

Learner Perceptions of Peer Feedback and Writing Activities: Associations with Writing Fluency and Sentence-Rewriting Performance among Japanese High School EFL Learners

Ikuya Yuasa

Graduate School of Humanities and Social Sciences, Nagoya City University, Nagoya, Japan
Email: 198yuasa@gmail.com

How to cite this paper: Yuasa, I. (2026). Learner Perceptions of Peer Feedback and Writing Activities: Associations with Writing Fluency and Sentence-Rewriting Performance among Japanese High School EFL Learners. *Open Journal of Modern Linguistics*, 16, 459-477.
<https://doi.org/10.4236/ojml.2026.165029>

Received: September 14, 2026

Accepted: October 19, 2026

Published: October 22, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study examined how learners' perceptions of peer feedback and learning benefits were associated with two distinct outcomes among beginner-level Japanese high school EFL learners: revised-draft word counts and posttest sentence-rewriting performance, after adjustment for the corresponding initial performance. Drawing on the classroom project reported in Yuasa (2025), the study analyzed writing tasks completed under self-review and peer feedback conditions, a 30-point sentence-rewriting test, and a post-task questionnaire. Analyses included negative binomial generalized linear mixed models, exploratory factor analysis, a paired-samples t-test, and regression analyses. The questionnaire factor analysis and word-count regression included 101 learners, whereas the sentence-rewriting comparison and regression included 95 learners. An exploratory interaction analysis indicated that, after adjustment for initial word counts, perceived peer feedback effectiveness was positively associated with revised-draft word counts when perceived learning benefits were high, but the association was not significant when perceived learning benefits were low. In contrast, neither perception measure was significantly associated with posttest sentence-rewriting performance after adjustment for pretest performance. Essay word counts increased during the peer feedback task and remained above initial levels at the delayed stage, while sentence-rewriting scores increased from pretest to posttest. These findings suggest that learners' perceptions were not uniformly associated with the two writing-related outcomes examined. Fixed task order and differences in writing prompts and test forms limit causal interpretation.

Keywords

Peer Feedback, Learner Perceptions, Writing Fluency,

1. Introduction

Writing instruction has become increasingly important in Japanese secondary English education. Japan's current Course of Study emphasizes learners' ability to communicate ideas logically through productive language activities, including writing (Ministry of Education, Culture, Sports, Science and Technology [MEXT], 2018). However, writing instruction remains challenging in Japanese secondary school contexts because of practical constraints such as large class sizes and teacher workload (Hirose & Harwood, 2019). Under these conditions, peer feedback has attracted attention as an approach to L2 writing instruction, with research indicating positive effects on writing development, particularly during revision (Vuogan & Li, 2023).

From a theoretical perspective, peer interaction may create opportunities for learners to notice features of language, reflect on language use, and collaboratively address language-related problems (Swain, 2006; Sato & Ballinger, 2012). However, relatively little research in secondary EFL contexts has examined multiple dimensions of performance by combining measures of written production with separate measures of form-focused language performance. Moreover, although some studies have examined the maintenance of peer-feedback-related gains using delayed measures (e.g., Kurihara, 2017), evidence from secondary EFL contexts remains limited. Accordingly, the present study focused on the associations between learners' perceptions of peer feedback and learning benefits and two distinct outcomes: revised-draft word counts and posttest sentence-rewriting performance, after adjustment for the corresponding initial performance.

2. Literature Review

2.1. Feedback in L2 Writing

Feedback and revision are widely recognized as important components of second language (L2) writing instruction. Response to student writing may focus on various aspects of a text, ranging from vocabulary and grammar to content, organization, and reader awareness (Ferris & Hedgcock, 2014). While attention to linguistic form may support accuracy, feedback on content and organization is also important because writing involves meaning-making, not only grammatical correction. Atkinson and Tardy (2018) caution that an excessive emphasis on written corrective feedback may narrow the focus of writing instruction to grammatical accuracy and obscure the meaning-making nature of writing. These perspectives suggest the importance of feedback activities that support both form and meaning in L2 writing.

2.2. Peer Feedback and Collaborative Revision

Peer feedback has attracted attention as an alternative or complementary ap-

proach to teacher feedback. Peer feedback refers to activities in which learners read and comment on each other's writing during the revision process. Previous studies have suggested that peer feedback may promote audience awareness, critical reflection, and learner autonomy (Rollinson, 2005). A recent meta-analysis by Vuogan and Li (2023) also showed that peer feedback has a significant positive effect on L2 writing development, particularly in revision contexts.

Research has further suggested that peer feedback may benefit both writers and reviewers. Lundstrom and Baker (2009) reported that learners who provided feedback improved more in their own writing than those who mainly received feedback, suggesting that reviewing peers' texts can help learners develop evaluative awareness. In Japanese EFL contexts, Yakame (2005) found that peer comments often helped learners recognize the need to add explanations or information to their drafts, leading to content elaboration during revision. These findings suggest that peer feedback can support writing development not only through receiving comments but also through evaluating others' texts.

At the same time, effective peer feedback may require appropriate support. Berg (1999) and Min (2006) reported that trained peer response or peer review improved the quality of learners' revisions and increased the incorporation of peer comments into revised drafts. Min (2006) also noted that inexperienced learners may struggle to provide concrete and useful feedback without sufficient guidance. In addition, Allen and Katayama (2016) showed that actual and perceived relative proficiency influenced the quantity and type of feedback learners provided and incorporated. These studies suggest that peer feedback is shaped not only by task design but also by learners' confidence, proficiency perceptions, and ability to evaluate peers' writing.

