Fruitful Feedback: Positive Affective Language and Source Anonymity Improve Critique Reception and Work Outcomes

Duyen T. Nguyen¹
duyenmarynguyen@fb.com

Thomas Gancarz¹
tgarncarz@gmail.com

Felicia Ng¹
fng@cs.cmu.edu

Laura A. Dabbish¹
dabbish@cmu.edu

Steven P. Dow²
spdown@ucsd.edu

¹Carnegie Mellon University
Pittsburgh, PA, USA

²UC San Diego
San Diego, CA, USA

ABSTRACT

Feedback is information that can improve task performance. Online communities, educational forums, and crowd-based feedback platforms all support feedback exchange among a more diverse set of sources than ever before, with greater control over how to moderate this exchange. In this work, we study how the power relationship between the source and receiver and the tone of language influence the receptivity, effort, and work performance resulting from online feedback exchange. We conducted an online experiment manipulating affective language and source of feedback on a writing task. We found that critiques with positive affective language increased positive emotions and reduced participants' annoyance and frustration, which led to an increase in work quality, compared to critiques without positive language. Feedback *without* positive affective language led to more edits, but not better work outcomes. Participants reacted more positively to feedback from an anonymous source than from a peer or an authority. Our findings provide design implications for platforms to support more fruitful feedback exchange.

Author Keywords

Feedback exchange; transparency; critique; emotional reactions to feedback; feedback systems

ACM Classification Keywords

H.5.3. Information interfaces and presentation (e.g., HCI): Computer-supported Collaborative Work

INTRODUCTION

Feedback on creative work can improve work quality and help people refine their skills [4, 19]. Online, people can exchange feedback with others all over the world. The rise of online MOOCs [20], work-sharing platforms like Dribbble or GitHub [25, 26], and crowd-based feedback platforms [23, 43] enable people to more quickly and conveniently obtain feedback from a range of sources—from experts to peers to anonymous sources—and with many

different contexts and styles. Although there are broader sources for feedback, classic challenges remain. People often find criticism unpleasant, react negatively to it, and perceive the feedback and source as unfair [5, 21]. Negative reactions to feedback can lead to reduced effort during revisions [10, 17], decreased motivation and commitment to task goals [17], and hindered cognitive abilities [33, 39].

Online feedback exchange happens through interfaces where designers can moderate how participants interact. Such interfaces can help recruit providers such as peers, experts, or anonymous crowd workers (e.g. [23, 42]) and provide language suggestions (e.g. [3, 23]). Given the abundance of possibilities for feedback language and sources, we focus on the following question: How can we design interfaces for feedback exchange to improve critique reception and work outcomes? In order to design interfaces for feedback, we need to understand how aspects of feedback, such as the source, style, and language influence feedback reception, and the subsequent effects on effort and performance.

Prior work suggests that when providing critical feedback, prepending some positive affective language before the criticisms, such as praise or mitigating language, results in more positive attitudes towards feedback [44] and better task outcomes than critical feedback without positive affective language [29]. What remains less clear is why positive affective language before critical feedback has these effects. Does the positive affective language increase motivation and effort, or does it affect how receivers process and learn from feedback?

Feedback source also plays an important role in feedback reception. People are more receptive to critique from higher status sources (e.g. from a manager or leader) versus peers [21]. People tend to value feedback from anonymous sources more than feedback from identified peers [22]. However, no study to our knowledge compares anonymous feedback sources with identified sources at different status levels (manager vs. peer). In other words, it is unclear if leaders or managers should also remain anonymous when giving feedback, or if anonymity only provides value for feedback exchanges among peers.

It is also important to understand how positive affective language in feedback interacts with information about source identity. Some studies have found that people are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CSCW '17, February 25–March 01, 2017, Portland, OR, USA
© 2017 ACM. ISBN 978-1-4503-4335-0/17/03...\$15.00
DOI: <http://dx.doi.org/10.1145/2998181.2998319>

more receptive to critical feedback from an authority than from peers [10, 21]. Does the influence of positive affective language depend on the source as well?

To examine the influence of positive affective language and feedback source on receptivity to feedback, we conducted a 2x3 factorial between-subjects experiment—with or without positive affective language x source identity (peer, authority, anonymous)—on Amazon’s Mechanical Turk (MTurk). 205 participants composed short news articles and received critical feedback on their writing. They then revised their articles, and self-reported emotional state and attitudes towards the content and source of feedback. We measured how much participants edited their articles after viewing the feedback, and evaluated the quality of their articles both before and after receiving feedback. Our results show that both using positive affective language and anonymizing the feedback source significantly improved feedback reception and final work performance. Participants who received critical feedback without positive language reported higher frustration and lower perceptions of fairness, and made significantly more edits during revision. The results provide implications for how online platforms can mediate more effective feedback exchanges.

FACTORS AFFECTING FEEDBACK EXCHANGE

In recent years, many online communities have sprouted up to facilitate worksharing and feedback exchange, such as Dribbble (<https://dribbble.com/>), Behance (<https://behance.com/>), Hunie and Forrst for design, or DeviantArt for visual art, PhotoSig for photography, GitHub for software development (<https://github.com/>), and crowd-feedback systems [23, 43]. These sites helped facilitate the exchange of honest, critical feedback among participants from novices to experts, from anonymous to semi-identified, or identified sources [23, 42]. Compared with face-to-face critiques, online settings allow more control over how feedback is presented or framed for the reviewer, including the general tone and availability of various identifying details about the source. Communication systems, for example, can analyze message tone and suggest changes to the framing or phrasing of messages (e.g. Crystal [3]). Platforms can mask feedback source identity or make it very prominent. Many platforms such as Dribbble or GitHub allow transparency, partial transparency, or anonymity of the identity of the feedback source [26]. Online platform design can enhance or downplay the source’s social status, by displaying of expertise levels, heat maps of contributions, or lists of past activities [26].

