

Original Article

University Students' Engagement with Peer Feedback: The Dissociation Between Revision Completeness and Reflection

Hee-Joo Im¹, Eunkyung Hwang^{2*}

¹ Professor, College of Liberal Arts, Dankook University, Korea (heejooim@dankook.ac.kr)

² Professor, Bangmok College of General Education, Myongji University, Korea (hwangek@mju.ac.kr)

*Corresponding Author : hwangek@mju.ac.kr

ABSTRACT

Objectives: This study reconceptualizes peer feedback effectiveness in a college general English writing activity through a three-dimensional engagement framework (behavioral, cognitive, affective), moving beyond whether feedback is taken up to how learners engage with it and what quality of revision results. **Methods:** Data were collected in the spring semester of 2026 (first drafts in Week 2; peer feedback and revision in Week 6). Paired first and second drafts underwent close comparative reading to code the completeness of each actionable feedback unit (successful, partial/incomplete, or no revision) and the revision operation involved (surface vs. meaning changes); each unit was initially coded with a large language model (LLM)-assisted procedure and exhaustively verified by the researcher against both drafts. Cognitive and affective engagement were captured from written reflections and post-course attitude items; the analyses covered 981 codable units from 231 students. **Results:** A striking dissociation emerged: the aspects of feedback students said they understood or acted upon in reflection were statistically unrelated to what they actually completed in revision. Meaning (content) revisions were significantly less likely to be completed than surface (form) revisions, and no attitude item significantly predicted completeness. Case analyses identified three patterns: aligned execution, aligned selective rejection, and reflection-execution dissociation. Whether feedback was revised to completion depended far more on the individual student than on proficiency level or feedback type; nearly half of the students who reflected most actively were nonetheless low completers. **Conclusions:** The findings caution against using reflection or self-report as proxies for revision quality and call for pedagogy that scaffolds the translation of engagement into completed, accurate revision.

Keywords: L2 writing, Learner engagement, Peer feedback, Reflection, Revision quality

Received 09 Jul 2026

Accepted 26 Jul 2026

Citation: Im HJ, Hwang E. University students' engagement with peer feedback: The dissociation between revision completeness and reflection. The Journal of Transdisciplinary Studies. 2026; 10(2): 139-152. <https://doi.org/10.22685/jts.2026.10.2.139>

Copyright: © 2026 by The Journal of Transdisciplinary Studies. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>)

1. Introduction

After completing peer feedback activities as part of in-class writing tasks in college general English courses, learners commonly report that they understood the feedback and corrected the points that were flagged, and uptake rates and satisfaction are typically high as well. Yet to what extent do such learner statements guarantee actual improvement of the text? This study begins with that question. Recent feedback scholarship understands feedback not as the transmission of information provided by the teacher but as a process in which learners interpret

and evaluate it and translate it into revision behavior [1-3]. This translation process has been elaborated into three dimensions of engagement: a behavioral dimension of converting feedback into actual revision, a cognitive dimension of understanding feedback and judging its validity, and an affective dimension of attitudes and emotions toward feedback [4,5]. Peer feedback, which involves learners as readers, evaluators, and revisers, is a prototypical context in which these three dimensions operate simultaneously [6,7].

The problem is that whether these three dimensions actually move together has not been sufficiently verified.

Two questions are intertwined here. First, does the fact that feedback was “taken up” mean that the writing was actually “improved”? In this study, uptake denotes whether a learner makes any observable attempt to act on a feedback point, whereas revision completeness denotes whether that attempt actually resolves the problem identified; the two need not coincide. Even when learners attempt to incorporate feedback, they may do so incompletely, introduce new errors in the process, or revise only at the surface level [8,9]. Second, does the engagement that learners report in reflection correspond to their actual revision behavior? Reflection is commonly taken as evidence of learning, but whether the understanding and execution articulated in reflection are actually realized in the text has rarely been verified. As long as these two questions remain unanswered, the effectiveness of peer feedback activities may be overestimated even when high uptake rates and satisfaction are reported.

In the Korean university general English context as well, peer feedback has been widely employed as part of process-oriented writing instruction [10-13]. Existing studies, however, have focused mainly on learner perceptions, feedback types and frequencies, and changes in writing scores; studies that verify at the unit level what quality of revision individual feedback points are actually converted into, and whether learners’ reflections on that process correspond to the actual textual changes, are difficult to find. Accordingly, this study integrates the first and second drafts, peer feedback comments, students’ overall written reflections, and post-course surveys produced in the peer feedback activities of a university general English course, measures the three dimensions of engagement from mutually independent data sources, and empirically examines their congruence. In sum, the present study differs from prior peer feedback research in two respects: it replaces the binary uptake criterion with a completeness criterion applied to every actionable feedback unit, and it directly tests, with independently sourced measures, whether the three dimensions of engagement actually cohere.

The research questions of this study are as follows.

- RQ1. At which revision operation levels (surface vs. meaning) do revisions in response to peer feedback occur, and how does revision completeness differ by operation and feedback type?
- RQ2. How are learners’ cognitive engagement (reflection) and affective engagement (attitudes) related to actual revision completeness?
- RQ3. Into what types can learners be classified by their reflection and revision completion, and what role do individual differences play in the variance of revision completeness?

2. Literature review

2.1. Peer feedback research in L2 writing

Peer feedback, or peer response, is a collaborative activity in which learners exchange comments and revision suggestions on one another’s writing as readers and evaluators; with the spread of process-oriented approaches to writing, it has become established as a core activity in L2 writing classrooms [6]. Although peer feedback carries less authority than teacher feedback, it holds distinctive pedagogical value in that it provides responses from real readers, fosters learners’ audience awareness and self-monitoring, and creates opportunities for negotiation and interaction [14-16].

Empirical research on the effectiveness of peer feedback has also steadily accumulated. Villamil and de Guerrero [17] showed that peer revision activities contributed substantively to the improvement of final drafts, and Paulus [18] observed that while both peer and teacher feedback triggered revision, revisions prompted by peer feedback could be relatively skewed toward the surface level. Min [19] reported that systematic peer feedback training improved both the quantity and quality of revisions, and Lundstrom and Baker [20] demonstrated that the act of giving feedback contributed to the writer’s own writing development no less than receiving it. Furthermore, the meta-analysis by Double, McGrane, and Hopfenbeck [21] confirmed that peer assessment has an overall positive effect on academic performance, and Vuogan and Li’s [22] meta-analysis specific to L2 writing likewise reported that peer feedback has a significant effect on writing performance, with the effect moderated by training and by how the feedback is processed. In the Korean university context as well, studies have reported on learner perceptions of peer feedback [10,11], analyses of feedback types [13], and effects on writing quality [12,23].