2.3. Theoretical Perspectives on Peer Feedback

The effectiveness of peer feedback has often been discussed from interaction-based perspectives of language learning. Swain's (2006) Output Hypothesis suggests that producing and revising language can promote noticing gaps between intended meaning and actual expression, hypothesis testing, and metalinguistic reflection. Schmidt's (1990) Noticing Hypothesis likewise emphasizes that conscious attention to linguistic form plays an important role in second language learning. Through peer feedback, learners may become more aware of vocabulary choice, grammatical problems, unclear meanings, and weaknesses in organization while evaluating both their own and others' writing.

From a sociocultural perspective, peer interaction may also promote language awareness and collaborative scaffolding through reciprocal learning processes (Sato & Ballinger, 2012). In this view, peer feedback is not simply a process of exchanging comments but a collaborative activity in which learners jointly construct understanding of language use and text improvement. These perspectives suggest that peer feedback may contribute to writing development by supporting both written production and attention to language form and meaning.

2.4. Previous Classroom Research and the Present Study

Yuasa (2025) examined self-review and peer feedback in the same classroom project involving 110 Japanese high school students. That study focused primarily on changes in essay quality, using ChatGPT-based ratings of content and ideas, vocabulary variety, and organization, and also reported selected questionnaire responses concerning learners' experiences.

The present study uses overlapping classroom data but addresses a distinct research focus. Rather than examining essay quality itself, it investigates how learners' perceptions of peer feedback and learning benefits are associated with two different outcomes: revised-draft word counts and posttest sentence-rewriting performance, after adjustment for the corresponding initial performance. Essay word counts are treated as an indicator of writing fluency, whereas the 30-point rewrite test is interpreted as a measure of sentence-rewriting performance rather than overall grammatical accuracy. By examining these two outcomes separately, the study explores whether learner perceptions show similar or different associations across written production and sentence-level reformulation.

Research Questions:

The following research questions guided the study:

RQ1. How are learners' perceptions of peer feedback and learning benefits associated with revised-draft word counts after controlling for initial word counts?

RQ2. How are learners' perceptions of peer feedback and learning benefits associated with posttest sentence-rewriting performance after controlling for pretest performance?

3. Method

3.1. Participants and Context

The present study draws on the same classroom project and overlapping data reported in Yuasa (2025), but addresses different research questions and uses outcome measures not examined as primary outcomes in that study. The methodological information necessary to understand the present analyses is summarized below.

The participants were 110 third-year students enrolled in a public vocational high school in Japan. The students belonged to three intact classes and attended school-designated English courses that met twice a week. Although the courses primarily focused on reading activities, writing activities were incorporated during the second semester.

The participants' English proficiency was approximately at the CEFR A1 - A2 level. The study was conducted from September 2024 to February 2025 over 15 class sessions (50 minutes each). All writing, feedback, and revision activities were completed during class time, and no homework revision was assigned.

Of the 110 students initially included in the classroom project, questionnaire data were available for 101, who were included in the exploratory factor analysis. The questionnaire was administered during class through Microsoft Forms. Students who were absent or did not successfully submit a response were not followed

up for an additional administration. The nine students without questionnaire records included students who had withdrawn from the school; the number attributable to each reason could not be established.

All 101 questionnaire respondents had initial and revised-draft word counts and were included in the word-count regression. Word-count records were linked to questionnaire responses by learner ID. Of these 101 learners, six had not taken the sentence-rewriting pretest because of absence or nonparticipation and were excluded from the rewriting analyses. Although their total-score cells contained zero, these entries represented missing assessments rather than valid scores. The remaining 95 learners had both pretest and posttest data and were included in the sentence-rewriting comparison and regression.

3.2. Materials and Procedure

Materials used in this study included two writing tasks, peer-feedback worksheets, a rewrite test, and a questionnaire.

3.2.1. Writing Tasks

The study used a within-subject design in which all participants completed the self-review task first and the peer feedback task second. The same teacher taught all three classes. Both tasks followed the writing sequence and time limits described below, but feedback condition was confounded with task order and writing prompt (**Table 1**).

In Task 1 (Self-Editing: My Ideal Room), students wrote an English essay describing their ideal room for a new lifestyle. After completing the initial draft, students revised their essays individually using a simple checklist focusing on content, vocabulary, and organization.

In Task 2 (Peer Feedback: A Room for Everyone), students wrote an English essay describing a multipurpose space in a workplace. After completing the first draft, students participated in peer feedback activities in fixed groups of four or five. Each student reviewed the essays of the other group members.

Both tasks followed the same sequence: initial draft writing, revision activity, revised draft writing, and a delayed writing task. Initial and revised drafts were written within 20 minutes. Delayed tasks were administered one week later without access to dictionaries, drafts, or feedback sheets to assess written production at the delayed stage.

Table 1. Overview of the instructional procedures and data collection.

Phase	Task 1: Self-Editing	Task 2: Peer Feedback
Initial writing	Initial draft writing (20 min)	Initial draft writing (20 min)
Revision activity	Self-editing	Peer feedback activity
Revision writing	Revised draft writing (20 min)	Revised draft writing (20 min)
Delayed task	Delayed writing task	Delayed writing task
Additional measures	Rewrite test (pretest)	Rewrite test (posttest)
		Questionnaire

3.2.2. Peer Feedback Procedure

All peer feedback activities were conducted during regular class time. Peer feedback groups were formed using Task 1 final-draft scores on a 15-point scale, with the aim of reducing differences in performance across groups. Group membership remained fixed throughout the Task 2 peer feedback activities.

