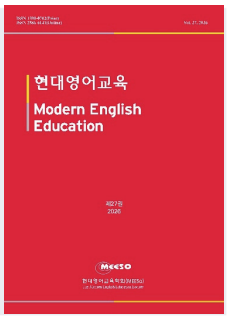
Modern English Education, Vol. 27, (2026) pp. 446-462
doi: https://doi.org/10.18095/meeso.2026.27.1.446



Modern English Education

ISSN: 1598-0782 / e-ISSN: 2586-6141
http://journal.meeso.or.kr

Selective Uptake of Peer Feedback Across Achievement Groups in a University General English Course

Eunkyung Hwang (Myongji University)

Received: 29 August 2026
Revised: 4 October 2026
Accepted: 30 October 2026

Hwang, Eunkyung. (2026). Selective uptake of peer feedback across achievement groups in a university general English course. *Modern English Education*, *27*, 446-462.

Keywords
Peer feedback, EFL writing, feedback uptake, feedback literacy
동료 피드백, 영어 쓰기, 피드백 수용, 피드백 문식력

Eunkyung Hwang
Professor
Bangmok College of General Education Myongji University
Email: hwangek@mju.ac.kr
ISNI: 0000 0004 7472 1150

Abstract

This study examined how peer comments were reflected in students' second drafts in a university general English course. The corpus comprised 1,116 comments received by 188 students, who were grouped into three levels according to their final course achievement. Each comment was traced across both drafts and coded into one of nine response categories, including direct implementation, broad uptake, possible broad uptake, and non-implementation. Content/logic comments were the most frequent comment type in all groups, and after correction for multiple comparisons, no pairwise between-group difference in the number of comments received was significant. Implementation, including instances that students extended through self-initiated revision, accounted for 62.0% of all comments and 71.1% of implementable comments, whereas broad and possible broad uptake together accounted for 8.3% of all comments. Second drafts were significantly longer than the corresponding first drafts. None of the uptake indicators differed significantly across achievement groups, although the confidence intervals were wide. An exploratory review of 66 traceable cases showed that some accurate comments were passed over and some inaccurate ones were declined; however, this subset could not indicate how often either pattern occurred. Broad uptake was concentrated in comments addressing content and organization. Because broad uptake can be attributed to specific comments only loosely, it should be reported separately from direct implementation.

INTRODUCTION

Peer feedback repositions learners in writing instruction. Rather than waiting for the teacher's evaluation, students read one another's texts as both readers and writers and look for ways to revise their own (Keh, 1990; Liu & Carless, 2006). Three decades of research have made the activity a standard component of second language (L2) writing instruction, and meta-analytic evidence supports its overall benefit (Vuogan & Li, 2023). A basic question about any single classroom nevertheless remains hard to answer. When a student receives six comments and hands in a longer, partly rewritten second draft, which comments were acted on, which were set aside, and was either decision the right one?

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

The usual answer is the uptake rate, the proportion of comments visibly incorporated into a later draft. That measure rests on three assumptions: that a comment's influence appears as a traceable local edit, that incorporation is success, and that non-incorporation is neglect. None of the three holds reliably. Peer comments are sometimes inaccurate, so incorporating one can damage a text and declining one can be sound judgment. Some comments, typically global advice about evidence or organization, are answered through indirect, text-level responses rather than a local edit. The writer adds evidence or restructures a passage without reproducing the comment's wording, so that no single revised sentence can be matched one-to-one to the suggestion. And receiving many comments, or producing a longer draft, does not in itself show that feedback was used well.

How uptake is defined therefore determines what a study can find. The choice carries practical weight in EFL general English courses, where individualized teacher feedback on every draft is rarely feasible and students differ widely in achievement. Korean students in such courses have reported positive attitudes toward using machine translation for English writing (H.-J. Im, 2017), and what a peer comment adds is the response of a real reader. Judging whether that response was used well requires a measure that separates a comment which was ignored from one that was answered indirectly or declined for good reason.

The present study traces individual peer comments from the first draft to the second in a required university general English course. It first compares the amount and types of feedback received across three course-achievement groups, providing a context for interpreting differences in uptake. It then examines which comments were implemented directly, which were associated with broader text-level revisions, and which were not implemented, and whether these response patterns differed across the groups.

LITERATURE REVIEW

Feedback and Peer Feedback in the Writing Process

Feedback ranks among the most powerful influences on learning, yet its effects depend on how learners process it (Hattie & Timperley, 2007). In process-oriented writing instruction in particular, feedback is not an evaluation of a finished product but a device for identifying problems in a draft and setting directions for revision. Keh (1990) viewed feedback as intervening in the writer's meaning construction and revision decision-making. Peer feedback extends this function beyond the teacher–student relationship to student–student relationships, and feedback on second language writing has since grown into a substantial research field in its own right (Hyland & Hyland, 2006). Throughout this study, peer feedback is understood in the sense established by Liu and Carless (2006) as a communication process in which learners enter into dialogue about the performance and standards of their writing.

Whether peer feedback leads to revision has been the field's central empirical question, and research has consistently found that it does. H.-T. Min (2005, 2006) showed that reviewer training changed the types of revisions students made and improved their revised texts, and S. Yu and Lee (2016) documented a steadily expanding research agenda over the following decade. Vuogan and Li's (2023) meta-analysis of 26 studies reported a significant positive overall effect of peer feedback on L2 writing performance, varying with factors such as task type and outcome domain, and a recent systematic review reports the same benefit alongside recurring implementation problems (Wei & Liu, 2024). The benefits also reach beyond the revised text, to a heightened sense of audience and ownership (Tsui & Ng, 2000) and to transfer on new writing (Q. Yu & Schunn, 2023).

Effectiveness, however, is not uniform across comments. Nelson and Schunn (2009), analyzing peer feedback on university students' writing, showed that specific features of a comment, such as localizing the problem and offering a concrete solution, were associated with whether the writer understood it, and that understanding was in turn associated with implementation. The practical implication is that analyses of peer feedback must distinguish local, form-focused comments from content- and structure-level comments about claims, evidence, examples, and organization, because the two differ in how readily they translate into revision.

Research in Korean universities shows a similar pattern. Domestic studies have documented Korean college students' perceptions of peer feedback alongside self-review and computer feedback (J. Y. Min, 2020), examined its effects on writing quality in first-year college writing classes (H. Y. Kim & J. Kim, 2020), and, most recently, surveyed students' perceptions of its effects on critical thinking and writing across English proficiency levels in a university critical English writing course (H. Ahn & J. Bang, 2024). As in the international literature, these studies evaluate the activity through learner perceptions, holistic writing quality, or score change. The response to each individual comment, including whether the comment was

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

accurate and whether it was implemented, has received less attention.

What this line of work leaves open is the status of the implementation judgment itself. Whether a comment that is reflected only indirectly, at the level of the whole text, should count as uptake, and what implementation is worth when the comment itself is inaccurate, remain unsettled.