In these studies, however, the incorporation of feedback was generally measured as a binary of uptake or as revision frequency, and effectiveness was proxied by writing scores or quantitative indices of the written product. Reviewing peer feedback research from 2005 to 2014, Yu and Lee [7] recommended that the focus of inquiry be expanded from the surface acts of providing and receiving feedback to the processes by which learners engage with feedback and the sociocultural contexts of that engagement. That is, research must move beyond asking “how much was taken up” to asking “how learners engaged, and what quality of revision resulted.” Extending this recommendation, the present study places at the center of its analysis the two issues that the uptake binary obscures: revision completeness and the congruence among the dimensions of engagement.

2.2. Three dimensions of learner engagement with feedback

The multidimensional model of learner engagement was first proposed by Ellis [4] as a framework for integrating research on oral and written corrective feedback and was subsequently elaborated for the written feedback context by Han and Hyland [5]. Behavioral engagement refers to revision actions taken in response to feedback and the observable strategy use that accompanies them; cognitive engagement refers to attention to feedback, depth of understanding, and metacognitive judgment; and affective engagement refers to emotional responses to feedback and attitudinal change. The key insight of this framework is that the three dimensions do not necessarily move together. In Han and Hyland's [5] case analyses, learners displayed different combinations of engagement even under identical feedback conditions: some learners were affectively receptive yet cognitively shallow in their processing, and the reverse also occurred.

Subsequent studies have repeatedly observed such asymmetries across dimensions. Zheng and Yu [24] showed that low-proficiency EFL learners engaged actively with teacher written feedback in behavioral and affective terms while their cognitive processing remained limited, such that revision behavior did not necessarily entail deep understanding. Zhang and Hyland [25], comparing engagement with teacher feedback and automated feedback, reported that the quality of engagement varied by feedback source and by learner, and Fan and Xu [26] revealed that in peer feedback contexts the dual role of receiver and provider complicates engagement patterns.

More recent studies have begun to test the relationship between engagement and execution in larger datasets. Yu, Zhang, Zheng, Yuan, and Zhang [27] showed that, in the context of peer feedback on theses, the three dimensions of engagement combined differently depending on learner and contextual factors, and Wu and Schunn [28] used large-scale data to trace how feedback features and learners' perceptions of them lead to actual revision, showing that whether feedback is understood is a key mediator. A follow-up study [29] hierarchized engagement with feedback into passive, active, and constructive levels, proposing that the "depth" of engagement differentially predicts learning outcomes. In the automated feedback context as well, Koltovskaia [30] showed that behavioral uptake may amount to no more than superficial cognitive processing, suggesting that asymmetry across dimensions is a general phenomenon that cuts across feedback sources.

Nevertheless, existing engagement research has largely relied on qualitative description of a small number of cases,

and studies that quantitatively test the congruence across dimensions (in particular, the correspondence between the cognitive engagement learners report and the behavioral outcomes actually realized in the text) in data from a larger number of learners remain rare. Cognitive engagement is typically measured through self-report such as interviews or written reflections; unless the understanding and judgments reported there are checked against textual evidence of whether they translated into completed revision, the measurement of engagement risks becoming circular. The present study addresses this gap by measuring the three dimensions of engagement from mutually independent data sources (behavioral = draft comparison; cognitive = reflection in the overall written reflection; affective = attitude survey) and directly testing their congruence.

2.3. Levels of revision operation and revision quality

Research on revision has long distinguished the levels of change that writers make to a text. Sommers [31] showed that inexperienced writers understand revision as local editing, such as replacing words and removing errors, whereas skilled writers understand it as the reworking of meaning and structure. Faigley and Witte [32] formalized this distinction as a taxonomy of surface changes and meaning changes. Surface changes are operations that do not alter the propositional content of the text, such as corrections of spelling, grammar, and form; meaning changes are operations that alter the informational structure of the text, such as adding, deleting, rearranging, and restructuring content. In cognitive process models of composing as well, reviewing is defined as a resource-intensive process of diagnosing and resolving mismatches between goals and text [33], and meaning-level revision entails a greater cognitive burden.

Studies of the relationship between feedback and revision consistently suggest that incorporation does not equal improvement. Ferris [34] showed that although a substantial proportion of teacher comments prompted revision, those revisions did not always improve the writing, and the quality of revision varied with the characteristics of the comments. Conrad and Goldstein [35] found that revision success depended less on the form of the comment than on the nature of the problem addressed (whether it was a surface problem or a meaning-level problem such as explanation or argumentation) and on student factors, and Nelson and Schunn [36] reported that specific feedback identifying the location of the problem and a solution was more likely to be implemented. Error correction research likewise shows that learners' corrections diverge into successful correction, incomplete correction, and the introduction

of new errors [8], and the need to examine the quality of feedback processing in depth within its sociocultural context has been raised [9].

The foregoing discussion indicates that analyzing the effectiveness of peer feedback requires two moves beyond the uptake binary. First, the outcome of incorporation must be evaluated in terms of completeness: whether the identified problem was actually resolved (successful revision), merely attempted (partial revision), or left unaddressed (no revision). Second, revision must be crossed with the level of operation (surface vs. meaning) to determine which levels of revision reach completion and which stall. The present study codes and analyzes these two axes at the level of individual peer feedback units.

2.4. Reflection, self-report, and actual performance

Reflection occupies a central place in process-oriented writing and self-regulated learning. In self-regulation models, self-monitoring and self-evaluation are essential steps in the cyclical loop that improves performance [37], and discussions of feedback literacy likewise emphasize learners' capacity to appraise and judge feedback and to take action on it [1,2,38]. The final link in this literacy is taking action: however sophisticated the judgment and reflection, unless they are translated into substantive improvement of the text, the feedback process remains incomplete [3].

