
Bento

Appendix D – Students Views on E.C. - Survey



Feedback

- Please say if you agree or disagree with the statements below.

以下の記述に賛成か反対かをお答えください。

- All of your answers are anonymous. Data will only be kept as statistics.

回答はすべて匿名です。データは統計としてのみ保存されます。

- The information you give will help us to improve how we teach English.

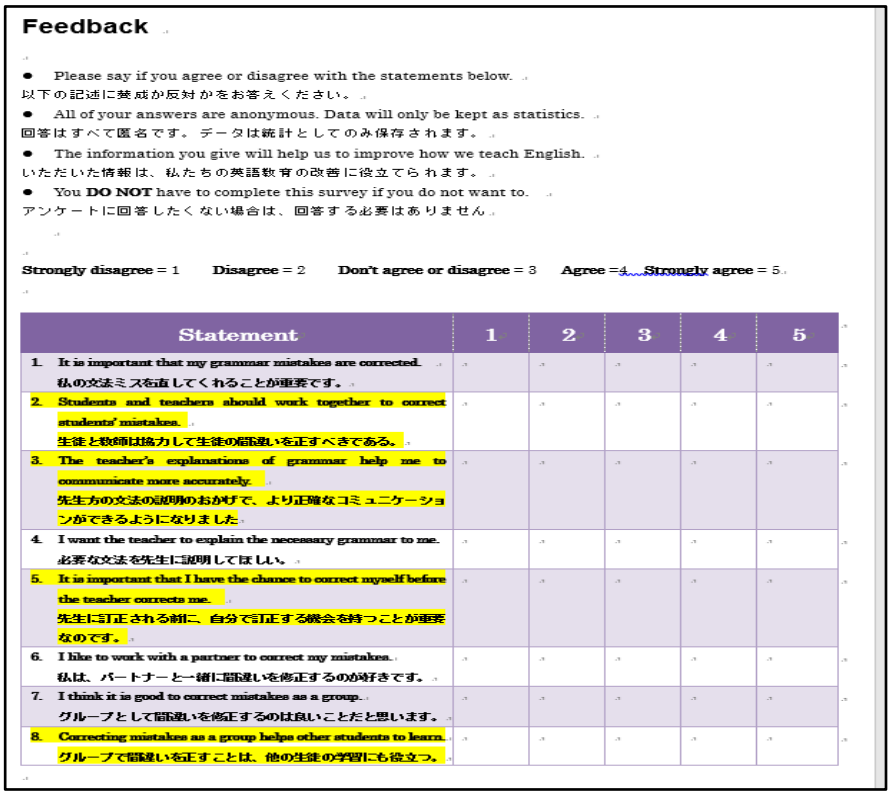
いただいた情報は、私たちの英語教育の改善に役立てられます。

- You **DO NOT** have to complete this survey if you do not want to.

アンケートに回答したくない場合は、回答する必要はありません

Strongly disagree = 1 **Disagree** = 2 **Don't agree or disagree** = 3 **Agree** =4 **Strongly agree** = 5

Statement	1	2	3	4	5
1. It is important that my grammar mistakes are corrected. 私の文法ミスを直してくれることが重要です。					
2. It is the teacher's job to correct my mistakes. 私の間違いを正すのは、先生の仕事です。					
3. Listening to the teacher correcting mistakes is quite boring. 先生が間違いを訂正しているのを聞くのは、かなり退屈です。					
4. I want the teacher to explain the necessary grammar to me. 必要な文法を先生に説明してほしい。					
5. I learn more if I correct my own mistakes than if the teacher corrects my mistakes. 先生が間違いを正してくれるより、自分で間違いを正した方が身につく。					
6. I like to work with a partner to correct my mistakes. 私は、パートナーと一緒に間違いを修正するのが好きです。					
7. I think it is good to correct mistakes as a group. グループとして間違いを修正するのは良いことだと思います。					
8. It is embarrassing if other students see my mistakes. 自分の失敗を他の生徒に見られると恥ずかしい。					



Feedback

- Please say if you agree or disagree with the statements below.

以下の記述に賛成か反対かをお答えください。

- All of your answers are anonymous. Data will only be kept as statistics.

回答はすべて匿名です。データは統計としてのみ保存されます。

- The information you give will help us to improve how we teach English.

いただいた情報は、私たちの英語教育の改善に役立てられます。

- You **DO NOT** have to complete this survey if you do not want to.