Feedback Uptake, Learner Engagement, and Feedback Literacy

Uptake refers broadly to what learners do with feedback in their subsequent work (Carless & Boud, 2018). Studies of written peer feedback have usually operationalized it more narrowly as implementation, a change in the revised draft that directly follows what a particular comment suggested (Nelson & Schunn, 2009). The present study retains implementation as its core measure but also records indirect responses, in which the writer addresses the problem identified by a comment elsewhere in the text or in a different way. These two forms are referred to as direct implementation and broad uptake, and they are reported separately because a broad-uptake revision cannot be linked to a specific comment with the same certainty. Selective uptake refers to the pattern in which a writer acts on some comments but not on others. The term directs attention to which comments are taken up, rather than treating uptake as a uniform response to all feedback received.

The effectiveness of peer feedback is not determined solely by what the feedback provider says; it requires a process in which the recipient understands the feedback, trusts it, and selects and reconstructs it in line with their own writing goals. This processing has been conceptualized as engagement along cognitive, behavioral, and affective dimensions since Ellis (2010) proposed a framework for corrective feedback, and case studies of written corrective feedback have documented wide individual variation in how L2 writers engage across these dimensions (Han & Hyland, 2015). Fan and Xu (2020) applied these three dimensions in a small-sample study of intermediate-level EFL college learners and found that engagement differed by the focus of the comment. Form-focused peer feedback elicited more positive emotional engagement and more extensive behavioral and cognitive engagement than content-focused feedback did. Engagement therefore varies both across learners and across comment types.

Recent studies have accordingly reconceptualized uptake as a matter of the quality and depth of engagement rather than a binary of incorporation. Wu and Schunn (2023) proposed a revised model that orders learning from peer feedback into passive, active, and constructive levels, arguing that constructive engagement, in which learners reinterpret feedback and transform it to fit their own texts, is more closely associated with learning than verbatim incorporation. Z. Zhang and Hyland (2023) traced L2 writers' reflective journals and revising practices and showed that responding to peer feedback is a continuous process of evaluating, selecting, and recontextualizing comments. Case-study evidence indicates that what university students do with peer feedback is shaped by their beliefs, motives, and context as much as by the comments themselves (S. Yu & Hu, 2017). Together, these studies suggest that visible incorporation is only one form of learning from a comment, and not necessarily the most meaningful one.

The concept of feedback literacy frames these findings at the level of learner capability. Carless and Boud (2018) identified four features of student feedback literacy: appreciating feedback, making judgments about the quality of one's own work and the work of others, managing affect, and taking action. In these terms, declining a comment is not automatically a failure. Passing over accurate feedback may cost a revision opportunity, whereas passing over inaccurate or unnecessary feedback may reflect the kind of judgment that feedback literacy describes.

Self-regulated learning offers one interpretive lens for these choices. Models of self-regulation describe how learners plan, act, and review the effects of their strategies (Panadero, 2017; Zimmerman, 2000). Oxford's (2017) strategic self-regulation (S2R) model likewise emphasizes the selection and coordination of resources in language learning. The response categories used here may be read as behavioral patterns consistent with such regulation, although the study did not measure learners' regulatory processes directly.

Engagement and recontextualization have mostly been examined through self-report or small qualitative samples. The space between direct incorporation and non-incorporation has rarely been classified at the level of individual comments in a full course cohort. Measuring that space requires a coding framework that keeps direct and indirect responses apart. Engagement and feedback literacy also raise the question of whether uptake differs among learners. Carless and Boud (2018) describe feedback literacy as a set of capacities that students continue to develop, and they note that lower-achieving students are often relatively weak at evaluating their own performance. Students who perform better in a course may therefore be better placed to judge and act on peer comments. Proficiency has also been examined in peer feedback research. In English academic writing, Gao et al. (2023) found that the combined benefits of providing and receiving feedback for revision were stronger among higher-proficiency students, whereas lower-proficiency students benefited more from the feedback they provided. In a Korean university writing course, H. Ahn and J. Bang (2024) reported that gains in students' confidence in

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

giving and receiving peer feedback were most marked among those of intermediate proficiency. These studies compared learners on revision outcomes or self-reported perceptions. Whether students at different achievement levels receive different feedback, and whether they take up what they receive in different ways, has rarely been examined comment by comment. Course achievement is also the learner characteristic most readily available to instructors who organize peer feedback, which gives the comparison practical as well as theoretical interest.

Selective Uptake and the Quality of Peer Comments

Whether a comment is taken up depends not only on the writer but on the comment itself. Collaboration and feedback let learners discuss their texts and explore possibilities for revision (Wigglesworth & Storch, 2012), but learners may decline feedback they find too general, hard to trust, or emotionally burdensome.

Trust depends in part on whether the comment is accurate. Unlike teachers, peers are not expert evaluators, so even specific comments are not always accurate or necessary, and suggestions grounded in faulty grammatical knowledge or insufficient understanding of the context can degrade rather than improve a draft. He and Gao (2023) gave this concern an empirical footing. They analyzed 5,606 implementable idea units produced by 110 Chinese EFL students across two argumentative writing tasks. Overall feedback quality was low, with accuracy at a medium level and revision potential low, and accuracy predicted implementation more strongly and more consistently than revision potential. What learners do with feedback therefore cannot be evaluated apart from the quality of the feedback itself.

This matters for how non-implementation should be read. If some comments do not deserve to be acted on, then declining them is an act of selection rather than a failure to respond, and an implementation count alone cannot tell the two apart. Selective uptake therefore has to be examined together with the quality of the comments that were selected or declined.

The Present Study

Three strands of the literature reviewed above lead to the design of the present study. First, peer feedback research has evaluated the activity mainly through learner perceptions, holistic text quality, or score change, so the fate of individual comments has often remained out of view. Each comment therefore needs to be traced into the revised draft. Second, engagement research suggests that writers may answer a comment indirectly, at the level of the whole text. Such responses have been documented mostly through self-report and small samples, and a coding framework needs a separate category for them, kept apart from direct implementation because their link to a specific comment is less certain. Third, research on comment quality and feedback literacy implies that not every comment deserves to be acted on and that non-implementation can reflect either neglect or judgment. The capacity for such judgment may also differ among learners. If students who perform better in a course make different use of peer comments, uptake patterns should differ across course-achievement groups. Because what a writer can take up depends on what the writer receives, that comparison must first establish whether the groups received comparable feedback.

An implementation rate alone cannot show which of these responses occurred. If a high rate can include suggestions that should have been refused, and a low rate can include sound refusals or responses made at the level of the whole text, then the figure tells a teacher or a researcher less than it appears to promise. This matters particularly in required general English courses, where peer feedback runs with little teacher mediation and the rate may be the only evidence about whether the activity worked. Two research questions guide the analysis, the first providing the basis for interpreting the second.

RQ1. How much and what kinds of peer feedback do students in the three course-achievement groups receive?

RQ2. How are those comments reflected in second drafts, and do the response patterns differ by group?

METHOD

Research Design and Participants

This was a quantitative, unit-level coding study supplemented by illustrative case analysis. Feedback types and response categories were summarized by course-achievement group, and selected cases were traced across drafts to show how the response categories operated in context. The cases were not treated as a separate, systematically sampled qualitative strand.