During the peer feedback activity, students read peers' essays and identified effective expressions as well as unclear or problematic sections. They then wrote comments and suggestions on a peer-feedback worksheet that included a checklist focusing on content, vocabulary, and organization.

After completing the written comments, students returned the worksheets to the writers and orally explained their feedback within the group.

3.2.3. Rewrite Test

The sentence-rewriting pretest was administered after all Task 1 activities had been completed and before the Task 2 peer feedback activities. The posttest was administered after the Task 2 delayed writing task. The interval between tests therefore encompassed the Task 2 instructional sequence, including peer feedback, revised-draft writing, and delayed writing, rather than peer feedback alone. No separate instruction on the pretest items or test-specific sentence-rewriting practice was provided between the two tests. The peer feedback procedures are described in Section 3.2.2. Each test consisted of 10 sentences developed from vocabulary and expressions commonly observed in the learners' essays. Learners were asked to rewrite the sentences while retaining their original meanings. The pretest concerned a personal room, whereas the posttest concerned a shared room or workplace space. Learners received only a brief oral explanation of the rewriting task; model answers and detailed explanations were not printed on the test sheets. The test items are provided in Appendix A.

Each item was scored from 0 to 3 points, yielding a maximum total score of 30. One point was awarded for lexical or phrasal substitution, two points for simplifying an expression, and three points for using a different sentence structure. Blank responses and unchanged sentences received zero points. Attempted rewrites could also receive zero points if they were not accepted as reformulations of the original sentence. For example, "My room is white walls" received zero points as a rewrite of "The room looks clean because the walls are white."

The author scored all pretest and posttest responses. A second teacher holding a TESOL qualification reviewed selected responses, including those for which scoring was uncertain, with access to the author's assigned scores. Alternative paraphrases were evaluated for preservation of the original meaning and the type of reformulation, with uncertain cases discussed with the second teacher. Both teachers were aware of whether responses came from the pretest or posttest. Because the second teacher's review did not constitute independent double scoring, inter-rater agreement was not calculated.

During item development, Microsoft Word's Flesch Reading Ease (FRE) and

Flesch-Kincaid Grade Level (FKGL) statistics were checked to help align the readability of the two test forms. Although the original output was not retained, a subsequent calculation using manually counted words and syllables yielded FRE scores of 99.48 and 94.19 and FKGL scores of 1.97 and 2.33 for the pretest and posttest, respectively. These values indicated that the two forms were broadly comparable in surface readability.

Scores were interpreted as sentence-rewriting performance within the classroom writing themes, rather than as a direct measure of overall grammatical accuracy or general rewriting ability across unrelated topics. Because the tests used different sentences, pre-post differences require cautious interpretation. The two measurements are referred to as *rewrite_pre* and *rewrite_post*, respectively.

3.2.4. Questionnaire

After all writing activities had been completed, learners responded to a questionnaire consisting of 29 items: 18 Likert-scale items, eight multiple-response items, and three open-ended questions. The Likert-scale items asked about learners' perceptions of writing activities, reading classmates' essays, and receiving peer feedback. Responses were collected using a 4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree).

The multiple-response items asked learners about revision focus, perceived improvement, use of advice, and reactions to peer feedback activities. In addition, the questionnaire included three open-ended questions asking learners about memorable advice they received, advice they gave to peers, and reflections on the peer feedback activities. Full English translations of all 18 Likert-scale items are provided in Appendix B.

3.3. Scoring and Data Analysis

Essay word count was used as an indicator of writing fluency. Word counts were calculated electronically, and misspelled words were counted if the intended meaning could be identified.

Rewrite test scores were analyzed as an indicator of sentence-rewriting performance. The scoring scheme distinguished types of reformulation and did not provide a separate measure of grammatical accuracy. A score of zero was a valid task score and was distinguished conceptually from a missing test record.

Before the exploratory factor analysis, sampling adequacy and factorability were examined using the Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity. The item distributions showed no severe univariate skewness, with skewness values ranging from -0.85 to 0.14 . Given these distributions, the four ordered response categories were treated as approximately continuous for this exploratory analysis, and maximum-likelihood extraction was used to estimate a common-factor model. Promax rotation was selected to allow the factors to correlate. All 101 respondents had complete data for the 18 items; therefore, no item-level imputation or casewise exclusion was required. Internal consistency was examined

using Cronbach's alpha and McDonald's omega.

For the regression analyses, composite scores were calculated by averaging the six items assigned to PF effectiveness perception (v12 - v17) and the eight items assigned to learning-effect perception (v2, v4, v5, and v7 - v11). These were un-weighted item means, not estimated factor scores; higher scores indicated more positive perceptions.

Because essay word counts constituted count data with overdispersion, generalized linear mixed models (GLMMs) with a negative binomial distribution were employed. Condition (Self-review vs. Peer review) and Time (Pre, Post, Delay) were included as fixed effects, and participants were treated as a random effect.

Sentence-rewriting analyses used 30-point scores for the 95 learners who had valid scores on both the pretest and posttest. Pre-post differences were examined using a paired-samples t-test. Posttest scores were regressed on pretest scores and the two questionnaire factors, with all predictors mean-centered.

To examine RQ1, hierarchical multiple regression analysis was conducted for revised-draft word counts in the peer feedback condition. Step 1 included PF effectiveness perception, learning-effect perception, and initial word count. Step 2 added all three two-way interactions among these predictors, and Step 3 added their three-way interaction. Because no directional moderation hypothesis had been specified in advance, the interaction terms were examined exploratorily. All predictors were mean-centered before the interaction terms were constructed. To examine RQ2, posttest rewrite scores were regressed on pretest scores and the two questionnaire factors, with all predictors mean-centered.

