Understanding and Supporting Peer Review Using AI-reframed Positive Summary

Chi-Lan Yang
chilan.yang@iii.u-tokyo.ac.jp
The University of Tokyo, Graduate School of Interdisciplinary Information Studies
Tokyo, Japan

Alarith Uhde
pcs@alarithuhde.com
The University of Tokyo, Tokyo College
Tokyo, Japan

Naomi Yamashita
naomiy@acm.org
1: NTT Communication Science Laboratories; 2: Kyoto University, Social Informatics
1: Keihanna; 2: Kyoto, Japan

Hideaki Kuzuoka
kuzuoka@cyber.t.u-tokyo.ac.jp
The University of Tokyo, Graduate School of Information Science and Technology
Tokyo, Japan

Abstract

While peer review enhances writing and research quality, harsh feedback can frustrate and demotivate authors. Hence, it is essential to explore how critiques should be delivered to motivate authors and enable them to keep iterating their work. In this study, we explored the impact of appending an automatically generated positive summary to the peer reviews of a writing task, alongside varying levels of overall evaluations (high vs. low), on authors' feedback reception, revision outcomes, and motivation to revise. Through a 2x2 online experiment with 137 participants, we found that adding an AI-reframed positive summary to otherwise harsh feedback increased authors' critique acceptance, whereas low overall evaluations of their work led to increased revision efforts. We discuss the implications of using AI in peer feedback, focusing on how AI-driven critiques can influence critique acceptance and support research communities in fostering productive and friendly peer feedback practices.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

Keywords

Peer feedback, critique acceptance, cognitive reframing, AI-mediated communication

ACM Reference Format:

Chi-Lan Yang, Alarith Uhde, Naomi Yamashita, and Hideaki Kuzuoka. 2025. Understanding and Supporting Peer Review Using AI-reframed Positive Summary. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3713219

CHI '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713219

1 Introduction

Peer reviews are crucial for quality and validity in scientific knowledge production, as they involve domain experts evaluating their peers' work [32]. However, this process is not without challenges, especially when it comes to delivering feedback. Unclear review guidelines, missed opportunities to follow review guidelines properly [54], the anonymity of the peer review [31, 64], heavy workload, lacking support for novice reviewers [55], and unclear review norms within the community [29] can sometimes lead to non-constructive or harsh criticism, which may be perceived as unfair or overly negative by the recipients [29, 41]. Such critical feedback can demotivate recipients (i.e., authors), potentially leading them to cease efforts to improve their work.

Despite the recognized importance of providing constructive feedback, the current practice in research communities places a significant cognitive burden on feedback providers (i.e., reviewers). They are expected not only to provide professional and high-quality reviews but also to offer critiques in a positive and constructive tone [1, 2, 4–6, 29, 32, 44]. This requirement demands additional effort, especially when the review workload is already heavy. This is where large language models (LLMs) can play a supportive role by assisting in reframing critical feedback in a more positive manner. Recent advancements in LLMs have made their application in the peer review process increasingly feasible [36, 55], particularly for tasks such as paraphrasing existing content[1]. When incorporating LLMs into the peer review process, it is crucial to effectively utilize LLMs while maintaining our originality and the autonomy that comes with being a human reviewer. Recent works on AI-assisted writing and AI-mediated communication indicated that people felt a diminished sense of control when they had minimal influence over the final outcomes of AI-assisted writing [16]. Additionally, they experienced a lack of authenticity when composing messages with AI for communication purposes [18]. Thus, instead of completely offloading the writing task to LLMs, we aim to investigate the possibility of using LLMs to "reframe" content based on what humans have written. We developed and evaluated the concept of an AI-reframed summary that reframes critical feedback more positively and appends it to the original critique written by the

[1] https://www.grammarly.com/paraphrasing-tool

reviewers to help authors engage more constructively with the review.

The main focus of this study was to see how AI-reframed positive summaries affected authors' perceptions and willingness to revise. Meanwhile, we cannot ignore the presence of "overall evaluation" scores, which are prevalent in peer reviews. In the current practice of peer review, authors receive critique with an "overall evaluation" of their work, which often includes scores or overall recommendations such as "accept with minor revision," "major revision," or "reject" [2, 3, 5]. The "overall evaluation" helps editors and reviewers make collective decisions, but related works have shown that receiving scores influenced authors motivation and task performance, even if scores were randomly assigned [73]. When we append an AI-reframed positive summary to critique that comes with an "overall evaluation", how these factors influence authors' further revision of the work is unclear. Hence, this study aims to examine how these scores interact with AI-reframed positive summaries.

Our study examined how AI-reframed positive summaries, combined with overall evaluations, influenced the way authors react to critiques. We define "overall evaluation" in this study as the perceived *relative performance* of the authors receiving feedback for their work, compared to other unknown peers. The evaluation scores are not necessarily linked to the actual quality of their work, as different reviewers may have varying standards. We conducted a two-by-two controlled experiment with 137 participants, varying the presence of *AI-reframed positive summaries* (with vs. without) and the *overall evaluation ratings* (high vs. low) to explore how these factors influenced participants' critique reception differently. Our findings show that critiques accompanied by an AI-reframed positive summary and a high overall evaluation led to increased positive emotions among authors. Moreover, these AI-reframed positive summaries increased authors' perceived autonomy and competence, and improved their perceived fairness of the feedback and the reviewers. Interestingly, while showing AI-reframed positive summaries did not significantly impact revision efforts, receiving a low overall evaluation did lead to significantly more revision efforts. Thematic analysis of the open-ended comments in the survey using the *attributional theory of motivation* revealed that both *intrapersonal* and *interpersonal* factors authors perceived from the critique influenced their motivation to revise their work.

This work makes three key contributions: First, it advances the understanding of how LLMs can be integrated into the peer review process by identifying both the opportunities and challenges involved; second, it provides empirical evidence on the effects of AI-generated positive summaries of reviewers' critical feedback and overall evaluation scores on feedback reception and authors' subsequent revision efforts; third, it offers broader insights into the implications of using LLMs in peer review and online feedback systems while suggesting directions for future research to further optimize and expand their role in enhancing these processes.

2 Background and Related Works

2.1 Supporting Peer Feedback

Peer review serves critical functions for both individuals and the community [25]. From an individual perspective, receiving peer feedback allows authors to gain new knowledge and refine and improve their work. From a community perspective, peer review acts as a gatekeeper, ensuring quality control and maintaining standards [32]. For example, code review helps programmers learn new implementation techniques while ensuring code quality for product deployment [25]. Similarly, peer review in research communities enables researchers to enhance science communication and promote the overall quality and trust within the research communities [32].

However, the effectiveness of peer feedback is often compromised by variability in the time and effort invested by the reviewers, especially since most of the peer review was done voluntarily. Additionally, the tone of the feedback, which can occasionally be excessively negative, may impact the emotional state and task performance of the authors [67]. In response to these challenges, several strategies have been developed to improve the process of delivering and receiving peer feedback for both reviewers and authors.

To support reviewers, Cambre et al. demonstrated that employing contrasting cases, which involve reviewers comparing two different submissions from different authors, improved the quality of peer feedback and engaged reviewers deeply [11]. Researchers also found that equipping reviewers with tools to quickly find examples to complement their feedback improved the quality of their critiques [30]. Additionally, it has been shown that rubrics [71] or structured scaffolding designed for novice reviewers facilitates the delivery of high-quality feedback during the peer review process [20, 23, 37, 55]. Moreover, encouraging reviewers to re-review their feedback based on the specific effective linguistic style, such as increasing specificity, before sending feedback [35] or to employ positive language, such as "good job" [41], has proven beneficial in improving the perceived helpfulness and emotional responses of authors and their reception to critiques.

On the other hand, to support authors, it has been found that encouraging and scaffolding authors to reflect on their own creative work and prioritize which comments should be addressed first during the revision process can mitigate the negative impact when receiving low-quality peer feedback [69, 70]. Additionally, some researchers have suggested employing various coping strategies, such as affirming one's self-worth before receiving critiques, having expressive writing, or diverting one's attention away from critique before revision, is effective in alleviating the negative effects when receiving critical feedback [67].

Although various strategies have been proposed to enhance the peer review process, they often require additional effort from both reviewers and authors. Given that the creation and refinement of review content are essential to the peer review process and critical to the advancement of scientific knowledge, many argue that this responsibility should not be delegated to technology such as AI. Therefore, in this paper, we focus on how AI can assist in improving the tone of reviews. Specifically, we explore the effects of adding a positively reframed summary generated by AI, using Large Language Models (LLMs), based on the original critical feedback provided by reviewers. Notably, the feedback itself remained unchanged, presented in its original form, with only the addition of a positive reframed summary at the end of the feedback. We then examined how this adjustment was perceived by the authors.

2.2 Cognitive Reframing for Positive Thinking

Receiving critical feedback has been found to elicit negative emotions in the peer review process [68]. These negative emotions can cause individuals to focus on negative aspects and feel discouraged, potentially hindering the iterative improvement of their work [56, 68]. It has been shown that providing praise in feedback is crucial for reviewers to reinforce positive behavior and enhance the self-esteem of authors [13], and it is often used to soften criticisms, thereby improving the relationship between reviewers and receivers [26].

Similar to the function of praise, we see cognitive reframing, also known as cognitive reappraisal, has the potential to soften criticisms. Cognitive reframing has been proven effective in various contexts. It can help in coping with anxiety during cognitive therapy [14], reducing emotional distress when receiving supportive communication [8], and offering reframed social support to people with negative thinking patterns [53]. Cognitive reframing involves identifying and changing one's perspective on situations and thoughts, with the goal of altering their emotional impact [58]. However, effectively employing cognitive reframing is cognitively demanding, especially when under stress [45], and it usually requires adequate training or professional assistance [58].

Therefore, we see the opportunity of using large language models (LLMs) to reframe critical feedback positively to support feedback reception. With the development of LLMs, researchers have begun to apply LLMs to support cognitive reframing. For instance, a recent study demonstrated the potential of LLMs to guide individuals in reevaluating their situations and finding alternative ways to manage upsetting circumstances [72]. A case study also revealed that LLMs can effectively help individuals overcome negative thinking and reduce emotional intensity [50].