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

The data were collected in a required general English course taken primarily by first-year university students, and informed consent was obtained before the study began. The course was offered in eight sections under a common syllabus. Students wrote a short essay on one of three argumentative prompts: whether university students should hold part-time jobs during the semester, whether social media has a positive or negative effect on real-life relationships, and whether it is better for university students to live in a big city or a small town. Of 243 students whose records were available, 188 met the inclusion criteria of writing at least 50 words in the first draft and receiving at least three peer comments, yielding 1,116 comments (High $n = 78$, Mid $n = 62$, Low $n = 48$). Within this corpus, eight comments, seven received by one mid-achievement student and one by a low-achievement student, could not be matched to the recipient's draft, indicating a matching error in the source data. These comments were coded as excluded and omitted from the implementable-response analyses.

Students receiving A+ or A0 were classified as high course achievers, those receiving B+ as mid achievers, and those receiving B0 or below as low achievers. These boundaries follow the grade bands assigned under the university's relative grading policy, which limits the proportion of students who may receive each grade. They were not derived from a theory of writing ability or feedback use, and they were not adjusted to equalize group sizes; the three groups are therefore unequal in size. The writing task itself was pass/fail and did not contribute a quality score to the final grade. Final course grades were based on attendance, midterm and final examinations, graded assignments, and regular course quizzes. The grouping variable is thus a composite of course performance and not a direct measure of writing proficiency or of the ability to use feedback. The separate 25-item diagnostic grammar quiz administered before the writing activity was ungraded and did not contribute to the final course grade. Scores on this diagnostic quiz followed the expected order across the three groups (High $M = 15.85$, $SD = 3.05$; Mid $M = 14.05$, $SD = 3.17$; Low $M = 12.25$, $SD = 3.83$; Kruskal–Wallis $H = 29.68$, $p < .001$; one score missing). The quiz thus offers limited, independent evidence that the groups already differed in grammatical knowledge before the writing activity, though the groups are still not read as writing-proficiency levels.

Of the 188 students, 75 (39.9%) were male and 113 (60.1%) were female. Most were first-year students (170, 90.4%); 15 (8.0%) were in the second year and 3 (1.6%) in the third year. Table 1 presents these characteristics by achievement group.

TABLE 1

Participant Characteristics by Achievement Group

Characteristic	High ($n = 78$)	Mid ($n = 62$)	Low ($n = 48$)	Total ($N = 188$)
Male	25	24	26	75 (39.9%)
Female	53	38	22	113 (60.1%)
First-year	76	57	37	170 (90.4%)
Second-year	1	4	10	15 (8.0%)
Third-year	1	1	1	3 (1.6%)

Data and Coding Procedures

The writing and peer feedback activities formed one component of the general English course rather than a dedicated writing course. First-draft writing, peer feedback, and second-draft writing were completed in class under instructor supervision using a common procedure across sections: first drafts were written in Week 2, and peer feedback and second drafts were completed in Week 6. AI tools and English dictionaries were not available during these sessions. Restricting these aids strengthened the attribution of textual change to peer feedback, though at a cost in realism, since students outside class routinely write with such tools at hand. The data used in this study comprised first drafts, second drafts, peer-feedback records, and final course grades. Feedback was coded as grammar/sentence structure, vocabulary/expression/mechanics, content/logic, organization/cohesion, praise/evaluation, or other; multiple coding was permitted. Within vocabulary/expression/mechanics, mechanics covers spelling, punctuation, capitalization, and word form (e.g., valueable → valuable), whereas sentence-level problems were coded as grammar/sentence structure. The corpus also supports two companion studies, one comparing self-rated proficiency groups on feedback perceptions, reported uptake, and revision patterns (E. Hwang, 2026) and the other examining revision completeness and reflection (H.-J. Im & E. Hwang, 2026).

Several terms require definition. Implementation refers to whether a peer comment is acted on in the second draft (Nelson & Schunn, 2009). Partial implementation, in which only some of the suggestions in a comment were acted on, was recorded during coding (cf. Ferris, 1997). It occurred in 112 of the 689 implemented or partially implemented comments (16.3%;

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

High 42, Mid 41, Low 29), and its share did not differ by group ($\chi^2(2) = 0.36$, $p = .835$). In the analyses, partial implementation is combined with full implementation, as in Leijen and Leontjeva (2012). Both involve a traceable change that corresponds to the comment, and whether a case counted as full or partial depended largely on how many separate suggestions the comment contained. Implementable feedback comprises comments that could in principle be acted on in writing, excluding praise and general evaluation, following He and Gao's (2023) implementable idea units.

The response scheme was derived from established unit-level analyses of peer feedback rather than constructed ad hoc. Coding implementation only for comments that request a change, and setting praise and general evaluation aside as non-implementable, follows Nelson and Schunn (2009) and He and Gao (2023). Partial implementation corresponds to Ferris's (1997) distinction between substantive and minimal revision attempts. Implementation with self-initiated revision reflects Villamil and de Guerrero's (1998) finding that writers make further self-directed revisions beyond what peers proposed. Broad uptake extends two established distinctions: Nelson and Schunn's (2009) coding of a comment's scope as global or local and of a problem as localized or not localized, and Faigley and Witte's (1981) separation of surface changes from text-base changes that alter content or structure. The label itself is introduced in the present study to name what happens when a writer addresses such a problem at the level of the text without reproducing the comment's wording. Finally, following Nelson and Schunn (2009), who set aside segments too ambiguous to code for implementation, cases with no defensible comment-to-change attribution were retained as unresolved rather than counted as non-implementation. The final response scheme comprised nine categories: implemented or partially implemented, implementation with self-initiated revision, broad uptake, broad uptake with local non-implementation, possible broad uptake, not implemented, unresolved attribution, non-implementable feedback, and exclusion for a feedback–draft mismatch.

Implementation with self-initiated revision required the core suggestion to be incorporated together with an additional change prompted by or extending that suggestion; unrelated editing did not qualify. Each decision was made with reference to the recipient's full set of comments so that a change prompted by another comment was not misclassified as self-initiated. When several comments made the same suggestion, each was credited only if the relevant change was present. Three categories mark the boundary of indirect response and were distinguished in sequence. Broad uptake was assigned when the problem a comment named, such as thin support, was answered by a substantive change, such as an added reason, an example, or a restructured passage. Possible broad uptake was assigned when the second draft contained a related change too slight to show that the problem had been addressed, for example a single restated sentence after a request for more evidence. Unresolved attribution was assigned when changes were so extensive, or several comments made overlapping general requests that could not be told apart, that no change could be linked to the comment in question, as when the whole essay was rewritten. Table 2 summarizes the operational definitions, inclusion and exclusion criteria, and verification procedures.