アンケートに回答したくない場合は、回答する必要はありません

Strongly disagree = 1 **Disagree** = 2 **Don't agree or disagree** = 3 **Agree** =4 **Strongly agree** = 5

Statement	1	2	3	4	5
1. It is important that my grammar mistakes are corrected. 私の文法ミスを直してくれることが重要です。					
2. Students and teachers should work together to correct students' mistakes. 生徒と教師は協力して生徒の間違いを正すべきである。					
3. The teacher's explanations of grammar help me to communicate more accurately. 先生方の文法の説明のおかげで、より正確なコミュニケーションができるようになりました。					
4. I want the teacher to explain the necessary grammar to me. 必要な文法を先生に説明してほしい。					
5. It is important that I have the chance to correct myself before the teacher corrects me. 先生に訂正される前に、自分で訂正する機会を持つことが重要なのです。					
6. I like to work with a partner to correct my mistakes. 私は、パートナーと一緒に間違いを修正するのが好きです。					
7. I think it is good to correct mistakes as a group. グループとして間違いを修正するのは良いことだと思います。					
8. Correcting mistakes as a group helps other students to learn. グループで間違いを正すことは、他の生徒の学習にも役立つ。					

Appendix E - Statistical Analysis Gap Fill Test

ANOVA Control Group

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	2829.499	9	314.3887	5.018553	0.001789	2.456281
Columns	95.52467	2	47.76233	0.762425	0.48102	3.554557
Error	1127.615	18	62.6453			
Total	4052.639	29				

ANOVA Teacher-led correction

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	4914.268	9	546.0298	8.520222	6.83E-05	2.456281
Columns	2374.766	2	1187.383	18.52787	4.27E-05	3.554557
Error	1153.554	18	64.08633			
Total	8442.588	29				

ANOVA Student-led Correction

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	9039.678	9	1004.409	23.2293	3.87E-08	2.456281
Columns	2203.631	2	1101.816	25.48207	5.62E-06	3.554557
Error	778.2995	18	43.23886			
Total	12021.61	29				

Control group and teacher-led intervention

t-Test: Two-Sample Assuming Unequal Variances

t Stat	-2.71088
P(T<=t) one-tail	0.007417
t Critical one-tail	1.739607
P(T<=t) two-tail	0.014834
t Critical two-tail	2.109816

Teacher and Student-led Intervention Groups

t-Test: Two-Sample Assuming Unequal Variances

t Stat	-0.31943
P(T<=t) one-tail	0.376539
t Critical one-tail	1.734064
P(T<=t) two-tail	0.753078
t Critical two-tail	2.100922

Appendix F - Statistical Analysis Timed Grammaticality Judgement Test

ANOVA Control Group

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	1473.119	9	163.6799	27.04575	1.13E-08	2.456281
Columns	68.36467	2	34.18233	5.64814	0.012478	3.554557
Error	108.9353	18	6.051963			
Total	1650.419	29				

ANOVA Teacher-led Intervention

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	3222.87	9	358.0966	17.61577	3.45E-07	2.456281
Columns	649.206	2	324.603	15.96812	0.000103	3.554557
Error	365.9073	18	20.32819			
Total	4237.983	29				

ANOVA Student-led Intervention

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	2159.579	9	239.9532	29.71407	5.21E-09	2.456281
Columns	356.5227	2	178.2613	22.07459	1.43E-05	3.554557
Error	145.3573	18	8.075407			
Total	2661.459	29				

Control Group and Teacher-led Intervention

t-Test: Two-Sample Assuming Unequal Variances

t Stat	-2.5625
P(T<=t) one-tail	0.010434
t Critical one-tail	1.745884
P(T<=t) two-tail	**0.02087**
t Critical two-tail	2.119905

Teacher-led and Student-led error correction

t-Test: Two-Sample Assuming Unequal Variances

	TLIPT	SLIPT
t Stat	0.812369	
P(T<=t) one-tail	0.213905	
t Critical one-tail	1.739607	
P(T<=t) two-tail	0.42781	
t Critical two-tail	2.109816	

Bibliography

Bartram, M., and Walton R. (1991). *Correction: A Positive to Language Mistakes* Heinle ELT.