However, the current research primarily emphasizes the potential of LLMs in assisting cognitive reframing to manage mental health issues, with limited exploration of its impact on receiving critiques in peer feedback. As previously mentioned, receiving critical feedback can potentially lead authors to experience negative emotions [33], low self-efficacy [39], and low autonomy [62]. Hence, this study seeks to broaden the scope of LLMs' reframing capabilities to the peer review process. Particularly, we aim to investigate how LLMs can positively reframe critiques to potentially influence the way feedback is received.

2.3 The Practice of Showing an Overall Evaluation in Peer Review

Research communities often adopt the practice of giving an overall evaluation in peer reviews to support reviewers in streamlining decision-making with other reviewers and guide authors to the expected next steps. To reviewers, the quantified score or categorization of recommendations provides them a rubric to assess the work and also facilitates reviewers and editors in making a consensus. To authors, the quantified relative performance could influence their task performance and motivation. For instance, previous studies have shown that individuals who received randomly assigned high performance as positive feedback demonstrated better task performance compared to those who were randomly evaluated as low performers [73]. Receiving scores also influences task motivation because individuals gauge their relative performance via scores by comparing themselves with others [10]. Though presenting overall evaluation is a common practice in peer review, we lack an understanding of how presenting this relative performance influences authors' perceptions and their revision outcomes, especially when combined with an AI-reframed positive summary. As far as we know, the rationale for keeping this practice of giving an overall evaluation is often grounded in established editorial policies and general best practices [2, 4–6] rather than empirical studies. Hence, we aim to explore how *the presence of AI-reframed positive summary* combined with *different overall evaluations* affects authors' critique reception.

2.4 Attributional Theory of Motivation and Authors' Motivation for Making Revisions

According to the *attributional theory of motivation* [19, 65], the way authors attribute the cause of the feedback can also affect their attitude and subsequent revision behavior. The attributional theory of motivation [65], first proposed by Bernard Weiner in the context of educational psychology, explains that people's motivation for a given task can be influenced by how they interpret the cause of success or failure. This includes their own responsibility (*intra*personal attribution) or others' responsibility (*inter*personal attribution). These two forms of attributions are not mutually exclusive, and the boundary between them can be unclear. However, this differentiation still enables researchers to better understand people's motivations. When individuals primarily adopt intrapersonal attribution, they tend to see the cause of their successes or failures as their own responsibility. This can lead to self-improvement or, if negative feedback is constantly received, to low self-efficacy [65]. Conversely, individuals who mainly adopt interpersonal attribution tend to attribute their learning outcomes to external factors, such as reviewers, feedback providers, instructors, teachers, or institutions. This can result in efforts to change external circumstances, such as the community, but may also lead to complaints about perceived unfairness within the community or society [65].

Based on the intrapersonal and interpersonal factors in the attributional theory of motivation, it is expected that when authors receive critical feedback, they may work on improving themselves to enhance self-efficacy, or they may end up believing they are unable to succeed if they adopt intrapersonal attribution. Alternatively, authors may attempt to change external factors, such as the practices of their research community, or resort to complaining and possibly leaving the community if they mainly adopt *inter*personal attribution.

Knowing how intrapersonal and interpersonal factors influence authors' motivation to make revisions enables researchers and designers to better provide support for peer feedback. Therefore, we also explored how authors describe their motivation to make revisions based on the critical feedback they received.

3 Research Questions and Hypotheses

In summary, the research questions were developed based on the *attributional theory of motivation* by considering the effect of the *presence of AI-reframed positive summary* and *overall evaluation*

on *intrapersonal* (RQ1), *interpersonal* (RQ2) factors, and revision outcomes (RQ3). We also examined potential motivational factors for authors to make revisions (RQ4). Because there was not enough past work to inform hypotheses for the factor of *overall evaluations*, we explored the impact of *overall evaluations* on critique acceptance more broadly using only research questions. Thus, we formed hypotheses for the factor of "presence of AI-reframed positive summary" but not for "overall evaluation". Together, we asked the following four research questions:

RQ1: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) affect the authors' intrapersonal perception when receiving critical feedback?

Under RQ1, we formed three hypothesis for the factor of *presence of AI-reframed positive summary* based on Section 2.1 and Section 2.2.

- H1a: Receiving an AI-reframed positive summary leads to more *positive emotion* compared with not receiving an AI-reframed positive summary.
- H1b: Receiving an AI-reframed positive summary makes people have less decrement in their *self-efficacy in writing* compared with not receiving an AI-reframed positive summary.
- H1c: People experience higher *perceived autonomy and competence* when the critique comes with an AI-reframed positive summary compared with not receiving an AI-reframed positive summary.

RQ2: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) affect the authors' interpersonal perception when receiving critical feedback?

Under RQ2, we formed two hypothesis based on Section 2.1. We formed the following hypothesis for the factor of *presence of AI-reframed positive summary* based on Section 2.1 and Section 2.2.

- H2a: Receiving an AI-reframed positive summary increases the perceived fairness and usefulness of the critique.
- H2b: Receiving an AI-reframed positive summary makes people perceive the reviewer was fairer and more knowledgeable compared with not receiving it.

RQ3: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) influence the authors' revision outcomes when receiving critical feedback?

RQ4: What factors influence the authors' motivation to revise their writing when receiving critical feedback?

4 Method

4.1 Experiment Design

To investigate the influence of showing *AI-reframed positive summary* and *overall evaluation* of the writing on authors' feedback reception, motivation, and revision outcome, we conducted a two-stage online controlled experiment, in which we asked participants to complete a writing task in stage 1 and then revise their writing based on the feedback they received in stage 2. In the writing task at stage 1, we instructed participants to write a 250-word cover letter for a job in social media marketing. We asked them to introduce themselves and describe how their characteristics, skills, and past experience make them suitable for this position. Then, we invited the same group of participants to join the study again three days later after completing their initial writing. We instructed participants to revise their initial writing based on the feedback we gave them at stage 2. We chose this writing task because we had to find a writing topic that was general enough for participants to complete in a relatively short period of time. Also, the task should have no right or wrong answers that allow us to give participants a randomized overall evaluation.

We manipulated and compared two types of feedback presentation (with and without AI-reframed positive summary) and two levels of overall evaluation (high-scored and low-scored overall ratings). This two-by-two design led to four conditions, including *NoPosFramed-HighScored, PosFramed-HighScored, NoPosFramed-LowScored,* and *PosFramed-LowScored* conditions. We randomly assigned participants to experience one of four conditions (between-subject design).

In *NoPosFramed-HighScored* condition, we provided critical feedback based on participants' writing at stage 1. To maintain the consistency of the feedback we gave to participants in terms of the aspects of criticism, we used the GPT4 model[2] to generate critical feedback, specifically focusing on the authors' *writing style, professional standing, work experience, and key accomplishments and their relevance* in applying for a specific job position in the writing task. As the response of the GPT models was designed to be expressed as positively as possible [42], we explicitly mentioned *being a strict reviewer* and *using a harsh tone* in the prompt for generating critical feedback. The detailed prompt can be referred to from the supplementary document. Please note that we did not inform participants about the source of their feedback. However, we asked them to identify who they thought might have provided the feedback, including ChatGPT, another crowd worker, a social media expert, or a member of our research team. 26% of participants (n=26) reported that they believed ChatGPT was the author of the feedback. In addition to providing customized feedback to participants, we also showed them an overall evaluation prior to the feedback, saying, "The score of your cover letter is 4 out of 5 among all candidates for this position." The purpose was to make people perceive that they received a high evaluation of their writing.

In *PosFramed-HighScored* condition, participants saw a positively reframed summary at the end of the critical feedback they received. We also used the same GPT4 model to generate a positively reframed summary based on the customized feedback we prepared for each participant. We followed the practice of encouraging cognitive reframing in cognitive therapy [27] to design the prompt by making the GPT4 model interpret the reviewers' intention by thinking about a benefit or upside to a negative situation using a positive tone. Also, the reframed content can involve identifying a lesson to be learned from a difficult situation. The example of critical feedback and AI-reframed positive summary is in Table 1. The purpose of showing a positively reframed summary was to encourage authors to interpret the critical feedback in a positive and constructive way regardless of the content and tone of the feedback they received.

[2]We used the gpt-4-0125-preview model for generating critical feedback.

Table 1: Example of critical feedback and AI-reframed positive summary a participant received. Participants saw the AI-reframed summary being appended to, not in place of, the critical feedback in the study.

Critical feedback	AI-reframed positive summary
1. Writing Style: The cover letter is excessively informal and lacks professional finesse. Starting with "To whom it may concern" is outdated, and the letter meanders without getting to the point. A sharp, engaging introduction tailored to the job and company is missing. 2. Professional Standing: Mentioning unrelated degrees (Marine Biology and Neurobiology) without effectively tying these to the required social media content creation skills shows a fundamental misunderstanding of the job's nature. The attempt to link these degrees with organizational skills is weak and unpersuasive. 3. Work Experience: Your experience in sales and as an environmental educator, while interesting, is only tangentially relevant. The crucial omission is your lack of direct experience in social media content creation or using analytics tools professionally. Asserting that you believe you can easily learn does not substitute for proven experience. 4. Key Accomplishments and Relevance: There's no indication of proficiency in graphic design tools, strong visual aesthetics, or specific accomplishments in content creation that would be directly relevant to this role. Simply stating personal use of social media and a vague familiarity with technology trends falls significantly short of demonstrating the expertise or passion for social media content creation this position demands.	Dear Applicant, Your cover letter is a promising foundation for showcasing your talents. A more dynamic opening could better reflect your creativity and personalizing it to the hiring manager shows you're truly invested. Highlighting specific achievements and providing concrete examples of your success will not only substantiate your claims but also bring your expertise to life. Don't forget to detail your experience with social media platforms, emphasizing how your skills meet the job's visual and design demands. This feedback is a splendid opportunity to refine your application into one that not only meets but exceeds expectations. Best regards,

In addition to providing customized feedback with a positively reframed summary to participants, we also showed them the same overall evaluation as we did in *NoPosFramed-HighScored* condition. The detailed prompt we used for generating the positively reframed summary can be referred to from the supplementary document.