TABLE 2

Response Coding Categories, Operational Definitions, and Decision Criteria

Category	Operational definition	Inclusion / exclusion criteria
Implemented or partially implemented	A change in the second draft corresponds to the comment's wording or suggested repair; partial = only some of the suggestions in the comment were acted on	Include: verbatim or adapted adoption of the suggestion at the indicated location; when several peers made the same specific suggestion, each is credited if the change is present. Exclude: a change at the location that does not follow the suggestion
Implemented with self-initiated revision	The core suggestion is incorporated together with an additional change prompted by or extending it	Exclude: unrelated editing, and changes requested by another comment (checked against the recipient's full set of comments)
Broad uptake	The problem identified by the comment is answered by a substantive content, evidence, or structural change without lexical correspondence	Include: an added reason, example, or restructured passage that answers a global comment. Exclude: a slight related change (possible broad uptake); wholesale rewriting (unresolved)
Broad uptake with local non-implementation	The global part of a comment is taken up as broad uptake, but a specific local suggestion in the same comment is not	Retained in the scheme; no unit met the criterion in the final coding

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

Category	Operational definition	Inclusion / exclusion criteria
Possible broad uptake	A related change is present but too slight to show that the problem identified by the comment was addressed	Include: e.g., a single restated sentence after a request for more evidence. Exclude: a substantive change (broad uptake)
Not implemented	The comment is implementable, but no related change appears in the second draft	Include: the flagged form or passage is left unchanged. Exclude: the passage disappears in a whole-text rewrite (unresolved)
Unresolved attribution	A change may be related, but no defensible link to this comment can be established	Include: wholesale rewriting, overlapping general comments that cannot be told apart, or general advice accompanied only by diffuse minor edits. Not counted as non-implementation
Non-implementable feedback	Praise, emotional support, or general evaluation with no revision-directed content	Exclude from the implementable denominator. A comment that offers a rewritten sentence is implementable even when phrased as praise
Excluded (feedback–draft mismatch)	The comment cannot be matched to the recipient's draft (source-data error)	Excluded from all response analyses

Note. Verification: the researcher assigned every final code after examining the comment against both drafts and the recipient's full set of comments. Model-assisted coding and inter-coder agreement are reported in the text.

A large language model (Claude Fable 5; Anthropic) was used at the preparatory stage. For each student, the model received the first and second drafts, the student's full set of peer comments, and the decision rules of the coding scheme. It returned a provisional category and a short written rationale for each comment. These provisional codes served as a starting point for review and were not analyzed as data. The researcher checked all 1,116 units against both drafts, accepted, revised, or rejected each provisional code, and made the final coding decisions. To gauge how closely the final codes followed the model's suggestions, the researcher also coded a separate stratified sample of 100 units without access to the model's output. Researcher–model agreement was $\kappa = .76$ for the nine-category scheme and $\kappa = .78$ for the binary distinction between direct implementation and all other categories. These values describe the model-assisted procedure and are not an index of inter-coder reliability. Inter-coder reliability was estimated separately with an independent human second coder, who saw neither the model's output nor the researcher's codes. After a ten-unit calibration round, the second coder coded a stratified 10% sample of 112 units using a blind coding sheet. Agreement with the researcher was $\kappa = .79$ for the nine-category scheme (raw agreement 86.6%) and $\kappa = .86$ for the binary distinction. The 15 disagreements were concentrated at the boundary between direct implementation and broad uptake; the main analyses use the researcher's final codes.

The accuracy review was restricted to comments whose response was unambiguous, so that accuracy could be examined alongside a clearly traceable outcome. An initial automated draft comparison, conducted before and independently of the model-assisted coding, returned a determinate judgment for 66 of the 1,116 comments, classifying 51 as implemented or partially implemented and 15 as not implemented; for the remaining comments the screening could not establish a response without further reading. These 66 comments formed the accuracy subset. During the accuracy review each response was re-examined against both drafts; three comments initially screened as not implemented proved to be at least partially implemented, and one could not be matched to the recipient's draft, yielding, for the accuracy review, 54 implemented or partially implemented, 11 not implemented, and one excluded. Each comment was classified as accurate, inaccurate, partially problematic, or excluded for a feedback–draft mismatch according to grammatical accuracy, contextual necessity, and the effect of incorporation on meaning and grammaticality. Because the subset consists of the most easily traced comments rather than a random sample, it is used to illustrate what can follow from a writer's decision and not to estimate how often such comments occur.

Data Analysis

Group differences in student-level indicators were examined with Kruskal–Wallis tests because Shapiro–Wilk tests showed departures from normality in most group-wise distributions ($p < .001$ for implementation rate and broad-uptake counts in all groups); median-based Levene tests did not indicate unequal variances (all $p > .10$). One-way ANOVA, run as a check, led to the same conclusions except for the number of comments received (Kruskal–Wallis $p = .045$; ANOVA $p = .114$). The uptake indicators were also re-estimated with generalized linear models: negative binomial regressions for the numbers of

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

implemented and broad-uptake cases, a Poisson model with implementable comments as an offset, and a binomial logistic regression with robust standard errors for the implementation rate. Draft length was compared with Wilcoxon signed-rank tests. Chi-square tests of feedback types and response distributions are treated as exploratory because comments are nested within students. A cluster-robust generalized estimating equation was fitted as a sensitivity analysis, and 95% confidence intervals for pairwise differences in the pooled implementation rate were obtained from a stratified bootstrap of students (5,000 resamples). Broad uptake is reported separately from direct implementation throughout. All analyses were conducted in IBM SPSS Statistics (Version 30).

Student-level indicators were calculated by aggregating final unit codes for each student: the number of implementable comments, implemented cases including self-initiated revision, broad or possible broad-uptake cases, and the ratio of implemented cases to implementable feedback.

RESULTS

Sample Characteristics and Revision Patterns

Table 3 presents essay characteristics and feedback exposure for the three course-achievement groups. First-draft length, second-draft length, and word-count change did not differ significantly. Second drafts were nonetheless significantly longer than first drafts, Wilcoxon signed-rank $z = 8.83$, $p < .001$, $r = .64$, and the increase was significant in each group (all $p < .001$); 145 of the 188 students (77.1%) lengthened their drafts. Second drafts averaged 122.02 words against 105.99 for first drafts, a modest gain of about 16 words. Length shows that students revised, but not that their texts improved, since text quality was not scored. The number of comments received differed significantly across the three groups overall ($p = .045$), but no two groups differed significantly from each other after correction for multiple comparisons. The high group received 5.51 comments on average, compared with 6.31 in the mid group and 6.15 in the low group; this difference is therefore treated as exploratory.

Table 4 presents student-level uptake indicators. None of the indicators differed significantly by course-achievement group. Mean individual implementation rates were .75, .71, and .69 in the high, mid, and low groups, respectively. The median number of broad-uptake cases was 0 in all three groups, so more than half the students in each group produced none at all. At the aggregate level, 65.9–74.3% of implementable comments were implemented across the three groups. Generalized linear models led to the same conclusion, showing no group effect on the number of implemented cases (negative binomial, likelihood-ratio [LR] $\chi^2(2) = 2.70$, $p = .259$), the number of broad-uptake cases (negative binomial, LR $\chi^2(2) = 0.97$, $p = .616$), implemented cases with implementable comments as an offset (Poisson, LR $\chi^2(2) = 1.49$, $p = .475$), or the proportion of implementable comments implemented (binomial logistic with robust standard errors, Wald $\chi^2(2) = 2.39$, $p = .303$).