Burden, T., (2011). An Investigation into the Effectiveness of the Keyword Method for a group of Japanese EFL Learners. *Journal of Regional Development Studies 75-97.*

Caine, N., (2020). Integrating active learning into EFDL Course Design. *University of Nagaski 23-39.*

Cohen, L., Manion, L. and Morrison, K. (2015). *Research Methods in Education (7th ed.)* New York: Routledge.

Conor, U., and Asenavage, K., (1994). Peer Response Groups in ESL Writing Classes: How much impact on revision. *Journal of Second Language Writing Vol. 3 No. 3 257-276.*

Creswell, J. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.)* Los Angeles, London, New Delhi, Singapore: Sage Publications.

Doody, O., and Doody, M., (2015). Conducting a pilot study: A case study of a novice researcher. *National Institutes of Health.*

Dzwigol, H., (2020). Pilot study in the research procedure. *Scientific Quarterly 'Organization and Management 1-8.*

Edge, J., (1997). *Mistakes and Correction* London, New York: Longman.

Ellis., R., (2009). Corrective Feedback and Teacher Development. *L2 Journal Vol. 1 3-18.*

Fraser, J., Fahlman, D., Arscott, J., and Guillot, I., (2018). Pilot Testing for Feasibility in a Study of Student Retention and Attrition in Online Undergraduate Programs. *International Review of Research in Open and Distributed Learning Vol. 19, No. 1.*

Fukuda, Y. (2004). Treatment of spoken errors in Japanese high school oral communication classes *(Master's thesis, California state)*.

Harmer, J., (1998). *How to Teach English: An introduction to the practice of English language teaching* Longman.

Hendrickson, J., M., (1978). Error Correction in Foreign Language Teaching: Recent Theory, Research, and Practice. *The Modern Language Journal, Vol. 62, No. 8. 387-398.*

In, J., (2017). Introduction of a pilot study. *Korean Journal of Anesthesiology. 601-605.*

Kartchava, E., (2016). Learners' Beliefs About Corrective Feedback in the Language Classroom: Perspectives from Two International Contexts. *The Canada Journal / Revue TESL Du Canada Vol. 33, No. 2 19-45.*

Katayama, A., (2007). Learners' perceptions toward oral error correction. In K. Bradford-Watts (Ed.), JALT2006 Conference Proceedings. Tokyo: JALT.

Li, S., (2010). The Effectiveness of Corrective Feedback in SLA: A Meta□Analysis. *Language Learning Vol. 60, No. 2. 309-365.*

Lightbowm, P., M., and Spada, N. (2017). *How Language Are Learned.* Oxford University Press.

Lyster, R., and Ranta, L., (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition Vol. 19, No. 1. 37-66.*

Lyster, R., and Saito, K., (2010). Oral Feedback in Classroom SLA: A Meta-Analysis. *Studies in Second Language Acquisition No. 32. 265-302.*

Miao, Y., Badger, R., and Zhen, Y., (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing. 179-200.*

Michael, J., (2006). Where's the evidence that active learning works?. *Advances in Physiology Education Vol. 30. 159-167.*

Paulus, T., M., (1999). The Effect of Peer and Teacher Feedback on Student Writing. *Journal of Second Language Writing, Vol. 8 No. 3. 265-289.*

Pawlack, M., (2014). *Error Correction in the Foreign Language Classroom: Reconsidering the Issues.* Springer.

Porter, S., R., Whitcomb, M., E., and Weitzer, W., H., (2004). Multiple Surveys of Students and Survey Fatigue *New Directions for Institutional Research No. 121. 63-73.*

Russell, J., & Spada, N. (2006). The Effectiveness of Corrective Feedback for the Acquisition of L2 Grammar: A Meta-Analysis of the Research. In J. M. Norris, & L. Ortega (eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 133-164). John Benjamins Publishing Company.

Russell, V., (2009). Corrective feedback, over a decade of research since Lyster and Ranta (1997): Where do we stand today?. *Electronic Journal of Foreign Language Teaching, Vol. 6, No. 1. 21-31.*

Yoshida, R., (2008). Teachers' Choice and Learners' Preference of Corrective Feedback Types. *Language Awareness Vol. 17, No. 1. 78-93.*

Zembytska, M., Romanova, Y., & Chumak, N. (2019). ESL students' perceptions of error correction techniques in oral production. *East European Journal of Psycholinguistics. 315-336.*

Contact email: aric.denfield@nihon-u.ac.jp