The problem is that reflection typically takes the form of self-report. Self-report presupposes that learners can accurately observe and judge their own performance, yet this presupposition may be fragile, particularly in L2 writing contexts in which learners must diagnose their own errors. Learners may state that they "understood and incorporated" the feedback while in fact having only partially resolved, or entirely failed to resolve, the problem that was pointed out; conversely, they may have carried out revisions that their reflection never mentions. Nonetheless, reflective writing and overall written reflections are frequently used in practice as evidence of engagement and learning, and in research they serve as measures of cognitive engagement.

Warnings about this practice have also accumulated. In a systematic review of research on feedback recipience, Winstone, Nash, Parker, and Rowntree [39] pointed out the risk that observable recipience behaviors may be misused as a proxy for actual learning-related processing. Reviewing feedback literacy intervention studies, Little, Dawson, Boud, and Tai [40] likewise criticized the fact that intervention outcomes have been measured largely through self-report indicators alone, leaving little evidence of actual behavioral change. Yet studies that directly compare reflection state-

ments with actual textual change at the unit and learner levels to test the validity of the self-report presupposition remain rare, both in Korea and internationally.

3. Methods

3.1. Research context and participants

This study analyzed data from peer feedback activities conducted in the spring semester of 2026 in eight sections of a first-year-oriented general English course at a university in Seoul (initial enrollment 256, with each section capped at 32 students); the writing and peer feedback tasks were carried out during regular class sessions. In Week 2, students completed a 25-minute first draft on one of three argumentative topics. In Week 6, they read anonymously posted peer essays on the learning management system (LMS) and wrote revision-oriented feedback comments in Korean, specifying the target of revision, the reason for the suggestion, and a possible way to improve the text; each student both gave and received multiple anonymous comments. After reviewing the feedback, students wrote a second draft for approximately 20 minutes and left an overall written reflection as a free-response comment, and attitude items concerning peer feedback were administered in a post-course survey at the end of the semester.

The unit of analysis was each individual peer feedback comment a student received, restricted to actionable feedback, that is, excluding units such as praise or general reactions that could not lead to revision. When two or more peers independently pointed out the same problem (e.g., the same subject-verb agreement error), each comment was counted as a separate unit, because the analysis concerns how learners respond to each comment they receive; verbatim duplicates were rare (7 of 2,255 within-student comment pairs among the analyzed units), and the within-student dependence among units is accommodated by the student random intercept in the mixed-effects model. Participants were 231 university students, drawn from an initial enrollment of 256, who completed both in-class writing sessions and the peer feedback activity and whose peer feedback could be conclusively coded against their paired drafts. All actionable feedback units were coded in their entirety. Written informed consent for research participation was obtained from all students before the study began, and all draft, reflection, and survey data were anonymized by replacing student identification numbers with random codes before analysis. The data sources and indicators for each engagement dimension are presented in Table 1, and the research context and procedures are summarized in Table 2.

Table 1. Data sources and indicators for each engagement dimension

Engagement dimension	Data source	Indicators
Behavioral	Close comparative reading of the paired first and second drafts	Revision completeness (successful revision / partial/incomplete revision / no revision), revision operation level (surface / meaning-level), per-student completion rate
Cognitive	Post-activity overall written reflection (free response)	Coding of 12 reflection categories (e.g., elaboration of content and evidence, improvement of structure and flow, improvement of grammar and expression, self-monitoring/metacognition, feedback appraisal and selective acceptance), reflection breadth (number of reflection categories)
Affective	Post-course survey	Positive perceptions of peer feedback, helpfulness for content, burden, overall helpfulness, intention to use peer feedback in the future, etc.

Table 2. Overview of the research context and procedures

Component	Details
Setting	First-year-oriented general English course at a university in Seoul; eight class sections (spring semester of 2026)
Participants	231 students (of 256 initially enrolled) who completed both in-class writing sessions and the peer feedback activity
Week 2	25-minute first draft on one of three argumentative topics (part-time jobs, social media, place of residence)
Week 6	Anonymous peer feedback written in Korean on the LMS; each student gave and received multiple comments
After feedback review	Second draft written in approximately 20 minutes; overall written reflection posted as a free response
End of semester	Post-course survey including peer feedback attitude items
Data analyzed	981 codable actionable feedback units; 229 written reflections; 220 attitude surveys

3.2. Behavioral engagement: Coding revision completeness and revision operations

For each actionable feedback unit, the researcher conducted a close comparative reading of the paired first and second drafts and coded revision completeness. A unit was classified as successful revision if the point raised was accurately incorporated into the second draft and improved; as partial/incomplete revision if the student attempted to incorporate the feedback but did so incompletely or only in part; and as no revision if no related change could be identified. Cases in which a judgment could not be made from draft comparison alone were treated as indeterminate and excluded from analysis. Cases in which a new error arose in the course of incorporating the feedback were flagged separately. In addition, following the distinction of Faigley and Witte [32], the revision required by each feedback unit was categorized as a surface operation (grammar/sentence structure, vocabulary/expression/form) or a meaning-level operation (content/logic, organization/cohesion), with multiple types permitted.

This close-reading coding made it possible to convert into actual judgments the many feedback units that automated draft comparison (e.g., word-count change, text similarity) had left indeterminate in the preceding analysis. Ultimately, 981 of the 1,076 units (231 students) for which

a judgment could be made were used in the statistical analyses; of the remaining 95 units, 49 could not be conclusively matched to the drafts and were treated as indeterminate, and 46, though initially listed as revision suggestions, proved on close reading to be praise or general reactions and were excluded as non-actionable. Coding proceeded in two stages. In Stage 1, the researcher developed a coding manual containing operational definitions and decision rules for the four completeness codes and the revision operation levels, and large language model (Claude, Anthropic)-assisted coding was applied to the anonymized data to produce an initial judgment for each feedback unit. In this LLM-assisted step, each anonymized unit was presented to the model together with the coding manual, the full text of the peer comment, and the corresponding excerpts of the paired first and second drafts, and the model returned a provisional completeness code, a provisional operation-level code, and a brief rationale for each judgment; these outputs were treated strictly as first-pass suggestions with no evidentiary status until verified in Stage 2. In Stage 2, the initial judgments were verified in two tracks: for 495 units, the researcher reviewed every judgment by closely comparing the original peer feedback text with the first and second drafts; for the remaining 581 units, an independent blind re-coding was conducted under the same decision rules (exact agreement 87.6–89.1%;

Cohen's $\kappa = .79-.82$), and the 47 units on which the two rounds disagreed about completion were resolved through the co-researcher's independent judgment and a final check against the drafts.