As a sensitivity analysis, the retained Step 2 model was re-estimated using standard errors clustered by the 27 peer feedback groups. A finite-sample correction was applied, and statistical tests used 26 degrees of freedom.

Responses to the three open-ended questions were reviewed and grouped into broad descriptive categories to summarize learners' retrospective accounts.

3.4. Ethical Considerations

The study protocol was reviewed and approved in August 2024 by the Research Ethics Committee of the Graduate School of Humanities and Social Sciences, Nagoya City University (approval ID: 24009). Permission to conduct the study was obtained from the participating school in accordance with the approved protocol. The classroom activities were conducted as part of regular instruction and were therefore not optional. Before the classroom data were used for research, the students received an oral explanation of the study and provided oral assent to the research use of their data. They were informed that allowing or declining the research use of their data would not affect their grades or treatment in class. Before analysis, names were replaced with identification codes, and the findings were reported without identifying individual students.

4. Results

4.1. Essay Word Count and Questionnaire Factor Analysis

4.1.1. Changes in Writing Quantity

The negative binomial GLMM showed a significant Condition $\times$ Time interaction, $\chi^2(2) = 46.83$, $p < 0.001$, indicating different patterns of change in essay word counts across the two conditions. Estimated mean word counts in the self-review condition were 53.0, 58.1, and 58.4 at the initial, revised, and delayed stages, respectively; corresponding means in the peer feedback condition were 48.6, 77.9, and 66.7, and 66.7.

Within the peer feedback condition, word counts increased from the initial to both the revised and delayed stages (both $p < 0.001$), but decreased from the revised to the delayed stage ($p < 0.001$). The initial-to-revised increase in the self-review condition was not significant ($p = 0.062$). Between-condition comparisons showed higher initial word counts for self-review ($p = 0.032$), but higher revised and delayed word counts for peer feedback (both $p < 0.001$; **Figure 1**).



Note: Error bars indicate 95% confidence intervals.

Figure 1. Estimated number of words across Time by instructional Method.

4.1.2. Exploratory Factor Analysis of the Questionnaire

The 18 questionnaire items showed adequate factorability. The Kaiser-Meyer-Olkin measure was 0.862, and Bartlett's test of sphericity was significant, $\chi^2(153) = 1034.66$, $p < 0.001$. An exploratory factor analysis (EFA) was conducted on the 18 questionnaire items using the maximum likelihood extraction method with promax rotation (Power = 4). The first and second eigenvalues were 7.046 and 2.480, respectively. Based on the scree plot and the MAP criterion, a two-factor solution was adopted.

However, four items (v1, v3, v6, and v18) showed relatively low communalities ($h^2 = 0.164, 0.163, 0.195, \text{ and } 0.175$, respectively). In addition, v1, v3, and v18 showed factor loadings below 0.40, whereas v6 showed a factor loading of 0.446 despite its low communality. Considering the stability and interpretability of the factor structure, these items were removed.

A second EFA was then conducted on the remaining 14 items (v2, v4, v5, and v7 - v17). The analysis again employed maximum likelihood extraction with promax rotation (Power = 4). The first two eigenvalues of the observed correlation matrix for the 14 items were 6.348 and 2.410. The MAP criterion supported retaining two factors. The two-factor solution obtained using maximum likelihood extraction accounted for 56.65% of the total standardized item variance, calculated from the sum of the unrounded communalities divided by 14.

The resulting pattern matrix is presented in **Table 2**. Factor 1 consisted of six items related to the perceived usefulness of classmates' comments and their contribution to improving writing (v12 - v17), with factor loadings ranging from 0.637 to 0.900. This factor was labeled PF effectiveness perception. Factor 2 consisted of eight items related to the perceived learning benefits of writing English compositions and reading classmates' writing (v2, v4, v5, and v7 - v11), with factor loadings ranging from 0.510 to 0.813. This factor was labeled learning-effect perception.

Table 2. Pattern matrix for the exploratory factor analysis of the 14 questionnaire items ($N = 101$).

Item	PF effectiveness perception (Factor 1)	Learning-effect perception (Factor 2)	h^2
v14: Peer comments helped enrich essay content	0.900		0.793
v16: Peer comments helped improve grammar and vocabulary	0.859		0.781
v15: Peer comments helped improve essay organization	0.833		0.653
v13: Peer comments were useful	0.805		0.634
v12: Peer comments were well communicated	0.777		0.516
v17: Gained something from peer comments	0.637		0.432
v8: Reading peers' essays generated more ideas		0.813	0.662
v11: Gained something from reading peers' essays		0.772	0.621
v10: Reading peers' essays helped improve grammar and vocabulary		0.769	0.697
v7: Reading peers' essays was useful		0.769	0.567
v9: Reading peers' essays helped improve essay organization		0.726	0.695
v5: Gained something from writing essays		0.681	0.381
v2: Writing essays was useful		0.531	0.218
v4: Writing essays helped review grammar and vocabulary use		0.510	0.282

Note: Extraction method = maximum likelihood; rotation method = promax rotation (power = 4). Factor loadings with absolute values below 0.40 are omitted. The factor correlation was 0.512. The two factors jointly accounted for 56.65% of the total standardized item variance, calculated using unrounded communalities. Full English translations of all 18 questionnaire items are provided in Appendix B.

The correlation between the two factors was moderate ($r = 0.512$). Internal consistency was high for both factors (Factor 1: $\alpha = 0.911$, $\omega = 0.915$; Factor 2: $\alpha = 0.885$, $\omega = 0.893$).