Table 2: Experiment design and sample size for each condition.

	critical feedback only (baseline, 150~200 words)	critical feedback (150~200 words) + AI-reframed positive summary (50 words)
See high overall evaluation	Score: 4/5 (N = 31)	Score: 4/5 + AI-reframed positive summary (N = 37)
See low overall evaluation	Score: 2/5 (N = 32)	Score: 2/5 + AI-reframed positive summary summary (N = 37)

In *NoPosFramed-LowScored* condition and *PosFramed-LowScored* condition, we followed the same approach to generate critical feedback and positively reframed summary for every participant, depending on their condition. The only difference was the overall evaluation we showed participants prior to the feedback, which said, "The score of your cover letter is 2 out of 5 among all candidates for this position." The purpose was to make people perceive that they received a low evaluation of their writing.

4.2 Participants

We recruited 137 participants (70 females, 63 males, 4 unspecified) with a mean age of 29.75 years old (SD = 8.81) from Prolific[3]. All participants were paid $10.5 GBP for participating in the approximately 65-minute study. The payment was determined based on Prolific's standard. We recruited participants by telling them this study required them to join two times by writing a cover letter for applying for a pseudo-job position (stage 1) and revising their writing a few days later (stage 2). The ethical review board of the authors' institute approved the study.

4.3 Procedure

The study was carried out on a website built with Google Cloud Firestore. In stage 1, participants joined the study directly via Prolific and accessed the Google form to read the pseudo-job description, which was a social media content creator for a tech company (Please refer to the details of the task instructions from the supplementary document). We instructed participants to fill out the Google form with their demographic information and basic information related to the job description, including their work experience and skills. The purpose was to engage participants by simulating the process of filling in an application form in an actual job application process. Next, we redirected participants to the website, where they were instructed to write a 250-word cover letter based on their work experience and skills, which they had just filled out on the Google form. After finishing the writing task, participants ended stage 1 by completing a survey via Google form. This process took approximately 30 to 40 minutes, depending on the participants' writing speed.

Approximately three to seven days later, we invited the same group of participants to join the study again. Participants would see the feedback with an overall evaluation depending on the condition to which they were randomly assigned. Afterward, they were presented with their initial writing and then instructed to revise it

[3]We recruited participants from https:/www.prolific.com/, an online participant recruiting and data collection platform

Figure 1: Procedure of the controlled online experiment.

based on the feedback they received. In this stage, we asked participants to revise their cover letters to at least 280 words (from 250). This word limit was intended to encourage participants to make some revisions at least. Once participants finished revising their cover letters, they filled in another Google form to end the study. At the end of the survey, we debriefed participants that the critical feedback and reframed summary were both generated by AI. This process took approximately 20 to 25 minutes, depending on the participants' revising speed (Figure 1).

For context, we did not forbid participants from using generative AI in the study to replicate real-life scenarios where people might use assistive technology for writing tasks. However, we disabled the copy-and-paste function in our web app to prevent direct copy-pasting. In the end, 29.2% of the participants (n = 40) reported using generative AI during the writing process.

4.4 Measurements

4.4.1 Perceived Emotional Valence of the Feedback (H1a). To examine how participants feel after receiving critical feedback, we asked them to indicate their feelings based on the following question adapted from Rasmussen and Berntsen [46]: "The feedback I received, as I recall are: 1-extremely negative, and 7-extremely positive"

4.4.2 Change of Self-Efficacy in Writing (H1b). We also evaluated participants' self-efficacy before and after receiving critical feedback with a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). The nine items for self-efficacy were adapted from Mitchell et al. [40]. The original scale was developed for evaluating people's self-efficacy for academic writing. We changed a few keywords and replaced them with writing a cover letter and self-introduction for applying for a job to fit the study context. Example items include *"I feel I have the skills to write a cover letter (self-introduction) for this job application"*, *"Writing a cover letter (self-introduction) for this job application comes easily to me."* The Cronbach's Alpha for the survey was α = .918. We calculated the average score from nine items from stage 1 and stage 2. The difference between the two scores from each stage was calculated to represent participants' change in self-efficacy.

4.4.3 Perceived Autonomy and Competence (H1c). Regarding how feedback influenced perceived autonomy and competence, we adapted the questions from Sheldon et al. [51]. Participants answered the following three items for autonomy with a 7-point Likert scale, ranging from 1 (not at all) to 7 (very much): *When reading the feedback and revising the text, I felt...that my choices were based on my true interests and values; free to do things my own way; that my choices expressed my "true self"*;, and another three items for competence: *When reading the feedback and revising the text, I felt...that I was successfully completing difficult tasks and projects; that I was taking on and mastering hard challenges; very capable in what I did.* The Cronbach's Alpha for the three autonomy items was α = .756, and for competence, it was α = .804. We averaged the scores from three items for each construct, representing participants' perceived autonomy and competence, respectively.

4.4.4 Perception of the Peer Feedback (H2a). To examine how much participants accepted the feedback they received, we adapted the survey from Nguyen et al. [41] and asked them to rate the perceived usefulness and fairness of the feedback. The questions related to perceived usefulness include, *The feedback I received influenced my revision of the work; I applied the feedback I received when I edited my work.* Whereas the questions related to perceived fairness include, *The feedback I received provides helpful suggestions for improving my work; The feedback I received is constructive; The feedback I received is fair; The feedback I received is relevant to my work.* The Cronbach's Alpha for the three items in *perceived usefulness* was α = .827, while the Cronbach's Alpha for the three items in *fairness of the feedback* was α = .911. Participants answered all survey items with a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree). We averaged the scores from all items from the sub-scale and used each of them to represent the *perceived usefulness* and *fairness of the feedback*, respectively.

4.4.5 Perception of the reviewer (H2b). We also adapted the survey from Nguyen et al. [41] to assess how participants perceived the reviewers, including their fairness and expertise. The questions regarding the fairness of reviewers were: *The person who gave me feedback was considerate; The person who gave me feedback was sincere; The person who gave me feedback was polite; The person who gave me feedback was disrespectful* [4]. The questions regarding the expertise of reviewers were: *The person who gave me feedback was qualified; The person who gave me feedback was an expert in the task; The person who gave me feedback was knowledgeable.* The Cronbach's Alpha for the three items in *perceived usefulness of the reviewer* was α = .87, while the Cronbach's Alpha for the three items in *expertise of the reviewers* was α = .936. These items were also measured with a 7-point Likert scale, where one indicated strongly disagree, and seven indicated strongly agree. We averaged the scores from all items from the sub-scale and used each of them to represent the *perceived fairness* and *expertise of the reviewers*, respectively.

[4]This is a reversed item

Table 3: Rubric for evaluating revision quality of the cover letter. Two independent raters, who were unaware of the research purpose, evaluated all 137 revisions based on this rubric. The maximum score for each revision is 20, while the minimum score is 4.

	5-Effective	3-Average	1-Need improvement
Format: grammar	No spelling, punctuation or grammar errors	Some spelling or grammatical errors found	Many errors that take focus away from content
Format: basic information	Covers basic information but offers an engaging and gripping way into the body of the letter and clearly connects the person to the position	Covers basic information but only a lackluster way of getting into the core content of the letter	May or may not cover basic information and only a tenuous or weak way into the body of the letter and establishes no link between the person and the position
Content: skills	Emphasizes skills or abilities the person has that relate to the job for which they are applying. Mentions work, volunteer, and education experiences.	Limited information on skills or abilities the person has that relate to the job they are applying for. Some unsupported assertions but also some good examples of the connections between the person and the position	List of Skills (i.e.- Communication, Flexibility, and Teamwork) with no evidence of work, volunteer, or education experience
Content: qualifications	Identifies one or two of the person's strongest qualifications and clearly relates how these apply to the job at hand.	Identifies one of the person's qualifications, but it is not related to the position at hand.	Does not discuss any relevant qualifications. Have not related the person's skills to the position applied for.

4.4.6 Revision Outcome (RQ3). In addition to their perception, we also measured their actual revision outcomes, including quantity and quality of revision.

Regarding the quantity of revision, we compared the document similarity between participants' initial writing and their final revision using cosine similarity, which is one of the common approaches for calculating semantic similarity between two documents [15, 21, 47]. We first converted text into numerical vectors using TF-IDF (Term Frequency-Inverse Document Frequency) and then calculated the cosine similarity between the two vectors. We used TF-IDF because this captured word-level information, and the weights assigned to words are relatively easy to interpret [5]. The cosine value ranges from 0 to 1. If the angle between two vector representations is small (close to 0 degrees), the cosine value is close to 1, meaning the two documents are very similar. If the angle is large (close to 90 degrees), the cosine value is close to 0, meaning the documents are not similar. In other words, a high cosine value indicates participants made few semantic revisions.

Regarding the quality of revision, we developed a rubric by synthesizing the rubrics we found from the career center that taught students how to write a cover letter [59–61]. The rubric consisted of four main aspects: grammar, basic information about the applicant, skills mentioned by the applicant, and qualifications (Table 3). A 5-point Likert scale was used to score each item, ranging from "needs improvement (1)" to "effective (5)." This allowed for a maximum score of 20 for each revision. We then recruited two external raters who were unaware of the research purpose to evaluate all 137 revisions based on the rubric. We paid each rater 78.15 USD (this is slightly above the hourly wage of the country where the study was conducted) for approximately five to six hours of work to assess all revisions. The averaged scores from two raters were analyzed as an index for quality of revision.

In addition to the behavioral index, we also collected participants' perceptions about their revision quality. We asked participants the following three questions with a 7-point Likert scale, ranging from 1 (low) to 7 (high): *How much effort did you invest in the writing task? How would you rate the quality of the writing task? How confident are you that the writing task fully satisfied the goals?* Three items were adapted from [69], and the Cronbach's Alpha for the three items was α = .858. The average score served as an index for participants' subjective assessment of their revision quality.

4.4.7 Potential Factors that Influence One's Motivation to Make Revision (RQ4). To investigate what factors determine participants' decision to make revisions, we asked participants to elaborate on their reasons for revising or not revising their cover letters in the open-ended responses of the survey. The data from participants' open-ended comments enabled us to identify participants' "claimed motivation". We followed the thematic analysis method [9] to analyze the open-ended responses. One of the authors open-coded all relevant concepts, assigned labels that featured the concepts, and grouped labels into different themes. Next, all the authors repeatedly discussed the quotes within each theme and the themes themselves. Finally, the developed themes were compared and adapted among all participants until they covered the data exhaustively.