All Korean-language excerpts from peer feedback comments and student reflections quoted in this article were translated into English by the author; learner English from the drafts is reproduced verbatim, including errors.

3.3. Indicators of cognitive and affective engagement

Cognitive engagement was measured using students' overall written reflections coded into 12 categories: perceived helpfulness/usefulness, recognition of weaknesses in one's own writing, improvement of grammar/expression/vocabulary, elaboration of content/evidence/examples, improvement of structure/flow/organization, reader perspective/objective self-review, execution of revisions, self-monitoring/metacognition, feedback appraisal and selective acceptance, learning from reading peers' writing, positive affect, and limitations/negative reactions. For each student, whether a given category was mentioned and the number of reflection categories (reflection breadth) were used as indicators. Affective engagement was measured with the peer feedback attitude items from the post-course survey (positive perceptions, helpfulness for content, absence of burden, overall helpfulness, intention to use peer feedback in the future, etc.; see the Appendix for the full items). Of the 231 participants, 229 provided an overall written reflection and 220 completed the attitude items; the corresponding analyses in Section 4.3 are based on these subsamples.

3.4. Analysis

The analysis proceeded in the following order. First, the distribution of revision completeness was described in cross-tabulation with revision operation level and feedback type and examined with chi-square tests. Second, for each reflection category, the corresponding completion rates were compared (Mann-Whitney U tests) with effect sizes (r) reported alongside, and Spearman correlations of reflection breadth and attitude items with completion rates were computed with 95% confidence intervals. Because multiple tests were involved, p -values were corrected with the Benjamini-Hochberg false discovery rate procedure. In addition, to examine directly whether the reflection-execution correlations were small enough to count as negligible, equivalence was assessed by checking whether the 90% confidence intervals of the correlations fell entirely within a predefined negligible range ($|\rho| < .25$); this con-

fidence-interval criterion is equivalent to two one-sided equivalence tests and can be computed from the correlation output in SPSS. Third, to identify learner types, students were classified by median splits on reflection breadth and on completion rate and cross-tabulated, and the independence of the two dimensions was tested with a chi-square test; the distribution of student-level completion rates was also examined. Finally, a mixed-effects logistic regression with a random intercept for each student estimated how much of the variance in revision completeness was attributable to individual differences. The intraclass correlation coefficient (ICC) was computed on the latent scale as the random-intercept variance divided by the total variance, $\sigma^2/(\sigma^2 + \pi^2/3)$. Quantitative results were presented alongside qualitative analyses of typical and deviant cases; for each response pattern, the first case is presented with paired draft excerpts and subsequent cases are summarized in prose. Statistical analyses were performed with SPSS Statistics 30, Python-based scripts were used for data processing, and all coding judgments and interpretations were finalized by the researcher through checks against the raw data.

4. Results

4.1. Revision operations and completeness

The first research question concerned the revision operations and the completeness of students' revisions. Of the 981 codable feedback units, 58.1% (570 units) resulted in successful revision, 29.1% (285 units) in partial/incomplete revision, and 12.8% (126 units) in no revision. The uptake rate in the broad sense (combining successful and partial/incomplete attempts) reached 87.2%, yet completed revisions that actually resolved the problems pointed out amounted to only slightly more than half. By revision operation (Table 3), completion rates were highest for grammar/sentence structure feedback, a surface operation (58.4%), and lowest for the meaning-level operations of organization/cohesion (53.4%) and content/logic (54.2%), although the descriptive gaps were modest; in the mixed-effects model reported below, a meaning-level (content) focus significantly lowered the odds of completion ($OR = 0.53$). What stood out instead was the uptake side: uptake rates were uniformly high (above 84% for every type) and highest of all for grammar/sentence structure feedback (94.2%), which therefore showed the widest gap between attempting a revision and completing it (35.8 percentage points). The differences among types thus arose not in whether students took up the feedback but in whether they carried it through to completion.

Table 3. Revision completeness by revision operation and feedback type

Operation/Feedback type	Units (<i>n</i>)	Completion rate (%)	Uptake rate (%)
Surface – grammar/sentence structure	190	58.4	94.2
Surface – vocabulary/expression/form	338	56.5	91.7
Meaning – organization/cohesion	189	53.4	88.4
Meaning – content/logic	603	54.2	84.6

Note. The uptake rate is the combined proportion of successful and partial/incomplete revisions. Because a single feedback unit could carry multiple types, the row counts sum to 1,320, exceeding the 981 codable units; 112 units whose type labels fell outside these four categories (105 “other,” e.g., spacing and formatting, and 7 praise/evaluation only) are not represented in the four rows.

A typical case in which a surface operation led to completion is as follows. Student A received a comment that flagged an unnatural connective and a subject-verb agreement error.

Peer feedback: “In the second sentence, an expression like ‘In this situation’ would sound more natural than ‘Because.’ Also, since ‘two reasons’ is plural, I think changing ‘is’ to ‘are’ would be grammatically correct.”

First draft: “Because, there is two reasons. First reason, University student is student. ... Part-time jobs has bad effect on your study.”

Second draft: “In this situation, there are two reasons. First reason, University student is a student. ... Part-time jobs have bad effect on your study.”

Student A not only accurately incorporated the two points raised (the connective expression and subject-verb agreement) but also spontaneously corrected adjacent errors that had not been pointed out (“is student” → “is a student”; “jobs has” → “jobs have”). Because surface feedback readily specifies the location of the problem and its solution, the target of revision was clearly defined and the revision tended to be carried through to completion. Student B’s case shows the same pattern. A single peer comment enumerated four local problems (the typo “essentia,” the agreement error in “it ... give,” a missing comma after “From now,” and the faulty parallelism in “from buying books or hang out”), and every one of them was corrected accurately in the second draft. In surface operations, it was thus common for multiple feedback points to be completed independently of one another.

Cases in which meaning-level operations reached completion were relatively rare, but when they occurred they involved restructuring of the text. Student C received feedback that the sentences broke off from one another and that the argument should be tied into a single flow, together with a second comment that singled out one ungrammatical sentence and modeled how it could be rewritten (“It could be revised to something like ‘Technology is making us interact more online...’”).

First draft: “Technology more and more we meet in online, more and more rude, so people confuse real and fake.” (a sentence whose subject and verb are difficult to recover)

Second draft: “Technology is making us interact more online, and people are becoming ruder, so it is becoming harder to tell what is real and what is fake.”