TABLE 3

Essay Characteristics and Feedback Exposure by Achievement Group

Variable	Total	High	Mid	Low	*p*
Number of students	188	78	62	48	–
Feedback units	1,116	430	391	295	–
Final course score	84.47 (14.22)	94.11 (2.39)	85.82 (2.48)	67.05 (17.91)	–
First-draft words	105.99 (38.09)	108.29 (37.00)	105.13 (34.48)	103.38 (44.38)	.491
Second-draft words	122.02 (41.64)	124.32 (38.58)	122.31 (44.91)	117.90 (42.60)	.463
Word-count change	16.02 (26.22)	16.03 (26.90)	17.18 (29.37)	14.52 (20.63)	.736
Feedback received	5.94 (2.38)	5.51 (2.40)	6.31 (2.50)	6.15 (2.12)	.045
Determinate response (implemented or not)	4.64 (2.33)	4.37 (2.39)	5.00 (2.46)	4.60 (2.02)	.156

Note. Standard deviations in parentheses. *p* values are from Kruskal–Wallis tests. High, Mid, and Low denote the high-, mid-, and low-achievement groups. Determinate response denotes comments whose implementation status (implemented or not implemented) could be established. For feedback received, post hoc Dunn tests with Holm correction showed no significant pairwise differences. No test is reported for final course score because it defines the groups.

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

TABLE 4
Student-Level Uptake Indicators by Achievement Group

Indicator	High (A+/A0)	Mid (B+)	Low (B0 or below)	*H*	*p*
Implementable feedback	4.91 (2.59); *Mdn* = 5.0	5.45 (2.51); *Mdn* = 5.0	5.25 (2.34); *Mdn* = 5.0	2.68	.261
Implemented (including self-initiated revision)	3.53 (2.14); *Mdn* = 3.0	4.05 (2.47); *Mdn* = 4.0	3.46 (1.90); *Mdn* = 3.0	2.69	.260
Broad uptake (incl. possible)	0.53 (0.96); *Mdn* = 0.0	0.40 (0.90); *Mdn* = 0.0	0.56 (1.01); *Mdn* = 0.0	0.95	.623
Implementation rate / implementable	.75 (.25); *Mdn* = .78	.71 (.28); *Mdn* = .75	.69 (.29); *Mdn* = .67	1.30	.523

Note. Values are *M* (*SD*); *Mdn*. *H* is the Kruskal–Wallis statistic; *df* = 2 for all tests. The implementation rate is the number of implemented cases (including self-initiated revision) divided by the number of implementable comments. Students with no implementable comments (n = 4) were assigned a rate of .00; treating them as missing left the result unchanged (H = 1.17, p = .557).

Distribution of Feedback Types

As Table 5 shows, content/logic feedback accounted for the largest share in all three groups, totaling 643 units (57.6%), followed by vocabulary/expression/mechanics (312 units, 28.0%) and praise/evaluation (222 units, 19.9%). Comments on grammar and sentence structure were comparatively few, at 192 units (17.2%), fewer than either of the two leading categories. None of the six type distributions differed significantly across course-achievement groups.

TABLE 5
Distribution of Peer Feedback Types by Achievement Group

Feedback type	Total *n* (%)	High *n* (%)	Mid *n* (%)	Low *n* (%)	*p*
Grammar/sentence structure	192 (17.2%)	71 (16.5%)	72 (18.4%)	49 (16.6%)	.734
Vocabulary/expression/ mechanics	312 (28.0%)	120 (27.9%)	117 (29.9%)	75 (25.4%)	.429
Content/logic	643 (57.6%)	259 (60.2%)	208 (53.2%)	176 (59.7%)	.089
Organization/cohesion	200 (17.9%)	81 (18.8%)	65 (16.6%)	54 (18.3%)	.697
Praise/evaluation	222 (19.9%)	86 (20.0%)	74 (18.9%)	62 (21.0%)	.792
Other	165 (14.8%)	55 (12.8%)	64 (16.4%)	46 (15.6%)	.318

Note. Because multiple coding was permitted, percentages may sum to more than 100%.

Response Distribution and Group Comparisons

Table 6 presents the final response distribution. Direct implementation was the most common outcome by a wide margin. In all, 689 comments (61.7%) were implemented or partially implemented, 112 of them (10.0% of all comments) only partially, with a further 3 (0.3%) extended by self-initiated revision. Broad uptake and possible broad uptake together accounted for 93 comments (8.3%), and 180 (16.1%) were not implemented. A further 135 comments (12.1%) were praise or general evaluation calling for no revision, 8 (0.7%) finally resisted classification, and 8 (0.7%) were excluded because of feedback–draft mismatches. The distribution did not differ significantly by course-achievement group, $\chi^2(6) = 8.00$, $p = .238$, Cramér's $V = .060$ (n = 1,100 after excluding unresolved and excluded units).

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

TABLE 6
Final Distribution of Feedback Responses

Response category	Total *n* (%)	High *n* (%)	Mid *n* (%)	Low *n* (%)	Interpretation
Implemented or partially implemented	689 (61.7%)	274 (63.7%)	249 (63.7%)	166 (56.3%)	Feedback–revision link directly confirmed
Implemented with self-initiated revision	3 (0.3%)	1 (0.2%)	2 (0.5%)	0 (0.0%)	Core suggestion incorporated, followed by additional student revision
Broad uptake	85 (7.6%)	41 (9.5%)	20 (5.1%)	24 (8.1%)	Weak direct correspondence, but feedback-related content/ structural revision
Broad uptake with local non-implementation	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	Global advice incorporated; local grammar suggestion not taken up
Possible broad uptake	8 (0.7%)	0 (0.0%)	5 (1.3%)	3 (1.0%)	Possible connection, but low certainty
Not implemented	180 (16.1%)	66 (15.3%)	59 (15.1%)	55 (18.6%)	No identifiable revision related to the comment
Unresolved attribution	8 (0.7%)	1 (0.2%)	3 (0.8%)	4 (1.4%)	Feedback–revision attribution not determinable
Non-implementable feedback (praise/general evaluation)	135 (12.1%)	47 (10.9%)	46 (11.8%)	42 (14.2%)	Praise, emotional support, or general evaluation
Excluded due to feedback–draft mismatch	8 (0.7%)	0 (0.0%)	7 (1.8%)	1 (0.3%)	Source data verification required

Note. Percentages are based on 1,116 units (High 430, Mid 391, Low 295).

Table 7 summarizes final response outcomes among implementable feedback comments by course-achievement group. Implementation, including self-initiated revision, accounted for 71.8% of implementable comments in the high group, 74.3% in the mid group, and 65.9% in the low group. The response distribution did not differ significantly, $\chi^2(4) = 6.03$, $p = .197$. Broad and possible broad uptake stayed within a narrow band, between 7.4% and 10.7%, in every group. Because comments are nested within students, a cluster-robust binomial generalized estimating equation was also fitted to the 965 implementable-response units from 184 students. The course-achievement-group effect remained nonsignificant, Wald $\chi^2(2) = 2.35$, $p = .309$ (independence working correlation), with odds ratios of 1.16, 95% CI [0.76, 1.77], for the mid group and 0.79, 95% CI [0.50, 1.25], for the low group, both relative to the high group.