5 Results

In the following analysis, we first tested the homogeneity of variances across groups using Levene's test for each of our dependent variables. We also tested the normality of the distribution of each dependent variable using the Shapiro-Wilk test. If both assumptions were met (i.e., the data showed homogeneity of variances according to Levene's test and normality according to the Shapiro-Wilk test), we proceeded with the standard two-way ANOVA with two independent variables, *with/without AI-reframed summary* and *high/low overall evaluation*, to analyze the data. If one or both assumptions were violated (i.e., if Levene's test indicated unequal variances or the Shapiro-Wilk test suggested non-normality), we applied the Aligned Rank Transform (ART) [66] to the dependent variables first. After applying ART, we then conducted the two-way ANOVA on the transformed data.

[5]We also tried BERT embeddings, which captured rich semantic relationships between words and contexts, with cosine similarity and got a similar statistical result. The conclusion of the revision quantity remained the same. We put the result in the supplementary file.

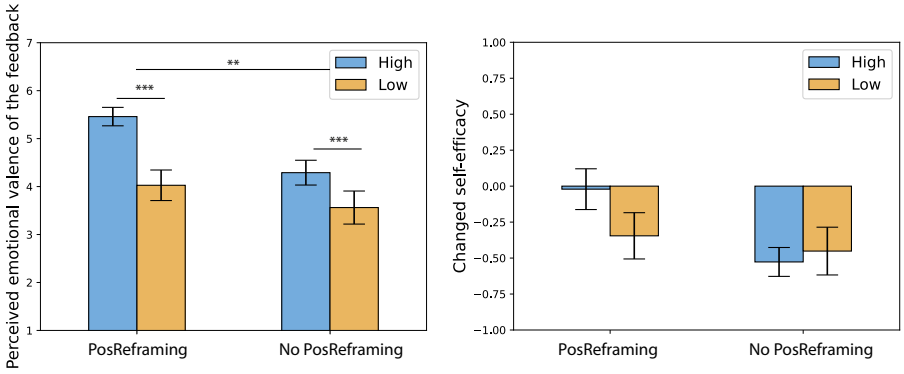

Figure 2: Mean (± standard error of the mean) of the perceived emotional valence of the feedback (Left; H1a) and the change in self-efficacy between initial writing and revision (Right; H1b). The X-axis shows whether participants received *AI-reframed positive summary*. The Y-axis of the left figure shows the perceived emotional valence. A high score indicates participants perceived positive emotion. The Y-axis of the right figure shows the change in self-efficacy. A score closer to zero indicates a smaller decrease in self-efficacy. (indicates $p < .01$, *** indicates $p < .001$)**

5.1 Effect on Intrapersonal Perception (RQ1)

To investigate the effect of *AI-reframed positive summary* and *overall feedback evaluation* on receivers' perception of themselves (RQ1), we compared their emotion, self-efficacy for writing, and perceived autonomy and competence across four conditions.

5.1.1 Positively Reframed Summary and High Scored Feedback Induced Positive Emotion (H1a). The result of a two-way ANOVA using aligned ranked transformation (ART) revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on participants' emotional valence ($F[1, 133] = 0.97$, *n.s.*). However, we found a significant main effect for both *AI-reframed positive summary* and *overall feedback evaluation* on participants' emotional valence. Regardless of receiving a high or low overall evaluation, participants had significantly more positive emotions when receiving AI-reframed positive summary than the original critical feedback (**H1a was supported**, $F[1, 133] = 7.71$, $p < .01$, $\eta_p^2 = .054$; PosFramed: $M = 4.74$, $SD = 1.55$; NoPosFramed: $M = 3.92$, $SD = 1.89$). Regardless of receiving AI-reframed positive summary or not, participants had significantly more positive emotions when receiving a high overall evaluation than a low overall evaluation of the feedback ($F[1, 133] = 14.95$, $p < .001$, $\eta_p^2 = .10$; HighScored: $M = 4.93$, $SD = 1.58$; LowScored: $M = 3.81$, $SD = 1.76$). (Figure 2, left)

Receiving *AI-reframed positive summary* and *a high overall evaluation* of the feedback led people to experience a more positive emotion than without receiving *AI-reframed positive summary* and *a low overall evaluation* of the feedback.

5.1.2 Reduced Self-Efficacy in Writing After Receiving Critical Feedback (H1b). The result of a two-way ANOVA with ART revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on participants' change of self-efficacy in writing ($F[1, 133] = 2.90$, *n.s.*). We also did not observe a significant main effect of either *AI-reframed positive summary* (**H1b was not supported**) or *overall feedback evaluation* on participants' change of self-efficacy in writing, though we observed the trend that all participants reduced self-efficacy in writing after receiving critiques. (Figure 2, right)

Regardless of the presence of *AI-reframed positive summary* and the score of *overall feedback evaluation*, participants experienced a decrease in self-efficacy after receiving critical feedback.

5.1.3 Positively Reframed Summary Induced High Autonomy and Competence (H1c). The result of a two-way ANOVA revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on participants' perceived autonomy ($F[1, 133] = 0.003$, *n.s.*) and competence ($F[1, 133] = 0.21$, *n.s.*) in making revisions. However, there was a significant main effect for *AI-reframed positive summary* on participants' perceived autonomy and competence. Regardless of receiving a high or low overall evaluation, participants experienced higher autonomy and competence for revision when the critical feedback came with an AI-reframed positive summary, compared with the original critical feedback (**H1c was supported**. Autonomy: $F[1, 133] = 6.95$, $p < .01$, $\eta^2 = .05$; PosFramed: $M = 4.98$, $SD = 1.22$; NoPosFramed: $M = 4.41$, $SD = 1.29$; Competence: $F[1, 133] = 9.12$, $p < .01$, $\eta^2 = .06$; PosFramed: $M = 5.02$, $SD = 1.13$; NoPosFramed: $M = 4.39$, $SD = 1.3$). (Figure 3)

Receiving critical feedback came with an *AI-reframed positive summary* significantly increased people's perceived autonomy and competence in making revisions, regardless of the score of the overall evaluation.

5.2 Effect on Interpersonal Perception (RQ2)

Next, we moved on to investigate the effect of *AI-reframed positive summary* and *overall feedback evaluation* on receivers' perception of the feedback they received (RQ2). We compared their perception of the feedback itself and the reviewers across four conditions.

5.2.1 Positively Reframed Summary Increased Perceived Fairness of the Peer Feedback (H2a). We asked participants to rate the usefulness and fairness of the feedback they received. The result of a

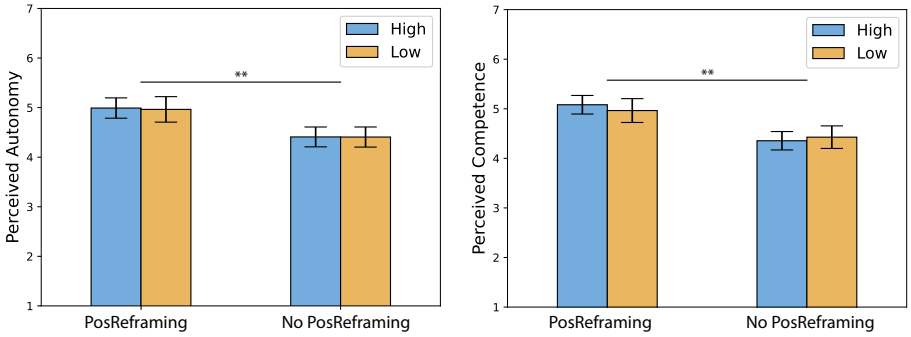

Figure 3: Two bar charts with a mean (± standard error of the mean) of participants' perceived autonomy (left) and competence (right) in making revisions. The X-axis shows the factor of the presence of *AI-reframed positive summary*. The Y-axis shows participants' perceived autonomy and competence. A higher score indicates a higher perceived autonomy and competence (indicates $p < .01$).**

two-way ANOVA with ART revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on neither the usefulness ($F[1, 133] = 0.07$, *n.s.*) nor fairness ($F[1, 133] = 0.005$, *n.s.*) of the feedback. However, we found the main effect for *AI-reframed positive summary* on participants' perceived fairness of the feedback. Regardless of receiving a high or low overall evaluation, participants thought the critical feedback was fairer when it came with an AI-reframed positive summary, compared with the original critical feedback (**H2a was partially supported**. $F[1, 133] = 8.62$, $p < .01$, $\eta_p^2 = .06$; PosFramed: $M = 6.17$, $SD = 0.85$; NoPosFramed: $M = 5.45$, $SD = 1.47$). The above effect was not found for participants' perceived usefulness of the feedback . (Figure 4A and B)

Receiving *AI-reframed positive summary* increased the perceived fairness of the critical feedback, but did not impact the usefulness of the critical feedback.

5.2.2 Positively Reframed Summary Increased Perceived Fairness of the reviewers (H2b). Next, we asked participants to rate the expertise and fairness of the reviewers. The result of a two-way ANOVA with ART revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on neither the expertise ($F[1, 133] = 1.40$, *n.s.*) nor fairness ($F[1, 133] = 0.20$, *n.s.*) of the reviewers. However, we found the main effect for *AI-reframed positive summary* on participants' perceived fairness of the reviewers. Regardless of receiving a high or low overall evaluation, participants thought the reviewers were fairer when the critical feedback came with an AI-reframed positive summary, compared with the original critical feedback (**H2b was partially supported**. $F[1, 133] = 4.962$, $p < .05$, $\eta_p^2 = .04$; PosFramed: $M = 5.8$, $SD = 1.05$; NoPosFramed: $M = 5.15$, $SD = 1.56$). A similar effect was not found for participants' perceived expertise with the reviewers. (Figure 4 C and D)

Receiving *AI-reframed positive summary* increased participants' perceived fairness of the reviewer, but did not impact the perceived expertise of the reviewer.