4.1.3. Associations between Questionnaire Factors and Writing Quantity

An exploratory hierarchical multiple regression analysis examined revised-draft word counts in the peer feedback condition ($N = 101$). Step 1 included PF effectiveness perception, learning-effect perception, and initial word count. Step 2 added all three two-way interactions among these predictors, and Step 3 added their three-way interaction. All predictors were mean-centered before constructing interaction terms.

Model R^2 values were 0.317, 0.424, and 0.427, and AIC values were 806.06, 794.81, and 796.22 for Steps 1 - 3, respectively. Step 2 significantly improved explained variance over Step 1, $\Delta R^2 = 0.107$, $\Delta F(3, 94) = 5.834$, $p = 0.001$. Step 3 provided no significant additional improvement, $\Delta R^2 = 0.003$, $\Delta F(1, 93) = 0.549$, $p = 0.461$. Step 2 was therefore retained. Among its interaction terms, only PF effectiveness perception $\times$ learning-effect perception was significant ($B = 11.611$, $p = 0.005$; **Table 3**).

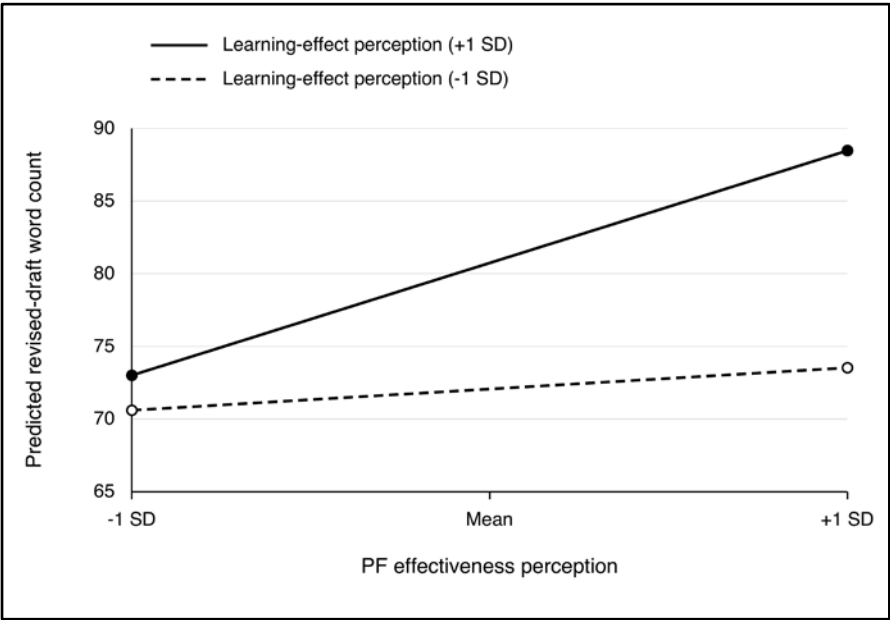
Table 3. Final regression model predicting revised-draft word counts in the peer feedback condition ($N = 101$).

Predictor	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Intercept	76.402	1.293	—	59.067	<0.001
PF effectiveness perception	8.856	2.669	0.304	3.319	0.001
Learning-effect perception	8.341	2.634	0.287	3.166	0.002
Initial word count	0.251	0.063	0.333	3.971	<0.001
PF effectiveness perception $\times$ Learning-effect perception	11.611	4.067	0.275	2.855	0.005
PF effectiveness perception $\times$ Initial word count	0.089	0.144	0.076	0.614	0.541
Learning-effect perception $\times$ Initial word count	-0.240	0.137	-0.206	-1.754	0.083

Note: $R^2 = 0.424$, Adjusted $R^2 = 0.387$, AIC = 794.81. Predictor variables were mean-centered prior to the creation of interaction terms.

Simple slopes were calculated from the retained Step 2 model, with initial word count held at its mean. PF effectiveness perception was positively associated with revised-draft word counts at high learning-effect perception (+1 *SD*; $B = 14.894$, $SE = 3.673$, $p < 0.001$), but the association was not significant at low learning-effect perception (-1 *SD*; $B = 2.817$, $SE = 3.115$, $p = 0.368$; **Figure 2**).

The sensitivity analysis produced the same coefficient estimates and showed that the focal interaction remained statistically significant after accounting for clustering within the 27 peer feedback groups ($B = 11.611$, cluster-robust $SE = 4.420$, $t(26) = 2.627$, $p = 0.014$, 95% CI [2.526, 20.696]). The simple-slope pattern was also retained. PF effectiveness perception was positively associated with revised-draft word counts at high learning-effect perception (+1 *SD*; $B = 14.894$, cluster-robust $SE = 3.529$, $p < 0.001$), but not at low learning-effect perception (-1 *SD*; $B = 2.817$, cluster-robust $SE = 2.624$, $p = 0.293$).