TABLE 7
Response Outcomes Among Implementable Feedback by Course-Achievement Group

Response category	High *n* (%)	Mid *n* (%)	Low *n* (%)	Interpretation
Implemented (including self-initiated revision)	275 (71.8%)	251 (74.3%)	166 (65.9%)	About 66–74% of implementable comments were acted on in every group
Broad or possible broad uptake	41 (10.7%)	25 (7.4%)	27 (10.7%)	Indirect, topic-related revision without direct correspondence
Not implemented	66 (17.2%)	59 (17.5%)	55 (21.8%)	No related change identified

Note. Percentages are based on implementable comments per group (High 383, Mid 338, Low 252); unresolved units (1, 3, and 4 per group) account for the remainder.

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

Differences in pooled comment-level implementation rates were −2.5 percentage points for the high–mid contrast, 95% CI [−10.8, 6.0]; 5.9 for high–low, 95% CI [−4.2, 16.0]; and 8.4 for mid–low, 95% CI [−2.7, 18.9]. The confidence intervals were obtained by resampling students within each group and recalculating the pooled rates, so that the clustering of comments within writers was preserved. Every interval spans values in both directions and is wide enough to remain consistent with differences of up to roughly 16 to 19 percentage points, so the absence of a significant effect should not be read as evidence that the groups performed equivalently.

Response Patterns by Comment Type

Response patterns did differ by the kind of comment. Among comments that addressed content or organization alone (n = 387), 58.4% were directly implemented, 13.4% were coded as broad or possible broad uptake, and 28.2% were not implemented. Among comments confined to grammar, vocabulary, or mechanics (n = 215), the corresponding figures were 86.5%, 5.1%, and 8.4%, $\chi^2(2) = 50.77$, $p < .001$, Cramér's V = .29 (n = 602); six comments with unresolved attribution were excluded from this comparison. A cluster-robust generalized estimating equation (independence working correlation) estimated the odds of broad or possible broad uptake as 2.88 times higher for content or organization comments than for grammar, vocabulary, or mechanics comments, 95% CI [1.47, 5.62], p = .002.

Feedback Accuracy and Selective Responses in the Exploratory Subset

Table 8 describes the selected 66-case accuracy subset. Of 55 accurate comments, 46 were implemented fully or partially and 9 were not, so roughly one accurate comment in six was passed over even in this most traceable subset. The subset also contained four inaccurate comments, two incorporated and two rejected, and six partially problematic comments, all incorporated at least in part.

TABLE 8

Feedback Accuracy and Writer Responses in the Exploratory Subset

Feedback accuracy	Implemented /partial	Not implemented	Other	Total	%	Interpretation
Accurate feedback	46	9	0	55	83.3%	Mostly incorporated, but some accurate feedback was not taken up
Inaccurate feedback	2	2	0	4	6.1%	Inaccurate feedback was sometimes incorporated, sometimes selectively rejected
Partially problematic feedback	6	0	0	6	9.1%	Direction partly valid, but quality limitations remained after incorporation
Excluded due to feedback–draft mismatch	0	0	1	1	1.5%	Verification needed as a peer-feedback unit

Note. The 66 cases were selected for clear traceability and are used only for exploratory illustration. Percentages are not prevalence estimates for the full corpus.

Illustrative Cases of Implementation and Non-Implementation

Table 9 presents selected cases, with the peer comment and the corresponding revision shown together. Mid-85, Low-55, and Low-78 illustrate the boundaries of indirect response. Mid-85 answered a request for stronger support by adding a new reason instead of developing the one that had been criticized; because the added reason addressed the request, the comment was coded as broad uptake. In Low-55, a request for more support for two reasons was followed only by one sentence restating one of them, a trace too slight to show that the problem had been addressed, so the comment was coded as possible broad uptake. In Low-78, the whole essay was rewritten, and no change could be attributed to this comment rather than to

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

the rewriting as a whole; the comment was coded as unresolved. Two further cases show that implementation does not guarantee improvement. Mid-10 adopted an inaccurate suggestion to replace 'have' with 'has' after 'Does', producing an ungrammatical question. In Mid-55, four of the five suggested corrections were made and the fifth only in part, yet an unflagged error of the same type remained. These examples clarify the coding decisions; they do not establish why individual students accepted or rejected a comment.

TABLE 9
Illustrative Cases of Feedback Implementation and Non-Implementation

Category	Case	Feedback and revision	Interpretation
Implemented with self-initiated revision	High-52	Peer comment: "Use 'Finally' instead of 'Third' for the last point." 'Third' became 'Finally', and the concluding sentence was moved to the end of the essay	Core suggestion incorporated; moving the conclusion, requested by no other comment, extends the suggestion
Implemented or partially implemented	High-20	Suggestion to insert 'me' in 'helps make' → 'helps me make social network'	Local expression feedback incorporated verbatim, the most common response in the final coding
Partially implemented	Mid-55	Peer comment suggested five corrections (a comma before 'because', 'have → has', 'valueable → valuable', sentence-initial capitals, and 'don't'). Four were made; the comma was added but the next word was capitalized ('big city, Because big city has…')	One of five suggestions incorporated only in part; coded as partially implemented. An unflagged error of the same type ('big city have chance') remained
Broad uptake	Mid-85	Peer comment that the first reason (lack of infrastructure) was thinly supported; instead of elaborating that reason, the student added a new transportation reason ('Second, traffic is well developed…')	Related strengthening of the argument without one-to-one correspondence to the comment; coded as broad uptake
Possible broad uptake	Low-55	Peer comment: "You gave two reasons why part-time jobs are negative; adding more support for them would help." The second draft added only 'I want them to focus more on school than part-time jobs,' restating the first reason	Related change present but too slight to show that the request for support was met; coded as possible broad uptake
Unresolved attribution	Low-78	Peer comment: "Opening with your own story together with your opinion would help readers." The second draft rewrote the entire essay, changing nearly every sentence	Wholesale rewriting; no change attributable to this comment rather than to the rewriting as a whole; not counted as non-implementation
Implementation of inaccurate feedback	Mid-10	Incorporated the inaccurate suggestion to replace 'have' with 'has' in 'Does social media have…', producing 'Does social media has…'	Problematic implementation showing that incorporation does not guarantee improvement
Rejection of inaccurate feedback	High-12	Did not adopt the suggestion 'in our daily lives → in daily life'	Original phrasing was natural; rejection interpretable as valid selection

Note. Cases were selected to illustrate the response categories and are not a representative qualitative sample.

DISCUSSION

The discussion follows the two research questions in order. Under RQ1, the three groups received comparable numbers and types of comments. The number of comments differed across the groups overall, but no two groups differed significantly after correction for multiple comparisons, and the distribution of feedback types did not differ. RQ1 thus provides descriptive

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

context for RQ2, but the analyses do not establish that feedback exposure or quality was equivalent across groups. The first section below discusses the feedback received, and the remaining sections turn to how it was taken up.

The Predominance of Content-Focused Comments

In all three achievement groups, content and logic comments were the most frequent. These comments addressed insufficient evidence, the need for examples, and the link between claims and conclusions. The pattern is notable because peer commentary among L2 learners is often expected to concentrate on surface correction; students new to peer response tend to address grammar, mechanics, spelling, and vocabulary rather than meaning (Leki, 1990). It becomes visible when the comments themselves are coded rather than inferred from learner perceptions, as in survey-based studies (e.g., J. Y. Min, 2020), and it is consistent with Paulus (1999), who found that feedback from readers prompted more meaning-level revision than students produced on their own.