Student C rebuilt the garbled sentence into a coherent complex sentence along the lines the peer had modeled, and smoothed other expressions in the same draft (“*example*” → “*for instance*”), realizing the intent of the feedback (connection and intelligibility) in the text; notably, the completion was directly traceable to the explicit model supplied in the comment itself. Student D likewise received feedback that “the supporting evidence is insufficient and a concluding sentence is needed,” and in the second draft added supporting sentences (“*This is because it helps you become more independent...*”, “*For example students need to plan their time...*”) and a conclusion (“*In conclusion, having a part-time job is helpful...*”); the length of the essay more than doubled. These cases show that when meaning-level feedback is completed, it is accompanied by quantitative changes (increases in word and sentence counts); yet, as shown below, quantitative change does not in itself signify completion.

4.2. Revisions that fell short of completion: Partial uptake, surface responses, and new errors

The 29-percentage-point gap between the uptake rate (87.2%) and the completion rate (58.1%) consists of revisions that were attempted but not completed. The first type is partial/incomplete revision, in which only part of the feedback is incorporated. Student E received feedback that the body offered no grounds for its claims and needed examples or logical elaboration, specifically, concrete instances of the “powerful policies and benefits” the essay invoked. The second draft appended only a single generic sentence (“*Because the big city has opportunities to get advantages more than others.*”). There was an outward gesture of adding

support, but the concrete elaboration identified as the problem was never supplied.

The second type involves new errors arising in the course of incorporation. Student F accepted the feedback that “‘must’ is too strong” and attempted to soften “must,” but the result was a new grammatical error: “*University student should be live big city*”. Student G received an explicit rewording suggestion.

Peer feedback: “‘students’ job is study’ is grammatically awkward, so I think it would be better to change it to ‘...job is studying’.”

First draft: “*Because first, students’ job is study, not a labor.*”

Second draft: “*Because students’ job is studing, not a labor.*”

Student G understood the suggestion accurately and attempted to incorporate it, but misspelled the word (*studying* → *studing*), thereby adding an error instead. Student H, following feedback to “improve only the minor grammatical connections,” appropriately introduced connectives (*However, Therefore*), but at the same time expanded the content, and errors crept into the newly added sentences (“*If we sick tonight, we can go hospital*”, “*small-town hospitals often don’t have available at night*”). These cases show that uptake does not guarantee improvement and that, particularly under the constraints of limited linguistic resources, revision attempts can be counterproductive.

The third type is no revision, in which only a surface response remains. Student I received the feedback that adding a concrete example of “*utilize chances*” would strengthen the grounds, but no such example appeared in the second draft; the visible change near the target sentence was the insertion of the labels “*first reason*” and “*second reason*” (prompted by a different comment), while the requested example never materialized. What is noteworthy is that such marker insertion looks, on the surface, like

incorporation. Indeed, in the preceding analysis based on automated comparison of the drafts, cases of this kind were sometimes classified as candidates for uptake. Close comparative reading of the paired first and second drafts makes it possible to distinguish the appearance of incorporation from its substance, and this distinction is precisely what explains the gap between the uptake rate and the completion rate.

4.3. The dissociation between reflection and execution

The second research question concerned the relationship between learners’ self-reported cognitive and affective engagement and their actual revision completeness. The results revealed a clear dissociation (Table 4). Students who reflected on grammar/expression improvement in their overall written reflections ($n = 142$) and those who did not ($n = 59$) showed form completion rates of .54 and .56, respectively, a nonsignificant difference ($p = .765$, $r = .02$), and content completion rates likewise did not differ between students who reflected on elaborating content/evidence and those who did not (.54 vs. .49, $p = .357$, $r = .06$). Reflection on improving structure/flow was also unrelated to content completion rates ($p = .914$). No group comparison remained significant after Benjamini-Hochberg correction. The sheer amount of reflection also failed to predict execution: the correlation between the number of reflection categories (reflection breadth) and completion rate was $\rho = .01$ (95% CI $[-.12, .14]$, $p = .934$), and the correlation between self-monitoring/metacognitive reflection and completion rate was only $\rho = -.11$ ($p = .111$); the 90% confidence intervals of both correlations ($[-.10, .11]$ and $[-.21, .00]$) fell entirely within the predefined negligible range of $|\rho| < .25$, satisfying the confidence-interval criterion for equivalence. In short, statements about what a student had understood and what they had fixed were statistically unconnected to the revisions that student actually completed.

Table 4. Reflection (cognitive engagement) and actual revision completeness

Cognitive engagement (overall written reflection)	Corresponding completeness measure	Reflected	Did not reflect	p	Effect size
Grammar/expression/vocabulary improvement	Form completion rate	.54 ($n = 142$)	.56 ($n = 59$)	.765	.02
Content/evidence/example elaboration	Content completion rate	.54 ($n = 104$)	.49 ($n = 107$)	.357	.06
Structure/flow/organization improvement	Content completion rate	.51 ($n = 66$)	.52 ($n = 145$)	.914	.01
Reflection breadth (number of categories)	Completion rate	$\rho = .01$	—	.934	95% CI [$-.12, .14$]

Note. Form and content completion rates were computed only for students who received the corresponding type of feedback. Group comparisons are Mann-Whitney U tests; p -values are reported prior to Benjamini-Hochberg correction (all comparisons remained nonsignificant after correction).

With respect to affective engagement, none of the post-course attitude items was significantly correlated with completion rate: positive perceptions of peer feedback showed at most a negligible positive association ($\rho = .10$, $p = .128$), and perceived helpfulness for content, burden, overall helpfulness, and intention to use peer feedback in the future were all unrelated to completeness (all Benjamini-Hochberg corrected $q \geq .52$). None of the forms of engagement that learners reported themselves showed a reliable link to actual behavioral completion.

When the correspondence between reflection and execution is examined case by case at the student level, three patterns can be distinguished. The first is aligned execution. Student J wrote in the overall reflection, "Many of the comments said the essay would be better with an introduction, so I tried opening with a question; I used 'that' to avoid repetition; and I accepted the feedback that my conclusion was vague and added a concluding sentence." The second draft did in fact contain a question-form introduction ("What do you think about social media?") and a concluding sentence, so that statement and text were consistent (content completion rate 1.00). Student C, discussed above, is likewise a case in which the overall reflection ("there were many comments on typos and grammar in my sentence construction, and I humbly accepted them and revised") matched execution.