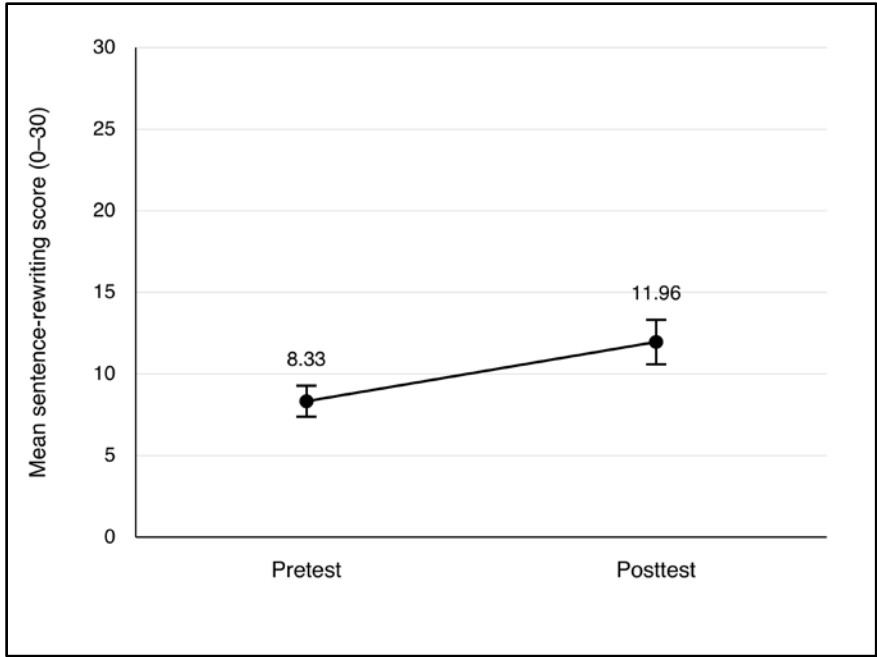


Note: Predicted values are based on the Step 2 model ($N = 101$), with initial word count held at its sample mean. Lines represent learning-effect perception at +1 SD and -1 SD; these are reference values of a continuous variable, not separate participant groups.

Figure 2. Interaction between PF effectiveness perception and learning-effect perception in predicting revised-draft word counts.

4.2. Rewrite Test Analysis

4.2.1. Changes in Sentence-Rewriting Performance



Note: Scores range from 0 to 30. Points represent observed means for the same 95 learners; error bars indicate 95% confidence intervals for each mean.

Figure 3. Mean pretest and posttest sentence-rewriting scores.

For the 95 learners analyzed, mean scores increased from 8.33 (SD = 4.64) at pretest to 11.96 (SD = 6.69) at posttest on the 30-point scale (**Figure 3**). A paired-samples *t*-test indicated a mean increase of 3.63 points, 95% CI [2.34, 4.93], $t(94) = 5.57$, $p < 0.001$.

4.2.2. Associations between Questionnaire Factors and Sentence-Rewriting Performance

The regression model was significant, $F(3,91) = 7.107$, $p < 0.001$, $R^2 = 0.190$ (adjusted $R^2 = 0.163$). Pretest scores were positively associated with posttest scores ($\beta = 0.443$, $p < 0.001$). Neither PF effectiveness perception ($\beta = 0.138$, $p = 0.203$) nor learning-effect perception ($\beta = -0.040$, $p = 0.708$) showed a statistically significant association after adjustment for the other predictors (**Table 4**).

Table 4. Multiple regression analysis predicting posttest rewrite scores.

Variable	<i>B</i>	β	<i>t</i>	<i>p</i>
Pretest rewrite score	0.638	0.443	4.592	<0.001
PF effectiveness perception	1.755	0.138	1.283	0.203
Learning-effect perception	-0.505	-0.040	-0.375	0.708

Note: $N = 95$. $R^2 = 0.190$; adjusted $R^2 = 0.163$. Predictors were mean-centered.

4.3. Supplementary Descriptive Findings

Responses to three open-ended questions were grouped into broad categories. The questions asked learners to describe memorable advice received and why it was memorable, advice they considered most useful to their partners and why, and what they wished they had done differently during the peer feedback activities. These findings reflect learners' retrospective accounts rather than direct analysis of the feedback exchanged.

The responses described advice on vocabulary, grammar, sentence reformulation, content elaboration, and organization (**Table 5**). Examples included suggesting alternative expressions, adding reasons for furniture choices, and clarifying introductions and conclusions. Reflections also identified difficulties in giving specific advice, particularly when learners lacked confidence in their grammar or vocabulary knowledge, as well as time constraints and difficulty incorporating suggestions. These accounts illustrate perceived benefits and challenges rather than independently demonstrated improvements in writing.

Table 5. Summary of themes in open-ended questionnaire responses.

Category	Illustrative response content
Vocabulary and expression choices	Suggested alternative words and expressions to convey meaning more clearly or avoid repetition.
Grammar and sentence reformulation	Described grammatical corrections, alternative sentence structures, and shortening or combining sentences.

Continued

Content elaboration	Suggested adding reasons, details, or explanations of the benefits of furniture and room arrangements.
Organization and reader awareness	Suggested clearer introductions, conclusions, sequencing, and connections between ideas.
Difficulties and desired support	Reported limited language knowledge, uncertainty about giving advice, insufficient time, and difficulty using suggestions.

Note: Entries summarize illustrative content from learners’ responses to the three open-ended questions; they are not verbatim quotations or frequency estimates. Reported improvements reflect learners’ perceptions and were not independently verified through analysis of their revisions.

5. Discussion

The present study investigated how learners’ perceptions of peer feedback and learning benefits were associated with two distinct outcomes: revised-draft word counts and posttest sentence-rewriting performance, after adjustment for the corresponding initial performance. The findings are discussed below in relation to the two research questions.

RQ1: How are learners’ perceptions of peer feedback and learning benefits associated with revised-draft word counts after controlling for initial word counts?

For RQ1, the exploratory analysis identified an interaction between the two perception measures after adjustment for initial word counts. PF effectiveness perception was positively associated with revised-draft word counts when learning-effect perception was high (+1 *SD*), whereas the association was not statistically significant when learning-effect perception was low (−1 *SD*). This pattern was retained in the sensitivity analysis using standard errors clustered by peer feedback group. The finding suggests that the association between perceived peer feedback effectiveness and written output may vary according to perceived learning benefits. Because the interaction was examined exploratorily rather than derived from a prespecified directional hypothesis, it requires confirmation in future research.