The small omnibus difference in feedback exposure cannot be explained from the recipient data alone. Stronger drafts may have offered fewer obvious points for comment, or some texts may have been harder to evaluate, since giving feedback is itself cognitively demanding (Gao et al., 2023; F. Zhang et al., 2023). Provider–recipient links were not analyzed, so these explanations remain possibilities.

Separating Direct Implementation from Broad Uptake

Broad uptake was derived from two established distinctions, and its distribution in the data is consistent with them. For comments addressing content or organization, which often call for broader changes than comments on language form, the odds of a broad or possible broad uptake response were nearly three times as high as for comments confined to grammar, vocabulary, or mechanics. This pattern fits the way the category was defined, although it does not independently validate it, since content and organization comments were also the comments for which a one-to-one match was least likely to be found.

Engagement research has argued that visible incorporation captures only part of what learners do with feedback (Wu & Schunn, 2023; Z. Zhang & Hyland, 2023), which might lead one to expect indirect response to be common in a corpus where content and logic comments outnumbered all others. Indirect responses were nonetheless uncommon. Even among content and organization comments, direct implementation exceeded broad uptake by a factor of four, and across the corpus broad or possible broad uptake accounted for 8.3% of comments.

Two readings of that figure are available, and these data do not decide between them. Indirect response may be genuinely uncommon when essays are short, comments are specific, and only one revision cycle intervenes, or this corpus may simply have supplied few occasions for it. Either way, the share of indirect response could be measured only because the coding scheme provided a category for it; a binary coding of uptake could not have reported that it was 8.3%.

Implementation Extended by Self-Initiated Revision

In High-52 a comment prompted a change beyond its own scope: besides replacing ‘Third’ with ‘Finally’, the writer moved the concluding sentence to the end of the essay. Villamil and de Guerrero (1998) observed similar behavior when writers continued to revise after peer sessions, and it is consistent with the recontextualization that Z. Zhang and Hyland (2023) describe.

Only three units (0.3%) met the criterion, and the low count follows from how the category was defined. An additional change was credited only when it extended the suggestion itself, and each judgment was made against the recipient’s full set of comments, so that a change requested by a different peer was not counted as self-initiated. Coding comments one at a time, without that context, would inflate the category.

The resulting count is therefore conservative. It shows that additional revision beyond what a comment requested, traceable to that comment alone, was uncommon, not that students rarely revised beyond what they were told. Changes that addressed the issue raised by a comment without direct correspondence were coded as broad uptake, and changes without a defensible link to any comment remained unresolved.

The Selectivity of Uptake

He and Gao (2023) showed that a comment’s accuracy predicts whether it is implemented. The accuracy subset illustrates what can follow, and why uptake needs finer treatment in peer feedback research, although it cannot show how often the

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

patterns it contains occur. The subset consisted of the 66 comments whose responses were easiest to trace and included only four inaccurate comments, two incorporated and two declined. Within these limits, the cases show that incorporating feedback does not always improve the text, because acting on an inaccurate comment can introduce a new grammatical error or weaken meaning. Conversely, when a student declines inaccurate or unnecessary feedback, the second draft records a non-implementation, yet the decision may be a sound one.

Selective uptake connects here with learner feedback literacy. Carless and Boud (2018) identify making judgments, the capacity to evaluate the quality of one's own work and the work of others, as one of four features of feedback literacy, alongside appreciating feedback, managing affect, and taking action. Deciding whether a peer's suggestion would improve a draft calls for this kind of judgment. Accepting an accurate comment and declining an inaccurate one are consistent with sound judgment, and the reverse with judgment that went astray, but the drafts alone cannot establish whether students deliberately evaluated the comments. An accurate suggestion may simply have been copied, and an inaccurate one may have been left aside for other reasons. The present data show that both kinds of outcome occurred, but not how often. If the pattern holds in larger, randomly sampled sets of comments, feedback literacy instruction could include a procedure for checking whether a comment fits the context, whether it aligns with the writer's argument, and whether the resulting revision improves the sentence. How common such selection is, and whether it can be taught, are questions for research designed to estimate prevalence and to test instruction directly.

The response patterns can be interpreted cautiously through the self-regulation frameworks proposed by Oxford (2017), Panadero (2017), and Zimmerman (2000). Choosing, adapting, or declining a comment is compatible with strategic regulation, but the present study observed textual outcomes rather than learners' planning, monitoring, or reasons for action. Direct evidence of these processes would require revision logs, interviews, or stimulated recall.

Achievement Groups and Uptake

The study set out to examine whether selective uptake differed across course-achievement groups. In this course, no statistically significant association between course achievement and uptake was detected. Implementation rates were similar across the three groups, and neither the number of implemented comments nor the distribution of response categories distinguished higher- from lower-achieving students. The advantage that the literature reviewed above would predict for students who perform better in a course was therefore not detected in this sample. The confidence intervals around these contrasts are wide, however. They remain consistent with differences of up to roughly 16 to 19 percentage points, so the results show neither that the groups used feedback equivalently nor that learner characteristics matter less than comment characteristics in general. Response patterns did differ by comment type, with content and organization comments answered indirectly or left unanswered more often than comments on grammar, vocabulary, or mechanics.

Several explanations are possible, and the present data cannot decide among them. Final course grades combine attendance, examinations, assignments, and quizzes, so the grouping variable may be too distant from the abilities that uptake requires. The grade-band boundaries may also have blurred differences within groups. Comments that were short and specific may have been similarly actionable for all students. Finally, a single short revision cycle may leave little room for differences in uptake to emerge.

In courses like this one, course achievement, the grouping variable most readily available to an instructor, did not by itself identify students who used feedback differently. The quality of the comments students write, and students' evaluation of the feedback they receive, are possible targets for future intervention studies; the present design did not compare them with differentiating the activity by achievement level.

CONCLUSION

This study asked what becomes of individual peer comments in a required first-year general English course and whether the answer differs across course-achievement groups. For RQ1, the three groups received broadly similar types of comments, and no pairwise difference in the number of comments was significant after correction. Content and logic comments predominated in every group and outnumbered comments on grammar and wording. The omnibus difference in comment counts did not yield a significant pairwise contrast after correction for multiple comparisons. For RQ2, most comments were acted on. Direct implementation was the usual response in every group, second drafts were generally longer than first drafts, and content and organization comments were answered indirectly or left unanswered more often than comments on language form. The title's comparison of achievement groups may lead readers to expect a difference, but the analyses did not detect

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

a statistically significant association between course achievement and how feedback was taken up. Uptake was selective in that writers acted on some comments and not others, yet the study did not detect group differences in this selectivity; this does not establish that the groups were equivalent.

Broad uptake captured a smaller set of whole-text responses for which the link to a particular comment was less certain. Keeping the two categories apart preserves that difference in evidentiary strength, and it is what allows the share of indirect responses to be reported rather than assumed.