The second pattern is aligned selective rejection. Student K specified a judgment for each piece of feedback in the overall reflection.

Student K's overall reflection: "There was a comment that 'a positive impacts to' was grammatically wrong; since the article 'a' does not fit the plural 'impacts' and the preposition that goes with 'impact' is 'on,' I accepted this and revised it to 'positive impacts on.' ... The suggestion to put a connective like 'but' before 'there are' in the fifth sentence I did not accept, because 'But' already appears in the very next sentence."

In the actual second draft, Student K carried out precisely the revisions that had been agreed to ("positive impacts on"; spacing after sentence-final periods; "violents" → "violences") and did not insert the connective that had been explicitly declined. When reflection recorded not only uptake but also reasoned non-uptake, statement and execution

were, if anything, exactly aligned. This is a case in which the "judgment" discussed in the feedback literacy literature was at work, showing that the accuracy of reflection can serve as a cue to the quality of engagement.

The third pattern, the one that constitutes the substance of the statistical dissociation, is reflection-execution dissociation. Student L confidently reported successful revision in the overall reflection: "Most of the comments were that it was odd to wrap up the conclusion by connecting it to privacy protection; I took this on board and revised it ..., and by adding a concluding sentence at the end I made sure the essay did not stray from the overall topic." In the second draft, however, the passage in question did no more than partially restate the flow problem that had been pointed out, and none of the content feedback was completed (content completion rate .00). Student M likewise reported accepting feedback about the repetitive flow of the essay, yet the repetition problem was not revised (content completion rate .00). These students perceived themselves as having completed their revisions when the text showed otherwise, revealing a gap between perception and reality. Of the 126 students who reflected on content or structure, 50 (39.7%) fell into this dissociation (content completion rate of .34 or below), whereas 46 (36.5%) showed alignment between statement and execution (completion rate of .67 or above).

4.4. Learner types and individual differences

The third research question asked whether learners could be grouped by their reflection and their revision completion, and how large individual differences were. Students were split at the median into high and low groups on reflection breadth and on completion rate, and the two classifications were cross-tabulated (Table 5; $n = 229$). The two were statistically independent ($\chi^2(1) = 1.98$, $p = .159$, $\phi = .09$): students who reflected broadly were no more likely to complete their revisions, and 64 of the 140 students in the high-reflection half (46%) fell in the low-completion half. Individual differences were also evident in the raw completion rates ($M = .56$, $SD = .34$): 35 students completed none of the feedback they received, and 49 completed all of it. What separated learners was the quality of execution, not the amount of reflection.

Table 5. Cross-tabulation of reflection breadth and revision completion (median splits)

Reflection breadth	High completion	Low completion	Total
High (≥ 5 categories)	76	64	140
Low (< 5 categories)	39	50	89
Total	115	114	229

Note. Median splits were made at a completion rate of .57 and a reflection breadth of five categories.

Finally, a mixed-effects logistic regression asked how much of the variance in completion lay in who the student was rather than in the feedback itself. The answer was about 29% ($ICC \approx .29$; random-intercept $SD = 1.16$). Among the predictors, only a content focus mattered, roughly halving the odds of completion ($OR = 0.53$, 95% CI 0.44-0.63); self-assessed proficiency ($OR = 1.10$), form focus ($OR = 1.01$), and feedback length ($OR = 0.90$) made no significant difference (all 95% CI s included 1). The same kind of feedback could thus succeed with one student and fail with another. Engagement is better understood as configured differently from learner to learner than as operating uniformly within individuals.

5. Discussion

This study reconceptualized the peer feedback activity in a university general English course in terms of the three dimensions of engagement, each measured from an independent data source. Three findings stand out. First, feedback was widely taken up but far less often carried to completion, the shortfall consisting of partial incorporation, surface-level responses, and new errors, with meaning-level operations less likely than surface operations to be completed. Second, neither the reflections learners reported nor their attitudes were statistically related to actual completeness, and the case analyses distinguished aligned execution, aligned selective rejection, and reflection-execution dissociation. Third, completeness was governed by individual differences rather than by proficiency level or feedback type.

The first implication of these results concerns measurement. Reflective writing and self-reports are widely used in practice and research as evidence of engagement and learning, but the dissociation found in this study shows that they can hardly serve as proxy indicators of revision quality. At the level of textual evidence, this finding supports both the warning against misusing uptake behavior as a proxy for learning-related processing [39] and the criticism that the outcomes of feedback literacy interventions have been measured predominantly through self-report [40]. Several mechanisms may explain why reflection diverges from execution. Reflection may be a statement of intention or a ritualized response to the activity rather than a report of actual performance; learners may lack the monitoring ability to determine whether their revisions actually resolved the problem, particularly for meaning-level problems; or the unit of reflection, an overall comment on the activity as a whole, may correspond only loosely to individual revisions. In any case, given that reflections that concretely recorded judgments (uptake and rejection), as

in Student K's case, were aligned with execution, it is reasonable to conclude that the problem lies not in the presence of reflection but in its quality and specificity.

The second implication concerns instructional design. The link between reflection and execution needs to be made explicit. For example, requiring a post-revision reflection (revision memo) that records which feedback was addressed, in which sentence, and how, together with quotations of the sentences before and after revision, turns reflection from a statement of intention into a verification of execution. Among the instructional measures proposed here, this revision memo is the one most directly warranted by the present findings, as it targets precisely the reflection-execution link whose fragility this study documented. In addition, the finding that meaning-level revisions are harder to complete than surface revisions suggests that content and structural feedback calls for scaffolding such as modeling sentence- and paragraph-level restructuring, having students construct comment-revision correspondence charts, and targeted teacher checks. As the new-error cases show, revision attempts are intertwined with the limits of learners' linguistic resources, so providing a separate opportunity for accuracy checking at the feedback-incorporation stage is also required.