This association should be interpreted in the context of the observed word-count patterns. Essay word counts increased from the initial to the revised stage in the peer feedback condition and remained above initial levels at the delayed stage, although they declined after revision. The self-review condition showed a smaller descriptive increase. These findings indicate increased written output during the peer feedback task, with word count serving as the study’s indicator of writing fluency.

The present finding extends the descriptive questionnaire results reported in Yuasa (2025) by relating learners’ perceptions to a writing outcome not examined as a primary outcome in that study. Yuasa (2025) focused primarily on ChatGPT-based evaluations of essay quality from the same classroom project, whereas the present analysis examines written production and its association with learner perceptions. However, greater output does not necessarily indicate better writing

quality. Moreover, the present analysis examined adjusted revised-draft word counts rather than individual gain scores, and perceptions were measured after Task 2. Positive evaluations could therefore reflect successful writing experiences rather than cause greater output. Because self-review preceded peer feedback and the tasks used different prompts, the observed word-count patterns also cannot be attributed solely to peer feedback.

RQ2: How are learners' perceptions of peer feedback and learning benefits associated with posttest sentence-rewriting performance after controlling for pretest performance?

For RQ2, neither PF effectiveness perception nor learning-effect perception showed a statistically significant association with posttest sentence-rewriting performance after adjustment for pretest performance. Thus, unlike the conditional association observed for revised-draft word counts, the analysis did not establish a corresponding relationship between learners' perceptions and sentence-rewriting performance.

The 95 learners analyzed obtained higher posttest than pretest scores on the 30-point sentence-rewriting test. This difference represents change across the intervening Task 2 sequence, including peer feedback, revised-draft writing, and delayed writing, and cannot be attributed specifically to peer feedback. Nor does it establish improvement in overall grammatical accuracy: the scoring scheme distinguished types of reformulation rather than assessing grammatical accuracy separately, and the two tests used different sentences.

Taken together, the findings for the two research questions suggest that learner perceptions were not uniformly associated with different writing-related outcomes. A conditional association was observed for revised-draft word counts, whereas no statistically significant association was found for posttest sentence-rewriting performance after adjustment for pretest performance. This contrast highlights the importance of distinguishing written production from sentence-level reformulation when examining how learners' perceptions relate to writing performance.

Pedagogical Implications and Limitations:

The different patterns observed across the two outcomes suggest that learners' perceptions of peer feedback and learning benefits should be interpreted in relation to specific aspects of writing performance. The conditional association with revised-draft word counts indicates that learners' evaluations of peer feedback and the broader writing activities deserve attention when implementing peer review, although no corresponding association was found for sentence-rewriting performance. Teachers can make the purposes of giving, receiving, and using feedback explicit and provide opportunities for learners to reflect on these experiences. The present findings do not, however, establish that improving learners' perceptions will increase written output or sentence-rewriting performance.

Learners' retrospective accounts described difficulties in giving specific advice and uncertainty about their grammar and vocabulary knowledge. These reports

suggest a role for teacher modeling and guided practice in providing specific, usable comments. Teacher support may also help learners evaluate the grammatical appropriateness of suggested rewrites, although the effectiveness of such support was not tested in this study.

Several limitations constrain interpretation. First, feedback condition was confounded with task order and writing prompt, preventing isolation of the contribution of peer feedback. Future studies should counterbalance conditions across task orders and prompts. Second, word counts measure the amount of written output and do not establish broader improvements in writing quality. Third, the rewrite tests used different sentences, so score differences may partly reflect differences between test forms; the scoring scheme also did not measure grammatical accuracy separately. No additional follow-up sentence-rewriting test was administered after the posttest. The EFA treated the four-category items as approximately continuous and used maximum-likelihood extraction. Although the item distributions did not show severe univariate skewness, future research should confirm the factor structure using methods designed specifically for ordinal data and a larger sample. Finally, the single-school context limits generalizability.

6. Conclusion

This study examined how learners' perceptions of peer feedback and learning benefits were associated with two distinct outcomes: revised-draft word counts and posttest sentence-rewriting performance, after adjustment for the corresponding initial performance. An exploratory conditional association was observed for written production: PF effectiveness perception was positively associated with revised-draft word counts when learning-effect perception was high, whereas the association was not statistically significant when learning-effect perception was low.

In contrast, neither perception measure was significantly associated with posttest sentence-rewriting performance after adjustment for pretest performance. Although word counts increased during the peer feedback task and sentence-rewriting scores were higher at posttest than at pretest, these performance patterns should be distinguished from the associations examined in the two research questions. Taken together, the findings suggest that learners' perceptions were not uniformly associated with different writing-related outcomes.

Given the fixed task order and differences in writing prompts and test forms, the results should be interpreted as associations within this classroom context rather than causal evidence of peer feedback effectiveness.