5.3 Effect on Revision Outcome (RQ3)

5.3.1 Low Scored Feedback Increased Quantity of Revision. We calculated the cosine similarity using TF-IDF between participants' initial writing and their final revision outcomes. This index indicates the quantity of participants' revisions. The result of a two-way ANOVA with ART revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on participants' quantity of revision ($F[1, 133] = 0.002$, *n.s.*). However, we found the main effect for *receiving overall evaluation* on participants' quantity of revision. Regardless of receiving an AI-reframed positive summary or not, participants made more revisions when the critical feedback was presented with a low-scored overall evaluation, compared with a high-scored overall evaluation ($F[1, 133] = 5.36$, $p < .05$, $\eta_p^2 = .04$; LowScored: $M = 0.81$, $SD = 0.13$; HighScored: $M = 0.87$, $SD = 0.11$). (Figure 5, Left)

5.3.2 Quality of Revision. The quality of the revision was the average score of two independent raters who evaluated the revisions based on a rubric we developed. The score for the quality of revision ranged from 4 to 20. The higher the score, the higher the quality of the revision. The result of a two-way ANOVA revealed that there was no significant interaction effect of *AI-reframed positive summary* and *overall feedback evaluation* on participants' quality of revision ($F[1, 133] = 0.043$, *n.s.*). We also did not find any main effect of *AI-reframed positive summary* ($F[1, 133] = 0.12$, *n.s.*) and *overall feedback evaluation* ($F[1, 133] = 0.65$, *n.s.*) on participants' quality of revision. (PosFramed-HighScored: $M = 13.15$, $SD = 3.15$; PosFramed-LowScored: $M = 13.45$, $SD = 2.54$; NoPosFramed-HighScored: $M = 12.87$, $SD = 2.73$; NoPosFramed-LowScored: $M = 13.38$, $SD = 3.16$)

However, we also evaluated participants' perception of their revision quality (Section 4.4.6) and found that *AI-reframed positive summary* significantly influenced participants' perception of the effort and quality of their revisions. The result of a two-way ANOVA with ART showed that regardless of receiving a high or low overall evaluation, participants perceived their effort and the revision quality was significantly higher when the critical feedback came with an AI-reframed positive summary, compared with the

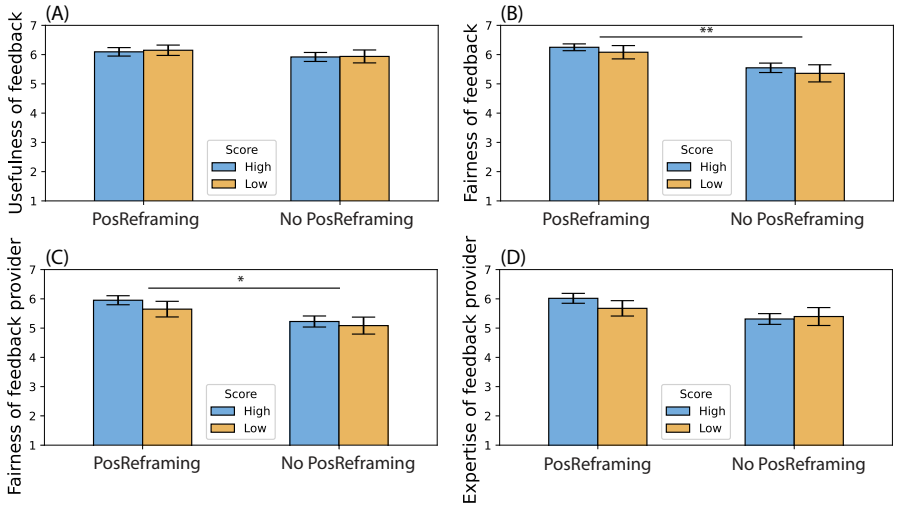

Figure 4: Bar charts with a mean (± standard error of the mean) of participants' perceived usefulness (left) and fairness (right) of the feedback (A, B), and fairness and expertise of the reviewer (C, D). The X-axis shows the factor of the presence of *AI-reframed positive summary*. The Y-axis shows participants' perceived usefulness and fairness of the feedback. A higher score indicates higher perceived usefulness (A), fairness (B, C), and expertise (D) of the feedback and reviewers (* indicates $p < .05$, ** indicates $p < .01$).

original critical feedback ($F[1, 133] = 5.56$, $p < .05$, $\eta_p^2 = .04$; PosFramed: $M = 5.62$, $SD = 1.08$; NoPosFramed: $M = 5.15$, $SD = 1.21$). (Figure 5, right)

Taken together, to answer RQ3, our result suggested that receiving feedback with *a low-scored overall evaluation* encouraged participants to revise their writing more than with a high-scored overall evaluation. However, we did not observe any significant effect of either *AI-reframed positive summary* or *overall feedback evaluation* on participants' quality of revision, though participants perceived their revision quality was increased when receiving *AI-reframed positive summary*.

5.4 Factors that Influence Receivers' Motivation to Revise (RQ4)

Based on the thematic analysis of the open-ended comments in the survey, we identified seven factors that potentially influenced participants' motivation to make revisions. Please note that we analyzed participants' "claimed motivation," as they may not explicitly articulate their reasons for making or not making revisions in an open-ended survey format. Additionally, this list of motivational factors is not exhaustive; it only includes those we identified based on our current data collection approach. We analyzed the data using the *attribution of motivation theory* to understand how people explain their motivation for making revisions. This analysis allows us to know what aspects can be taken care of if we want to leverage LLMs to polish critiques to support the delivery of peer feedback. Among the seven factors, we found that four of the factors were related to *intrapersonal attribution of motivation*, where people attribute their motivation to revise to themselves. We also found that three factors were related to *interpersonal attribution of motivation*, where people attribute their motivation to revise to external things, including the feedback itself and the reviewers. Note that the thematic analysis was done with all data without comparing the percentage of each time quantitatively across conditions.

We present *intrapersonal factors* that influenced participants' motivation for revision first.

5.4.1 *Intrapersonal factor: Actionability of the feedback.*

Participants' decisions to revise or not could be influenced by how feasible they thought they could address the issue in the revision process. They would balance their expected quality and efforts to make revisions. For instance, P28 shared, *"Overall, I did not want to throw away too much of the initial cover letter, so I tried to add the feedback I thought was easier to insert into the cover letter."* (P28, *NoPosFramed-HighScored* condition.) P70 also stated, *'"Some of the feedback was too generic, so I didn't implement it. When it was specific, I did it."* (P70, *NoPosFramed-LowScored* condition.)

5.4.2 *Intrapersonal factor: Alignment with personal goals/values.*

Some other participants shared that their motivation for revision was influenced by whether and how much the critique aligned with their personal goals or values for the given writing task. Their motivation to challenge themselves, reflect their authentic self in the writing, or their own standard about the writing can all influence their decision for revision. For example, P02 noted, *"I completely ignored the rudest parts. My writing is fine and not too casual for a cover letter, so I didn't listen to that. I did add some more specific information regarding the social media numbers I hit, but I disagree that a cover letter is the right place for that. Concise is better."* (P02, *NoPosFramed-HighScored* condition.) P126 also shared that *"It depended on the way in which I worded my letter and still being true to myself. "*(P126, *PosFramed-LowScored* condition.)

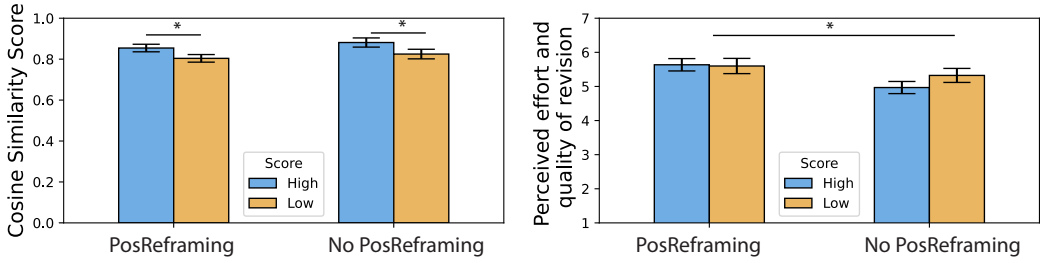

Figure 5: Two bar charts with a mean (± standard error of the mean) of cosine similarity score for the revision outcome (left) and perceived effort and quality of revision (right). The X-axis shows the factor of the presence of *AI-reframed positive summary*. The Y-axis for the left figure shows the cosine similarity score, ranging from zero to one. A score closer to one indicates a larger difference between the initial writing and final revisions. The Y-axis for the right figure shows participants' perceived effort and quality of revision. A higher score indicates increased effort and quality for the revision (* indicates $p < .05$).

5.4.3 ***Intrapersonal factor: Emotional response.*** Some participants explicitly mentioned their emotions when deciding what to revise or not. When emotion was mentioned in the open-ended comments, almost all of them were negative emotions, such as feeling "overwhelmed", "disrespected", "disappointed", "messed up", or "annoyed". This result was consistent with the finding in Section 5.1.1, which showed that participants experienced negative emotions significantly more when not receiving AI-reframed positive summaries with the critiques. As exemplified in P88's comment, *"I didn't like the way because I felt attacked by the person who replied to my letter. I only changed a few things about what I wrote initially because I didn't have enough interest in completing the task as it was asked of me."* (P88, *NoPosFramed-LowScored* condition.) P76 also described, *"I had feelings of annoyance since I believe that the person who reviewed my essay does not understand my point of view."* (P76, *NoPosFramed-LowScored* condition.)

5.4.4 ***Intrapersonal factor: Motivation to learn and improve.*** Participants' motivation to improve themselves and learn new things tended to make them have a positive attitude while interpreting the critiques they received. Those participants explicitly mentioned that critiques helped them improve their writing or enabled them to know which aspects should be improved. For example, P40 shared, *"I tried to follow all instructions because I felt I could really learn from them and apply them to my real life. I had to write many cover letters when I was searching for a job, [...]. I feel I can improve my writing skills from now on. Thank you."* (P40, *PosFramed-HighScored* condition) Similarly, P95 noted, *"I tried to take almost every feedback because it was important to me to improve my letter."* (P95, *NoPosFramed-LowScored* condition)

Next, we present *interpersonal factors* that influenced participants' motivation for revision.