The third implication concerns individual differences. The finding that revision completeness is dominated by individual differences rather than group-level factors reveals the limitations of evaluating the effects of peer feedback activities by averages or approaching them with uniform, proficiency-based treatments. Why the individual outweighed both proficiency and feedback type calls for theoretical interpretation. From a self-regulated learning perspective [37], carrying a revision to completion requires a full monitoring-and-evaluation cycle, checking the revised text against the problem identified, and learners differ substantially in this self-regulatory capacity even at comparable proficiency levels. The feedback literacy literature [1,2,38] likewise locates the decisive competencies not in linguistic knowledge *per se* but in appraising feedback and acting on it, and Student K's aligned selective rejection illustrates precisely such literacy at work. Han and Hyland's [5] case analyses similarly showed that engagement is configured learner by learner rather than dictated by feedback conditions, and Conrad and Goldstein [35] identified student factors as decisive for revision success. Within this frame, the roughly 29% of variance attributable to the student is best read as stable between-learner differences in revision competence, the capacity to translate an understood problem into a completed textual solution, rather than as noise. Behavioral indicators such as per-student completion rates reveal learners' actual engagement more

accurately than reflective writing; used diagnostically, they could enable early identification of low-completion learners (those who perform reflection but whose execution does not follow) and the provision of individualized support. This is also one way of concretizing the learner-centered feedback design long advocated in engagement research [5,25] into an operationalizable indicator.

Finally, these findings are not confined to peer feedback. Now that generative AI can supply immediate, large-scale feedback [41], the bottleneck lies less in providing feedback than in converting it into completed revision; whatever the source of feedback, the quality of execution needs to be verified independently of self-report, and revision-completeness coding offers one operationalizable tool for doing so.

This study has the following limitations. First, although the coding was verified through the researcher's exhaustive review and an independent blind re-coding of the LLM-assisted judgments, and the units on which the two rounds disagreed were resolved by the co-researcher, independent human double coding of the full sample was not conducted; the agreement indices reported in the Methods concern automated re-coding rather than human coders, and full-scale reliability verification with independent human coders remains for future work. Second, because reflection was coded at the level of the overall written reflection on the activity as a whole and does not correspond one-to-one with individual feedback points, the dissociation reported in this study should be interpreted as a learner-level result, and re-verification with reflection data matched to the feedback unit (think-aloud protocols, stimulated-recall interviews) is needed. Third, some units could not be conclusively matched to the drafts and were treated as indeterminate (49 of the 1,076 units, 4.6%), and a further 46 units were reclassified on close reading as praise or general reactions. Fourth, this study covers a single revision cycle (first to second draft), so whether learner types and revision completeness transfer to subsequent writing requires separate longitudinal analysis. Fifth, participants' English proficiency and academic major were not examined as substantive analytic variables; although self-assessed proficiency was included as a covariate in the mixed-effects model, the possibility that objectively measured proficiency or disciplinary background moderates how learners process peer feedback and complete revisions remains a question for future research.

6. Conclusion

This study asked whether the engagement learners report and the revisions they actually complete tell the same

story, and found that they often do not. Peer feedback was widely taken up, yet only about six in ten actionable points were carried through to accurate completion, and neither the breadth of a learner's reflection nor their expressed attitudes predicted that completion. What a learner stated they had understood and corrected was statistically unconnected to what the second draft actually showed, and whether a revision reached completion depended far more on the individual learner than on proficiency level or feedback type.

The contribution of this study lies in decomposing "uptake," commonly treated as a single undifferentiated category in peer feedback research, into revision completeness and revision operation, and in measuring the behavioral, cognitive, and affective dimensions of engagement from independent data sources so that their dissociation could be demonstrated in data from a substantial number of learners. Its central lesson is that reflection and self-report cannot stand in for revision quality: the pedagogical value of peer feedback rests on designs that convert engagement into completed, accurate revision and on diagnostics that identify learners whose reflection outruns their execution. As feedback increasingly comes from peers, teachers, and AI alike, verifying the quality of execution independently of what learners report is less a refinement than a requirement, and text-based indicators of revision completeness offer one workable way to meet it.

Author Contribution

Conceptualization: Hee-Joo Im and Eunkyung Hwang
Data collection and analysis: Hee-Joo Im and Eunkyung Hwang

Writing—original draft: Hee-Joo Im and Eunkyung Hwang

Writing—review & editing: Hee-Joo Im and Eunkyung Hwang

ORCID

Hee-Joo Im: <https://orcid.org/0000-0002-3952-8190>

Eunkyung Hwang: <https://orcid.org/0000-0002-6493-7412>

Conflicts of Interest

The authors declare no conflict of interest.

References

- Carless D, Boud D. The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*. 2018;43(8):1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Molloy E, Boud D, Henderson M. Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*. 2020;45(4):527-540. <https://doi.org/10.1080/02602938.2019.1667955>
- Winstone NE, Carless D. Designing effective feedback processes in higher education: A learning-focused approach. London: Routledge; 2019.
- Ellis R. Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*. 2010;32(2):335-349. <https://doi.org/10.1017/S0272263109990544>
- Han Y, Hyland F. Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*. 2015; 30:31-44. <https://doi.org/10.1016/j.jslw.2015.08.002>
- Liu J, Hansen JG. Peer response in second language writing classrooms. Ann Arbor: University of Michigan Press; 2002.
- Yu S, Lee I. Peer feedback in second language writing (2005-2014). *Language Teaching*. 2016;49(4):461-493. <https://doi.org/10.1017/S0261444816000161>
- Ferris DR. Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In: Hyland K, Hyland F. *Feedback in second language writing: Contexts and issues*. 1st ed. Cambridge: Cambridge University Press; 2006: 81-104.
- Storch N. Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching*. 2018;51(2):262-277. <https://doi.org/10.1017/S0261444818000034>
- Kim YJ. Exploring English writing feedback employed Korean collegiate students. *The Mirae Journal of English Language and Literature*. 2013;18(1):177-205.
- Min JY. Korean university students' perceptions of feedback activities on English composition: Self-review, peer feedback, and computer feedback. *Korean Journal of English Language and Linguistics*. 2020;20:335-362. <https://doi.org/10.15738/kjell.20..202008.335>
- Ahn HS, Bang J. The impact of peer feedback on critical English writing skills. *Journal of the Korea English Education Society*. 2024;23(2):53-74. <https://doi.org/10.18649/jkees.2024.23.2.53>
- Lee YH, Kim HK. Investigation into the revision types of EFL college learners' peer feedback and their perceptions towards the peer feedback. *Studies in British and American Language and Literature*. 2021;142: 177-200. <https://doi.org/10.21297/ballak.2021.142.177>
- Hyland K, Hyland F. Feedback on second language students' writing. *Language Teaching*. 2006;39(2):83-101. <https://doi.org/10.1017/S0261444806003399>
- Mendonça CO, Johnson KE. Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*. 1994;28(4):745-769. <https://doi.org/10.2307/3587558>
- Tsui ABM, Ng M. Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*. 2000;9(2):147-170. [https://doi.org/10.1016/S1060-3743\(00\)00022-9](https://doi.org/10.1016/S1060-3743(00)00022-9)
- Villamil OS, de Guerrero MCM. Assessing the impact of peer revision on L2 writing. *Applied Linguistics*. 1998;19(4):491-514. <https://doi.org/10.1093/applin/19.4.491>
- Paulus TM. The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*. 1999;8(3):265-289. [https://doi.org/10.1016/S1060-3743\(99\)80117-9](https://doi.org/10.1016/S1060-3743(99)80117-9)
- Min HT. The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*. 2006;15(2):118-141. <https://doi.org/10.1016/j.jslw.2006.01.003>
- Lundstrom K, Baker W. To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*. 2009;18(1):30-43. <https://doi.org/10.1016/j.jslw.2008.06.002>
- Double KS, McGrane JA, Hopfenbeck TN. The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*. 2020;32(2):481-509. <https://doi.org/10.1007/s10648-019-09510-3>
- Vuogan A, Li S. Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *TESOL Quarterly*. 2023;57(4):1115-1138. <https://doi.org/10.1002/tesq.3178>
- Kim HY, Kim JN. The effect of peer feedback on writing quality in L2 college classes. *The New Korean Journal of English Language and Literature*. 2020;62(1):197-219. <https://doi.org/10.25151/nkje.2020.62.1.010>
- Zheng Y, Yu S. Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*. 2018;37:13-24. <https://doi.org/10.1016/j.asw.2018.03.001>
- Zhang ZV, Hyland K. Student engagement with teacher and automated feedback on L2 writing. *Assessing*