Acknowledgements

The author would like to thank the students and teachers who participated in and supported the classroom project.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Allen, D., & Katayama, A. (2016). Relative Second Language Proficiency and the Giving and Receiving of Written Peer Feedback. *System*, 56, 96-106.
<https://doi.org/10.1016/j.system.2015.12.002>
- Atkinson, D., & Tardy, C. M. (2018). SLW at the Crossroads: Finding a Way in the Field. *Journal of Second Language Writing*, 42, 86-93.
<https://doi.org/10.1016/j.jslw.2018.10.011>
- Berg, E. C. (1999). The Effects of Trained Peer Response on ESL Students' Revision Types and Writing Quality. *Journal of Second Language Writing*, 8, 215-241.
[https://doi.org/10.1016/s1060-3743\(99\)80115-5](https://doi.org/10.1016/s1060-3743(99)80115-5)
- Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 Composition: Purpose, Process, and Practice* (3rd ed.). Routledge.
- Hirose, K., & Harwood, C. (2019). Factors Influencing English as a Foreign Language (EFL) Writing Instruction in Japan from a Teacher Education Perspective. In L. Seloni, & S. Henderson Lee (Eds.), *Second Language Writing Instruction in Global Contexts: English Language Teacher Preparation and Development* (pp. 71-90). Multilingual Matters.
<https://doi.org/10.21832/9781788925877-008>
- Kurihara, N. (2017). Do Peer Reviews Help Improve Student Writing Abilities in an EFL High School Classroom? *TESOL Journal*, 8, 450-470. <https://doi.org/10.1002/tesj.282>
- 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, 30-43.
<https://doi.org/10.1016/j.jslw.2008.06.002>
- Min, H. T. (2006). The Effects of Trained Peer Review on EFL Students' Revision Types and Writing Quality. *Journal of Second Language Writing*, 15, 118-141.
<https://doi.org/10.1016/j.jslw.2006.01.003>
- Ministry of Education, Culture, Sports, Science and Technology (2018). *Kōtō Gakkō Gakushū Shidō Yōryō (Heisei 30-Nen Kokujī) Kaisetsu: Gaikokugo-Hen, Eigo-Hen [Commentary on the Courses of Study for Upper Secondary Schools (2018 Notification): Foreign Languages and English]*. https://www.mext.go.jp/content/1407073_09_1_2.pdf
- Rollinson, P. (2005). Using Peer Feedback in the ESL Writing Class. *ELT Journal*, 59, 23-30. <https://doi.org/10.1093/elt/cci003>
- Sato, M., & Ballinger, S. (2012). Raising Language Awareness in Peer Interaction: A Cross-Context, Cross-Methodology Examination. *Language Awareness*, 21, 157-179.
<https://doi.org/10.1080/09658416.2011.639884>
- Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11, 129-158. <https://doi.org/10.1093/applin/11.2.129>
- Swain, M. (2006). Language, Agency and Collaboration in Advanced Second Language Proficiency. In H. Byrnes (Ed.), *Advanced Language Learning: The Contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.
- Vuogan, A., & Li, S. (2023). Examining the Effectiveness of Peer Feedback in Second Language Writing: A Meta-Analysis. *TESOL Quarterly*, 57, 1115-1138.
<https://doi.org/10.1002/tesq.3178>
- Yakame, H. (2005). The Role of Peer Feedback in the EFL Writing Classroom. *ARELE: Annual Review of English Language Education in Japan*, 16, 101-110.
- Yuasa, I. (2025). An Empirical Study Comparing Peer Feedback and Self-Revision in Japanese High School English Writing. *JACET Chubu Journal*, 23, 101-116. (in Japanese)
https://doi.org/10.34545/jacetchubu.23.0_101

Appendix A

Sentence Rewriting Test Items

Both forms asked learners to paraphrase the following sentences. Model answers and detailed explanations were not printed on the distributed test sheets. Brief oral instructions were provided. Each item was scored from 0 to 3 points, for a maximum total of 30.

Pre Test

1. There is a large bed and a computer in my room.
2. The room looks clean because the walls are white.
3. There are a lot of books on my bookshelf.
4. The desk is next to the window, so I can see outside.
5. The bed is very soft, and I can sleep well.
6. I have a chair near my desk, and I sit there to study.
7. The room is small, but it is comfortable.
8. I chose blue curtains because I like the color blue.
9. The closet is big, so I can put all my clothes in it.
10. I use my desk to study and play games.

Post Test

1. This room has space to work and rest.
2. The walls are green to help people focus.
3. There are plants on the shelves to make it look nice.
4. The windows are big and let in sunlight.
5. There is a soft sofa for taking breaks.
6. The coffee machine is near the table for easy use.
7. The room is small, but it feels cozy for everyone.
8. I picked soft chairs for sitting a long time.
9. The whiteboard is used for writing ideas.
10. The round table is good for group talks or meetings.

Appendix B

Questionnaire Items

The following are English translations of the original Japanese questionnaire items. Responses were recorded on a four-point scale from 1 (strongly disagree) to 4 (strongly agree).

Item	English translation
v1	I enjoyed writing English essays.
v2	I felt that writing English essays was useful.
v3	Writing English essays helped me generate more ideas.
v4	Writing English essays helped me review my use of grammar and vocabulary.

Continued

v5	I gained something from writing English essays.
v6	I enjoyed reading my classmates' English essays.
v7	I felt that reading my classmates' English essays was useful.
v8	Reading my classmates' English essays helped me generate more ideas.
v9	Reading my classmates' English essays helped me improve the organization of my own essays.
v10	Reading my classmates' English essays helped me improve my grammar and vocabulary.
v11	I gained something from reading my classmates' English essays.
v12	I thought that my classmates communicated their written comments on my English essays well.
v13	I felt that my classmates' written comments were useful.
v14	My classmates' written comments helped me enrich the content of my English essays.
v15	My classmates' written comments helped me improve the organization of my English essays.
v16	My classmates' written comments helped me improve the grammar and vocabulary in my English essays.
v17	I gained something from my classmates' written comments.
v18	Did you become more aware of how your English essays would come across to readers?

Note: Items v1, v3, v6, and v18 were excluded from the final exploratory factor analysis. PF effectiveness perception comprised v12 - v17; learning-effect perception comprised v2, v4, v5, and v7 - v11.