5.4.5 ***Interpersonal factor: Perceived quality of the feedback.*** The perceived fairness and constructiveness also influenced participants' motivation to make revisions. Participants often described the feedback as "on point", "reliable", "constructive", "honest", "specific", or "fair" when they decided to follow the feedback for revision. For instance, P07 commented, *"Feedback was very constructive and genuine, which helped me to fix the errors and improve my writing."* (P07, *NoPosFramed-HighScored* condition). P53 also noted, *"I took most of the feedback and was grateful for the criticism. I appreciated that the faults were not only pointed out but possible solutions were also provided."* (P53, *PosFramed-HighScored* condition). This theme, combined with the quantitative finding in Section 5.2.1, suggested that high-quality feedback could motivate people to make revisions when receiving critiques.

5.4.6 ***Interpersonal factor: Perceived expertise/effort of the reviewers.*** Some participants mentioned that their motivation to revise was mainly due to the perceived expertise or the perceived effort of the reviewers. For those who mentioned the perceived characteristics of the reviewers, most of them thought reviewers were "skilled", "trustworthy", "professional", "experienced", "knowledgeable", or "making excellent points". As P123 shared, *"I was very happy with the feedback as it highlighted all the aspects that my writing lacked. I realized that the reviewer was very skilled in the points and clear explanations they provided. Adding to what I previously lacked made me feel that the cover letter now stood a greater chance of being approved."* (P123, *PosFramed-LowScored* condition) This theme, combined with the quantitative finding in Section 5.2.2, suggested that the perceived expertise of the reviewers could also play a part in motivating people to make revisions when receiving critiques.

5.4.7 ***Interpersonal factor: Tone of the feedback.*** In addition to the content of feedback, the tone of the feedback can also influence participants' motivation to make revisions. When describing the tone of the feedback, participants mentioned that the feedback sounded "aggressive", "lacked empathy", "harsh", or "authoritarian". AS P94 wrote, *"It was quite aggressive and made me less interested in the role [of the job post]. Some positive feedback is much nicer, and the delivery was very poor."* (P94, *NoPosFramed-LowScored* condition) Moreover, P96 described, *"The feedbacks were on point but lacked a bit of empathy. I didn't lie in my cover letter regarding my experience, which is not related to this field, and that makes the exercise more difficult, I believe. [...] "* (P96, *NoPosFramed-LowScored* condition).

Note that all factors we presented above were found in all conditions, though the distribution of the factors was slightly different across conditions. Table 4 and Figure 6 showed the distribution of each factor under each condition. In Table 4, we quantified each theme for each condition to show its distribution so readers could

Table 4: Distribution of the percentage of each factor that influenced authors' decisions to make revisions.

Factors that influence feedback revision			Condition				Total
			NoPosFramed-HighScored	PosFramed-HighScored	NoPosFramed-LowScored	PosFramed-LowScored	
Intrapersonal factor	Actionability of the feedback	Count	1	1	4	2	8
		% within column	3.2 %	2.7 %	12.5 %	5.4 %	5.8 %
	Alignment with personal goals/values	Count	8	2	2	4	16
		% within column	25.8 %	5.4 %	6.3 %	10.8 %	11.7 %
	Emotional response	Count	4	1	5	3	13
		% within column	12.9 %	2.7 %	15.6 %	8.1 %	9.5 %
	Motivation to learn and improve	Count	7	3	5	2	17
		% within column	22.6 %	8.1 %	15.6 %	5.4 %	12.4 %
Interpersonal factor	Perceived expertise/effort of the reviewers	Count	3	6	2	3	14
		% within column	9.7 %	16.2 %	6.3 %	8.1 %	10.2 %
	Perceived quality of the feedback (fairness, constructive)	Count	3	16	5	14	38
		% within column	9.7 %	43.2 %	15.6 %	37.8 %	27.7 %
	Tone of the feedback	Count	0	1	2	2	5
		% within column	0.0 %	2.7 %	6.3 %	5.4 %	3.7 %
	Others	Count	5	7	7	7	26
		% within column	16.1 %	18.9 %	21.9 %	18.9 %	19.0 %
	Total	Count	31	37	32	37	137
		% within column	100.000 %	100.000 %	100.000 %	100.000 %	100.000 %

have extra information to interpret the themes we identified above. In Figure 6, we merged "Actionability of the feedback", "Alignment with personal goals/values", "Emotional response", and "Motivation to learn and improve" together to label them as "Intrapersonal factor". For "Interpersonal factor", we merged "Perceived expertise/effort of the reviewers", "Perceived quality of the feedback (fairness, constructive)", and "Tone of the feedback" together. Although we could not come to any conclusion about whether an AI-reframed positive summary or showing an overall evaluation significantly influenced certain ways for participants to make attribution of motivation for making revisions, this qualitative analysis gave us insights into what factors influence authors' motivation to iterate their writing when receiving critiques. In the next section, we discuss ways to leverage LLMs to polish or reframe critiques based on seven common factors that influence people's attribution of their motivation for making revisions.

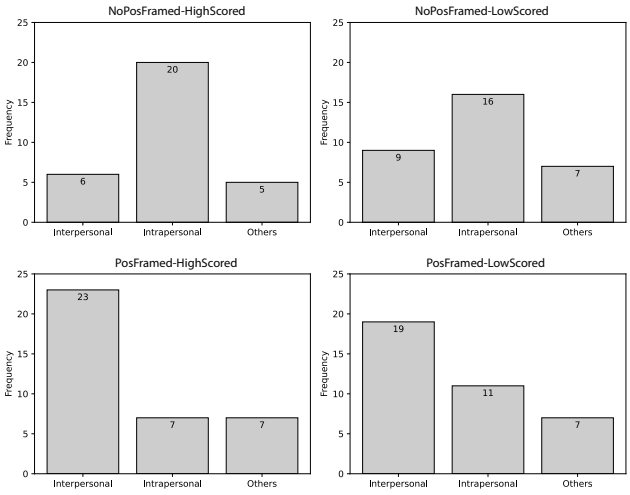

Figure 6: Distribution of different ways participants attribute their motivation for making revisions.

6 Discussion

Our results showed that receiving critique with an AI-reframed positive summary and seeing a high overall evaluation led authors to have positive emotions. The AI-reframed positive summary also enabled authors to experience increased autonomy and competence, and they felt that both the quality of the feedback itself and the fairness of the reviewers were better than only receiving critical feedback. Surprisingly, the AI-reframed positive summary did not have much influence on the revision outcome. Instead, receiving a low overall evaluation made authors revise more than receiving a high overall evaluation. Through the thematic analysis, we also identified seven key factors that influenced authors' motivation to make revisions or not after receiving critiques. We summarized our key findings in Table 5.

6.1 The Effect of Showing AI-reframed Positive Summary on Critique Reception at Perceptual-Level

Consistent with prior literature documenting that the use of positive language can encourage authors when receiving critiques [41], our study extends these findings by showing that reframing critiques by highlighting the positive aspects further enhances critique reception (Section 5.1, 5.2). Past research has shown that incorporating praise alongside critiques can help soften the tone of criticism [26], reinforcing appropriate behaviors, reducing the perceived threat in feedback giver-receiver relationships, and fostering the feedback receivers' self-esteem [13]. However, related works also suggest that general praise or veiled praise that shows indirect criticism makes authors feel confused and discourages revision [12]. Instead of directly praising the authors, our proposed AI-reframed positive summary was designed to help people see the positive side of criticism while keeping reviewers' original critiques. The result suggested that positive summaries reframed by AI not only softened the critiques but also empowered authors in the revision process through increased autonomy and competence.

Interestingly, appending an AI-reframed positive summary to critiques also positively influenced authors' perception of the critique and reviewers (Section 5.2, Section 5.4.5, and Section 5.4.6). This finding highlights once again that peer feedback not only helps individual authors improve their work but also establishes a positive impression of the review itself and reviewers. This positive perception could lay a strong foundation for establishing trust with other peer reviewers/experts, which can benefit the research community or the peer feedback platforms more broadly in the long run. Both anecdotal evidence and the literature have suggested that low-quality and poor peer review practice caused a negative impact on the evolution of science, including demotivating researchers from staying in a research community [24, 38, 52]. Thus, to prevent people from getting discouraged by critical feedback during peer reviews and potentially leaving the research community, which could result in only those who are emotionally tough or indifferent prevailing, we invite the research community to think about the possibility of integrating AI-generated positive summaries with the critical feedback in peer reviews. This would help to preserve diversity in the research community and broaden the way topics are studied [48].

Meanwhile, we should make AI's involvement transparent to the authors when deploying AI-assisted feedback outside the controlled experiment setting. Literature on AI-mediated communication has shown that distrust from receivers to senders arises when receivers are aware that AI was involved in generating text that could also be generated by a mixture of humans and AI [28]. To alleviate potential distrust that authors might feel towards reviewers when integrating AI-reframed positive summaries into the peer review process, one possible strategy is to present the AI-generated section separately from the human-written review, while clearly indicating AI's role. Although the impact of revealing AI's involvement in AI-assisted peer feedback has not been explored in the current scope, and we could only disclose the use of AI at the end, future research can examine the effects of disclosing AI's role at different review writing stages in the peer review process. This could lead to a better understanding of how such disclosures affect the receivers' (authors') trust toward the sender (reviewer) in AI-assisted peer feedback.

Unexpectedly, receiving a positive summary reframed by AI only changed authors' perceptions, not their actual quality of revision (Section 5.3). Prior research also did not find any significant improvement in the quality of writing when authors received either positive (i.e., praise) or negative (i.e., blame) feedback, but positive feedback led to authors' increased favorable attitudes and motivation [57]. One possible reason for the lack of improvement in revision quality in our study may be that the cognitive reframing strategy we employed was quite generic. In the current design of the prompt for the AI-reframed positive summary, we encouraged participants to adopt a positive attitude and consider the positive aspects of a negative situation. However, the effectiveness of cognitive reframing could vary depending on whether individuals' learning styles [63] match with the specific cognitive reframing strategy or depending on the different focus of reframing strategies, such as rationality or actionability [49]. Based on our results, future work could explore different reframing strategies for LLMs to generate various types of reframed summaries. This could involve considering authors' learning styles or preferred thinking styles and examining how different types of AI-reframed positive summaries lead to specific behavioral changes in the context of peer feedback.