- Writing. 2018;36:90-102. <https://doi.org/10.1016/j.asw.2018.02.004>
26. Fan Y, Xu J. Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*. 2020;50:100775. <https://doi.org/10.1016/j.jslw.2020.100775>
 27. Yu S, Zhang Y, Zheng Y, Yuan K, Zhang L. Understanding student engagement with peer feedback on master's theses: A Macau study. *Assessment & Evaluation in Higher Education*. 2019;44(1):50-65. <https://doi.org/10.1080/02602938.2018.1467879>
 28. Wu Y, Schunn CD. From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology*. <https://doi.org/10.1016/j.cedpsych.2019.101826>
 29. Wu Y, Schunn CD. Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*. 2023;73:102160. <https://doi.org/10.1016/j.cedpsych.2023.102160>
 30. Koltovskaia S. Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*. 2020;44:100450. <https://doi.org/10.1016/j.asw.2020.100450>
 31. Sommers N. Revision strategies of student writers and experienced adult writers. *College Composition and Communication*. 1980;31(4):378-388. <https://doi.org/10.2307/356588>
 32. Faigley L, Witte S. Analyzing revision. *College Composition and Communication*. 1981;32(4):400-414. <https://doi.org/10.2307/356602>
 33. Flower L, Hayes JR. A cognitive process theory of writing. *College Composition and Communication*. 1981;32(4):365-387. <https://doi.org/10.2307/356600>
 34. Ferris DR. The influence of teacher commentary on student revision. *TESOL Quarterly*. 1997;31(2):315-339. <https://doi.org/10.2307/3588049>
 35. Conrad SM, Goldstein LM. ESL student revision after teacher-written comments: Text, contexts, and Individuals. *Journal of Second Language Writing*. 1999;8(2):147-179. [https://doi.org/10.1016/S1060-3743\(99\)80126-X](https://doi.org/10.1016/S1060-3743(99)80126-X)
 36. Nelson MM, Schunn CD. The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*. 2009;37(4):375-401. <https://doi.org/10.1007/s11251-008-9053-x>
 37. Zimmerman BJ. Becoming a self-regulated learner: An overview. *Theory Into Practice*. 2002;41(2):64-70. https://doi.org/10.1207/s15430421tip4102_2
 38. Chong SW. Reconsidering student feedback literacy from an ecological perspective. *Assessment & Evaluation in Higher Education*. 2021;46(1):92-104. <https://doi.org/10.1080/02602938.2020.1730765>
 39. Winstone NE, Nash RA, Parker M, Rowntree J. Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*. 2017;52(1):17-37. <https://doi.org/10.1080/00461520.2016.1207538>
 40. Little T, Dawson P, Boud D, Tai J. Can students' feedback literacy be improved? A scoping review of interventions. *Assessment & Evaluation in Higher Education*. 2024;49(1):39-52. <https://doi.org/10.1080/02602938.2023.2177613>
 41. Steiss J, Tate T, Graham S, Cruz J, Hebert M, Wang J, et al. Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*. 2024;91:101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
 42. Hwang E. Peer feedback perception, uptake, and writing revision patterns across self-rated English proficiency levels in a college general English class. *Journal of the Korea English Education Society*. 2026;25(3):127-146. <http://dx.doi.org/10.18649/jkees.2026.25.3.127>

[Appendix]

Post-course attitude items used in this study. The items were administered in Korean as part of the end-of-semester survey and are translated here; responses were given on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Items 1–9 were averaged to form the positive-perception index; items 3, 10, 11, and 12 served as single-item indicators of perceived helpfulness for content, absence of burden, overall helpfulness, and intention to use peer feedback, respectively. The items are reproduced here solely to document the measures used in the present analyses; the full questionnaire and item-level descriptive results are reported in a companion study based on the same corpus [42].

1. Peer feedback helped me identify the strengths and weaknesses of my essay.
2. Peer feedback showed me how my writing is understood by readers.
3. Peer feedback helped me develop the content and ideas of my writing.
4. Peer feedback helped me examine paragraph organization and the flow of my writing.
5. Peer feedback helped me discover problems I had not noticed myself.
6. The feedback my peers gave me was easy to understand.
7. The feedback my peers gave me was specific.
8. The feedback my peers gave me actually helped me revise my writing.
9. Reading my peers' essays and giving them feedback also helped my own writing.
10. The process of receiving peer feedback was not burdensome.
11. This semester's peer feedback activity helped me revise my English essays.
12. I would like to continue using peer feedback in future English writing classes.