The findings suggest several practical steps, which the present study did not test and which should be read as proposals. Peer-feedback prompts can direct attention to content, evidence, examples, and organization as well as language form. Teachers can distinguish direct implementation from broader text-level response when reviewing second drafts. Deciding what to do with a comment requires judging its quality, a capacity central to learner feedback literacy (Carless & Boud, 2018), but the drafts alone could not show whether students made such judgments. Students can therefore be asked to check whether a suggestion is accurate and fits the local context, and to record briefly why they accepted or declined it. Such records would also provide evidence of evaluative judgment that was unavailable in this study, and declining a comment could then be treated as a possible act of judgment rather than automatically as neglect. Training for feedback providers can emphasize locating a problem, explaining it, and offering a usable suggestion. In completion-based writing activities, presenting peer feedback as low-stakes reader response may support substantive commentary without implying that every suggestion should be adopted.

The study has six limitations. First, the groups reflect final course achievement, a composite of attendance, examinations, assignments, and quizzes. It measures neither writing proficiency nor the ability to use feedback, and it is not equivalent to learners' own assessments of their proficiency. The writing task contributed no quality score to the grade, and the ungraded diagnostic quiz offers only limited evidence of prior differences. The group boundaries follow the grade bands of the university's relative grading policy, not a theoretically derived threshold, and collapsing grade bands into three groups may have obscured differences within them. Both features may have reduced the study's ability to detect group differences. Second, the sample supports only limited precision for the group contrasts, and the reported intervals leave room for moderate differences that this study could not detect. Third, broad uptake marks a possible comment–text connection and cannot establish causation. Fourth, inter-coder reliability was estimated on a 10% sample rather than through full double coding. Fifth, the accuracy subset consists of the comments whose responses were easiest to trace and contains only four inaccurate comments. It therefore cannot indicate how often selective uptake reflected the evaluative judgment that feedback literacy describes, whether by accepting accurate feedback or by declining inaccurate feedback, nor why students accepted or rejected particular comments. Sixth, excluding AI tools and dictionaries supported attribution within the classroom task, but it also means the setting was narrower than writing contexts in which students use machine translation and similar tools (H.-J. Im, 2017). Future work could link providers and recipients to examine who gives which types of comments, follow writers across several drafts, and use revision logs or interviews to investigate the reasoning behind uptake decisions. Research conducted with machine translation and generative tools available would also show how peer comments interact with other sources of writing support.

The main contribution of this study is a detailed, comment-level account of how peer feedback is reflected in revision in a general English course. It does not provide evidence about learners' underlying decision-making or self-regulatory processes. The study separates clearly traceable implementation from broader and less certain text-level response and finds no reliable differences in overall patterns across course-achievement groups. For teachers and researchers, the practical lesson is that a peer feedback activity is worth judging by which comments students acted on and how, and not by an implementation rate alone.

References

Ahn, Hyeseong, & Bang, Jyun. (2024). The impact of peer feedback on critical English writing skills. *Journal of the Korea English Education Society*, *23*(2), 53–74. https://doi.org/10.18649/jkees.2024.23.2.53

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, *32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*(4), 400–414. https://doi.org/10.58680/ccc198115887

Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing*, *50*,

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

Article 100775. https://doi.org/10.1016/j.jslw.2020.100775
Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*(2), 315–339. https://doi.org/10.2307/3588049
Gao, Y., An, Q., & Schunn, C. D. (2023). The bilateral benefits of providing and receiving peer feedback in academic writing across varying L2 proficiency. *Studies in Educational Evaluation*, *77*, Article 101252. https://doi.org/10.1016/j.stueduc.2023.101252
Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487
He, W., & Gao, Y. (2023). Explicating peer feedback quality and its impact on feedback implementation in EFL writing. *Frontiers in Psychology*, *14*, Article 1177094. https://doi.org/10.3389/fpsyg.2023.1177094
Hwang, Eunkyung. (2026). Peer feedback perception, uptake, and writing revision patterns across self-rated English proficiency levels in a college general English course. *Journal of the Korea English Education Society*, *25*(3), 127–146. https://doi.org/10.18649/jkees.2026.25.3.127
Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399
Im, Hee-Joo. (2017). The university students' perceptions or attitudes on the use of the English automatic translation in a general English class: Based on English writing lessons. *Korean Journal of General Education*, *11*(6), 727–751.
Im, Hee-Joo, & Hwang, Eunkyung. (2026). University students' engagement with peer feedback: The dissociation between revision completeness and reflection. *The Journal of Transdisciplinary Studies*, *10*(2), 139–152. https://doi.org/10.22685/jts.2026.10.2.139
Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, *44*(4), 294–304. https://doi.org/10.1093/elt/44.4.294
Kim, Hyo Young, & Kim, Jini. (2020). The effect of peer feedback on writing quality in L2 college classes. *The New Korean Journal of English Language and Literature*, *62*(1), 197–219. https://doi.org/10.25151/nkje.2020.62.1.010
Leijen, D. A. J., & Leontjeva, A. (2012). Linguistic and review features of peer feedback and their effect on implementation of changes in academic writing: A corpus based investigation. *Journal of Writing Research*, *4*(2), 177–202. https://doi.org/10.17239/jowr-2012.04.02.4
Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *The CATESOL Journal*, *3*(1), 5–19. https://doi.org/10.5070/B5.36653
Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290. https://doi.org/10.1080/13562510600680582
Min, H.-T. (2005). Training students to become successful peer reviewers. *System*, *33*(2), 293–308. https://doi.org/10.1016/j.system.2004.11.003
Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118–141. https://doi.org/10.1016/j.jslw.2006.01.003
Min, Ju Young. (2020). Korean university students' perceptions of feedback activities on English composition: Self-review, peer feedback, and computer feedback. *Korean Journal of English Language and Linguistics*, *20*, 335–362. https://doi.org/10.15738/kjell.20..202008.335
Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x
Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge. https://doi.org/10.4324/9781315719146
Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422
Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265–289. https://doi.org/10.1016/S1060-3743(99)80117-9
Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, *9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9
Villamil, O. S., & de Guerrero, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, *19*(4), 491–514. https://doi.org/10.1093/applin/19.4.491
Vuogan, A., & Li, S. (2023). Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *TESOL Quarterly*, *57*(4), 1115–1138. https://doi.org/10.1002/tesq.3178
Wei, Y., & Liu, D. (2024). Incorporating peer feedback in academic writing: A systematic review of benefits and challenges. *Frontiers in Psychology*, *15*, Article 1506725. https://doi.org/10.3389/fpsyg.2024.1506725
Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback? *Journal of Second Language Writing*, *21*(4), 364–374. https://doi.org/10.1016/j.jslw.2012.09.005
Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology*, *73*, Article 102160. https://doi.org/10.1016/j.cedpsych.2023.102160

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, *147*, Article 107857. https://doi.org/10.1016/j.chb.2023.107857

Yu, S., & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing*, *33*, 25–35. https://doi.org/10.1016/j.asw.2017.03.004

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461–493. https://doi.org/10.1017/S0261444816000161

Zhang, F., Schunn, C., Chen, S., Li, W., & Li, R. (2023). EFL student engagement with giving peer feedback in academic writing: A longitudinal study. *Journal of English for Academic Purposes*, *64*, Article 101255. https://doi.org/10.1016/j.jeap.2023.101255

Zhang, Z., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, *58*, Article 100784. https://doi.org/10.1016/j.asw.2023.100784

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

www.kci.go.kr

© 2026 The Modern English Education Society (MEESO)