6.2 The Effect of Showing Low Overall Evaluation on Critique Reception at Behavioral-Level

Our result showed that regardless of receiving AI-reframed summaries, when people received a low overall evaluation for their writing, they made more revisions than those who received a high overall evaluation (Section 5.3). Prior research has shown that feedback highlighting discrepancies between the authors' current performance and their desired standards can lead to more substantial efforts for authors to close the gap [34]. In our study, participants who received low scores might be able to directly see the gap between "where they are" and "where they aim to be" [22], thus taking feedback more seriously and engaging more deeply with the revision process. Past studies suggested that individuals who receive negative feedback often engage in deeper cognitive processing to understand the topic and rectify their mistakes [22]. Thus, it is possible that receiving a low-scored overall evaluation made authors scrutinize the feedback more thoroughly and revise extensively to reduce the discrepancies between their writing and expected goals.

Though receiving a low-scored overall evaluation led to authors' increased revisions, receiving too much negative feedback can be harmful and discouraging, especially for those who have low self-efficacy in writing [43, 57]. It has been found that people with low writing self-efficacy were more likely to be negatively affected by critical feedback [43]. These writers showed decreased motivation and less engagement in revising their work, indicating that their perception of their abilities influenced how they responded to feedback [43]. When authors doubt their capabilities, overwhelming criticism can lead to feelings of inadequacy, reduced motivation, and possibly a decline in writing performance. Therefore, we suggest that when reviewers want to motivate authors to significantly revise their writing during the revision process, combining a low overall score critique with positive reframed summaries could be effective in maintaining the right balance between authors' perceptions and revision behavior.

6.3 Incorporating Intrapersonal and Interpersonal Factors in AI-assisted Peer Feedback

Based on the seven motivational factors we identified in Section 5.4 and the distribution of the motivational factors in Table 4, it revealed that authors considered both *intrapersonal* and *interpersonal* factors when deciding whether to make revisions. Current practice in peer feedback often provides reviewers with a rubric to guide the review writing process. Our findings pointed out that in addition to following rubrics, incorporating *intrapersonal* and *interpersonal* factors into peer reviews or meta-reviews might be an effective way to motivate authors to make revisions. For example, future AI-assisted peer review systems could scaffold and help reviewers as they write meta-review by incorporating intrapersonal

Table 5: Research questions and summary of the key findings

Research questions	Main findings
RQ1: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) affect the authors' intrapersonal perception when receiving critical feedback?	**H1a**: Receiving AI-reframed positive summary and a high overall evaluation of the feedback led people to **experience a positive emotion**. (supported) **H1b**: Whether receiving AI-reframed positive summary or receiving high/low overall evaluation did not change people's **self-efficacy in writing** when receiving critical feedback. (not supported) **H1c**: Receiving critical feedback came with an AI-reframed positive summary **increased people's autonomy and competence in making revisions**. (supported)
RQ2: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) affect the authors' interpersonal perception when receiving critical feedback?	**H2a**: Receiving AI-reframed positive summary **increased the perceived fairness of the critical feedback**. (partially supported) **H2b**: Receiving AI-reframed positive summary **enhanced the perceived fairness of the feedback provider**. (partially supported)
RQ3: How does the combination of the presence of AI-reframed positive summary (present/absent) and different overall evaluations (high/low) influence the authors' revision outcomes when receiving critical feedback?	Receiving feedback with a low-scored overall evaluation encouraged people to **revise their writing more** than with a high-scored overall evaluation. Although little influence was found on participants' revision quality, participants **perceived their revision quality was significantly better** when receiving an AI-reframed positive summary.
RQ4: What factors influence the authors' motivation to revise their writing when receiving critical feedback?	**Intrapersonal factors:** Actionability of the feedback, alignment with personal goal/value, emotional response, motivation to learn and improve; **Interpersonal factors:** perceived quality of the feedback, tone of the feedback, perceived expertise/effort of the reviewer.

factors, such as "enhancing the actionability of the feedback," and "encouraging authors to make improvements," as well as interpersonal factors, such as "reframing the critique in a friendly tone," and "showing the efforts reviewers have put into writing reviews", into the meta-review.

Meanwhile, we acknowledge that we could not fully capture all possible motivational factors with the current online controlled experiment. Although we attempted to qualitatively analyze the claimed motivators and present their quantified distribution, future studies could explore motivations by building on the factors identified in this study and incorporating additional relevant measures (e.g., semi-structured interviews) to enhance the robustness of the findings.

7 Limitations and Future Work

Though we identified the positive impact of AI-reframed positive summaries on critique reception, we acknowledge that our study has some limitations. Firstly, we designed a writing task that asked participants to write a cover letter to apply for a job post. In light of the potential challenges of having researchers write about their research and then being processed by GPT, which could compromise the confidentiality of their work during the study, we have chosen to employ a writing task that can be well-controlled in an experimental setting. Writing a cover letter allowed participants to write something that they care about to a certain extent so that we can simulate the situation where people receive critique for something that matters to them. Moreover, writing a cover letter is a relatively short task that can be completed within a reasonable amount of time. However, we acknowledge that writing a cover letter is different from writing research articles, and we need to be cautious when interpreting the generalizability of our findings. To enhance the generalizability of our findings, future work could expand the scope of the study from a controlled experiment to a large-scale field study or explore the possibility of analyzing the dataset of peer review from some research communities. With a field study or dataset analysis, we may also have a sufficiently large sample size to include the factor of initial writing quality and enhance the current understanding.

Secondly, though we found that appending a positively reframed summary to peer feedback supported critique reception, future studies can explore the best way to present the positively reframed content. We appended the positively reframed summary at the end of the critique in the current setting, which may be influenced by recency bias [7, 17], where people tend to give greater importance to recent events (i.e., positive summary) than historical ones (i.e., critical feedback). Future work can examine whether recency bias occurs in a peer review context and how to cope with it with the suitable presentation of the positively reframed summary.

This study explored and evaluated the possibility of using LLMs to reframe human-written critiques in a positive way and demonstrated its positive impact on authors' emotions, perception of the

critique, and perception of the reviewer. Based on the current finding, we encourage the research community to explore different ways in which LLMs could be beneficial in supporting reviewers in delivering constructive peer feedback together. For instance, incorporating an option in the peer review system for reviewers and/or authors to activate the AI-reframed feature while composing and receiving critiques.

8 Conclusion

When critiques are delivered in a harsh tone in peer review, it can make people feel negative emotions and discourage them from revising their work. We proposed using AI to reframe human-written critique in a positive way and investigated its effect on the perception of the critique when it is combined with high or low overall evaluations. We found that adding AI-reframed positive summaries to critique made authors experience positive emotions, increased autonomy and competence, and perceive both the critique and reviewers as fair. In contrast, receiving low-scored overall evaluations increased the amount of revisions people made. Furthermore, people tended to attribute their motivation for making revisions to interpersonal factors when receiving AI-reframed positive summaries. We discussed the opportunity for combining a low-scored overall evaluation and an AI-reframed positive summary of critical feedback for a constructive and friendly peer review process.

Acknowledgments

We extend our gratitude to all the anonymous participants who took part in this study. We also appreciate the constructive feedback provided by the reviewers, which has significantly enhanced this work. Lastly, we would like to thank Jiayi Yang and Xiaotong Li for their valuable assistance in evaluating all the written outcomes from the participants.

References

[1] ACM Conference on Computer-Supported Cooperative Work and Social Computing. 2016. Instructions for CSCW Student Mentor Program. https://cscw.acm.org/2016/volunteer/InstructionsCSCWStudentMentorProgram.pdf. Accessed: 2024-08-15.

[2] ACM Conference on Human Factors in Computing Systems. 2024. Guide to Reviewing Papers for CHI 2024. https://chi2024.acm.org/submission-guides/guide-to-reviewing-papers/. Accessed: 2024-08-15.

[3] ACM SIGGRAPH 2024. 2024. Technical Papers Reviewer Instructions and Ethics Guidelines. https://s2024.siggraph.org/technical-papers-reviewer-instructions-ethics/. Accessed: 2024-08-15.

[4] ACM Symposium on Virtual Reality Software and Technology. 2023. Reviewing Guidelines for VRST 2023. https://vrst.acm.org/vrst2023/reviewing/. Accessed: 2024-08-15.

[5] ACM Transactions on Computer-Human Interaction. 2024. Reviewer Guidelines for ACM TOCHI. https://dl.acm.org/journal/tochi/reviewer-guidelines. Accessed: 2024-08-15.

[6] American Geophysical Union. 2024. Review Criteria. https://www.agu.org/publications/reviewers/review-criteria. Accessed: 2024-08-15.

[7] Norman H Anderson and Stephen Hubert. 1963. Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior* 2, 5-6 (1963), 379–391.

[8] Anika Batenburg and Enny Das. 2014. An experimental study on the effectiveness of disclosing stressful life events and support messages: When cognitive reappraisal support decreases emotional distress, and emotional support is like saying nothing at all. *PLoS One* 9, 12 (2014), e114169.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Ruth Butler. 1987. Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of educational psychology* 79, 4 (1987), 474.

[11] Julia Cambre, Scott Klemmer, and Chinmay Kulkarni. 2018. Juxtapeer: Comparative peer review yields higher quality feedback and promotes deeper reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[12] Maria Cardelle and Lyn Corno. 1981. Effects on second language learning of variations in written feedback on homework assignments. *Tesol Quarterly* 15, 3 (1981), 251–261.

[13] Brian Cavanaugh. 2013. Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children* 36, 1 (2013), 111–136.

[14] David A Clark. 2013. Cognitive restructuring. *The Wiley handbook of cognitive behavioral therapy* (2013), 1–22.

[15] Srayan Datta, Chanda Phelan, and Eytan Adar. 2017. Identifying misaligned inter-group links and communities. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–23.

[16] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors. *ACM Transactions on Computer-Human Interaction* 31, 2 (2024), 1–40.

[17] Joseph P Forgas. 2011. Can negative affect eliminate the power of first impressions? Affective influences on primacy and recency effects in impression formation. *Journal of Experimental Social Psychology* 47, 2 (2011), 425–429.

[18] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[19] Sandra Graham. 2020. An attributional theory of motivation. *Contemporary Educational Psychology* 61 (2020), 101861.

[20] Michael D Greenberg, Matthew W Easterday, and Elizabeth M Gerber. 2015. Critiki: A scaffolded approach to gathering design feedback from paid crowdworkers. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 235–244.

[21] Marlo Häring, Wiebke Loosen, and Walid Maalej. 2018. Who is addressed in this comment? Automatically classifying meta-comments in news comments. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[22] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[23] Catherine M Hicks, Vineet Pandey, C Ailie Fraser, and Scott Klemmer. 2016. Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 458–469.

[24] Hugo Horta and Jisun Jung. 2024. The crisis of peer review: Part of the evolution of science. *Higher Education Quarterly* (2024), e12511.

[25] Wenjian Huang, Tun Lu, Haiyi Zhu, Guo Li, and Ning Gu. 2016. Effectiveness of conflict management strategies in peer review process of online collaboration projects. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 717–728.

[26] Fiona Hyland and Ken Hyland. 2001. Sugaring the pill: Praise and criticism in written feedback. *Journal of second language writing* 10, 3 (2001), 185–212.

[27] International Journal of Psychiatry Research. 2021. Article from International Journal of Psychiatry Research. https://www.psychiatricjournal.net/article/view/44/2-2-18. Accessed: 2024-08-15.

[28] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[29] Yvonne Jansen, Kasper Hornbæk, and Pierre Dragicevic. 2016. What Did Authors Value in the CHI'16 Reviews They Received?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 596–608.

[30] Hyeonsu B Kang, Gabriel Amoako, Neil Sengupta, and Steven P Dow. 2018. Paragon: An online gallery for enhancing design feedback with visual examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[31] Fatma Kaya. 2021. Emotions related to identifiable/anonymous peer feedback: A case study with Turkish pre-service English teachers. *Issues in Educational Research* 31, 4 (2021), 1088–1100.

[32] Jacalyn Kelly, Tara Sadeghieh, and Khosrow Adeli. 2014. Peer review in scientific publications: benefits, critiques, & a survival guide. *Ejifcc* 25, 3 (2014), 227.

[33] Eun Jung Kim and Kyeong Ryong Lee. 2019. Effects of an examiner's positive and negative feedback on self-assessment of skill performance, emotional response, and self-efficacy in Korea: a quasi-experimental study. *BMC medical education* 19 (2019), 1–7.

[34] Avraham N Kluger and Angelo DeNisi. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin* 119, 2 (1996), 254.

[35] Markus Krause, Tom Garncarz, JiaoJiao Song, Elizabeth M Gerber, Brian P Bailey, and Steven P Dow. 2017. Critique style guide: Improving crowdsourced design feedback with a natural language model. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 4627–4639.

[36] Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Yi Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Scott Smith, Yian Yin, et al. 2024. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI* (2024), AIoa2400196.

[37] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P Dow. 2015. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 473–485.

[38] Kyle McCloskey and Jon F. Merz. 2022. A Pilot Survey of Authors' Experiences with Poor Peer Review Practices. *bioRxiv* (2022). https://doi.org/10.1101/2022.12.20.521261 arXiv:https://www.biorxiv.org/content/early/2022/12/28/2022.12.20.521261.full.pdf

[39] Marlee Mercer and Duygu Biricik Gulseren. 2024. When negative feedback harms: a systematic review of the unintended consequences of negative feedback on psychological, attitudinal, and behavioral responses. *Studies in Higher Education* 49, 4 (2024), 654–669.

[40] Kim M Mitchell, Tom Harrigan, Torrie Stefansson, and Holly Setlack. 2017. Exploring self-efficacy and anxiety in first-year nursing students enrolled in a discipline-specific scholarly writing course. *Quality Advancement in Nursing Education-Avancées En Formation Infirmière* 3, 1 (2017), 4.

[41] Thi Thao Duyen T Nguyen, Thomas Garncarz, Felicia Ng, Laura A Dabbish, and Steven P Dow. 2017. Fruitful Feedback: Positive affective language and source anonymity improve critique reception and work outcomes. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1024–1034.

[42] OpenAI. 2023. GPT-4 System Card. https://cdn.openai.com/papers/gpt-4-system-card.pdf. Accessed: 2024-08-15.

[43] Frank Pajares and Margaret J Johnson. 1994. Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English* 28, 3 (1994), 313–331.

[44] Proceedings of the National Academy of Sciences. 2024. Reviewer Guidelines. https://www.pnas.org/reviewer. Accessed: 2024-08-15.

[45] Candace M Raio, Temidayo A Orederu, Laura Palazzolo, Ashley A Shurick, and Elizabeth A Phelps. 2013. Cognitive emotion regulation fails the stress test. *Proceedings of the National Academy of Sciences* 110, 37 (2013), 15139–15144.

[46] Anne S Rasmussen and Dorthe Berntsen. 2009. Emotional valence and the functions. *Memory & Cognition* 37 (2009), 477–492.

[47] Eugenia Ha Rim Rho, Gloria Mark, and Melissa Mazmanian. 2018. Fostering civil discourse online: Linguistic behavior in comments of# metoo articles across political perspectives. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–28.

[48] Sandra L Schneider. 2001. In search of realistic optimism: Meaning, knowledge, and warm fuzziness. *American psychologist* 56, 3 (2001), 250.

[49] Ashish Sharma, Kevin Rushton, Inna Lin, David Wadden, Khendra Lucas, Adam Miner, Theresa Nguyen, and Tim Althoff. 2023. Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 9977–10000.

[50] Ashish Sharma, Kevin Rushton, Inna Wanyin Lin, Theresa Nguyen, and Tim Althoff. 2024. Facilitating self-guided mental health interventions through human-language model interaction: A case study of cognitive restructuring. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–29.

[51] Kennon M Sheldon, Andrew J Elliot, Youngmee Kim, and Tim Kasser. 2001. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology* 80, 2 (2001), 325.

[52] Nyssa J Silbiger and Amber D Stubler. 2019. Unprofessional peer reviews disproportionately harm underrepresented groups in STEM. *PeerJ* 7 (2019), e8247.

[53] C Estelle Smith, William Lane, Hannah Miller Hillberg, Daniel Kluver, Loren Terveen, and Svetlana Yarosh. 2021. Effective strategies for crowd-powered cognitive reappraisal systems: A field deployment of the flip* doubt web application for mental health. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[54] Lu Sun, Aaron Chan, Yun Seo Chang, and Steven P Dow. 2024. ReviewFlow: Intelligent Scaffolding to Support Academic Peer Reviewing. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 120–137.

[55] Lu Sun, Stone Tao, Junjie Hu, and Steven P Dow. 2024. MetaWriter: Exploring the Potential and Perils of AI Writing Support in Scientific Peer Review. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.

[56] Amy Rupiper Taggart and Mary Laughlin. 2017. Affect Matters: When Writing Feedback Leads to Negative Feeling. *International Journal for the Scholarship of Teaching and Learning* 11, 2 (2017), n2.

[57] Winnifred F Taylor and Kenneth C Hoedt. 1966. The effect of praise upon the quality and quantity of creative writing. *The Journal of Educational Research* 60, 2 (1966), 80–83.

[58] Allison S Troy, Frank H Wilhelm, Amanda J Shallcross, and Iris B Mauss. 2010. Seeing the silver lining: cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion* 10, 6 (2010), 783.

[59] Humboldt State University. n.d.. Cover Letter Scoring Rubric. https://acac.humboldt.edu/sites/default/files/COVER%20LETTER%20SCORING%20RUBRIC.pdf Accessed: 2024-11-07.

[60] Indiana State University. n.d.. Cover Letter Rubric. https://www.indstate.edu/sites/default/files/media/documents/pdf/cover-letter-rubric.pdf Accessed: 2024-11-07.

[61] Western Carolina University. n.d.. Cover Letter Rubric. https://www.wcu.edu/WebFiles/CCPD-Cover-Letter-Rubric.pdf Accessed: 2024-11-07.

[62] U.S. Merit Systems Protection Board. 2024. The Roles of Feedback, Autonomy, and Meaningfulness in Employee Performance Behaviors. https://www.mspb.gov/studies/researchbriefs/The_Roles_of_Feedback_Autonomy_and_Meaningfulness_in_Employee_Performance_Behaviors_1548113.pdf. Accessed: 2024-08-29.

[63] Karlijn van Doorn, Freda McManus, and Jenny Yiend. 2012. An analysis of matching cognitive-behavior therapy techniques to learning styles. *Journal of Behavior Therapy and Experimental Psychiatry* 43, 4 (2012), 1039–1044.

[64] Elizabeth Walsh, Maeve Rooney, Louis Appleby, and Greg Wilkinson. 2000. Open peer review: a randomised controlled trial. *The British Journal of Psychiatry* 176, 1 (2000), 47–51.

[65] Bernard Weiner. 2001. *Intrapersonal and Interpersonal Theories of Motivation from an Attribution Perspective*. Springer US, Boston, MA, 17–30. https://doi.org/10.1007/978-1-4615-1273-8_2

[66] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[67] Y Wayne Wu and Brian P Bailey. 2018. Soften the pain, increase the gain: Enhancing users' resilience to negative valence feedback. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[68] Y Wayne Wu and Brian P Bailey. 2021. Better feedback from nicer people: Narrative empathy and ingroup framing improve feedback exchange. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–20.

[69] Yu-Chun Grace Yen, Steven P Dow, Elizabeth Gerber, and Brian P Bailey. 2017. Listen to others, listen to yourself: Combining feedback review and reflection to improve iterative design. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 158–170.

[70] Yu-Chun Grace Yen, Joy O Kim, and Brian P Bailey. 2020. Decipher: an interactive visualization tool for interpreting unstructured design feedback from multiple providers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[71] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1005–1017.

[72] Hongli Zhan, Allen Zheng, Yoon Kyung Lee, Jina Suh, Junyi Jessy Li, and Desmond C Ong. 2024. Large language models are capable of offering cognitive reappraisal, if guided. *arXiv preprint arXiv:2404.01288* (2024).

[73] Jing Zhou. 1998. Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of applied psychology* 83, 2 (1998), 261.