Previous work suggests that the source and tone of feedback matters in online contribution settings. Zhu et al. [44] found that in Wikipedia, feedback from newcomers was received negatively and easily rejected. They also found that feedback with positive and social tone could improve work motivation but had little effect on improving focal task performance, while feedback with negative and directive tone reduced work motivation but improve focal

task performance. They thus recommended that only senior members should deliver feedback to newcomers, and that depending on the objective of feedback (to improve motivation or to increase focal task performance), feedback should be framed with more positive or more critical tone. Previous work also suggests that the level and amount of information available about other participants in open contribution settings influences receptivity to contributions. Marlow & Dabbish [26] found that GitHub users formed impressions of, and even made decisions to collaborate with one another based on various personal details in the GitHub users’ profiles. Participants negatively viewed code change suggestions from newcomers to the site with little information on their profiles. Deeper understanding of how these two factors affect feedback reception and work performance will shed light on how feedback platform design can encourage learning and improve work quality.

Positive affective language

Studies suggest that the extent to which people value and follow feedback may depend on how it is expressed. Nelson & Schunn [29] defined affective language as “statements reflecting emotion, which may influence the implementation of feedback”. Examples of affective language in feedback include stand-alone praise (e.g. “Great job!”), inflammatory language (which is rare, e.g. “The writing quality is abysmal.”), and mitigating language applied to criticisms or suggestions (e.g. “This is a good start, but you should add some examples.”). When criticisms are offered with positive affective language, such as a mitigating statement (e.g. “Good effort!”) before listing the issues, people perceive the feedback source more positively, seeing them as more likeable with higher integrity [30], and improve their work more after feedback [29, 35].

Although previous work indicates that positive affective language leads to work improvements, it does not explain the mechanisms behind these improvements. Positive affective language can increase positive emotions and reduce negative feelings when receiving criticism [30]. This may improve people’s motivation [19], and hence, increase the effort they spend revising after feedback. Conversely, criticisms without mitigating language may be causing emotional dissonance that precludes people from learning from feedback. Other studies [e.g. 10, 11] also suggested (but did not prove) that positive affective language improved perception about the feedback provider, and thus would motivate participants to improve their work. Our research examines the link between participants’ emotional reactions to feedback and their behaviors and performance after receiving feedback.

RQ1: How does affective language influence receptivity to feedback, and subsequent effort and work outcomes?

Based on previous studies [e.g. 29, 30], we expect that compared to those who receive critical feedback without positive affective language, participants who receive critical feedback with positive language will be more receptive,

make more edits to their work, and improve their work quality. We also hypothesize that positive language will improve performance by enhancing perceptions of the feedback provider, as suggested by previous work [10, 11].

Source identity

Information about the feedback provider's identity also notably influences receptivity. The literature on feedback receptivity suggests that people perceive feedback from peers as less accurate and less useful than that from a source of higher social status, such as their managers or leaders [10, 21]. However, many interactive feedback systems leverage peers (or novices, more generally), e.g. [23, 43] since they are more readily available than experts. The bias against peer feedback presents a challenge for online platforms and interactive feedback systems. Anonymity may present a design solution. Previous work suggests that people are more receptive to anonymous sources than peers. In Lu & Bol [22], students rated anonymous peer feedback from their classmates, including critical feedback, as more useful for improving their writing than feedback from identified classmates. Anonymous feedback sources are also perceived to be more honest and objective than identified peers [22, 24].

Although previous studies suggest that source anonymity increases receptivity to feedback compared to identified peers, it is less clear how this compares with receptivity to feedback from an authority. Past work has not examined how source identity affects feedback receptivity and the amount of effort exerted in response to critique. Therefore, we ask:

RQ2: How does feedback from an anonymous source, from a source of higher status, and from a source of the same status influence participants' feedback receptivity, post-feedback behavior, and final work performance?

We expect that compared to people who receive feedback from an identified peer, those who receive feedback from an anonymous source will be more receptive to the feedback, make more edits to their work, and improve their work quality after feedback.

It is unclear from previous studies how positive affective language and the feedback source's identity will interact to influence feedback receptivity, revision behavior and resulting task performance. Critical feedback from superiors is better received than that from peers [10, 11, 21], and criticisms from junior members are more easily rejected than those from senior members [44]. At the same time some management scholars argue that positive affective language from supervisors, in the form of the feedback 'sandwich' can have negative effects on supervisor-employee relationships [40]. Does the influence of affective language depend on the feedback source? And how will positive affective language influence the reception of critical feedback from anonymous sources?

RQ3: Are the effects of positive affective language on participants' feedback reception, post-feedback behaviors, and work performances different when the feedback comes from different sources?

In the next section we describe the study we conducted to examine our research questions on feedback receptivity.

FEEDBACK RECEPTIVITY STUDY: METHOD

We conducted an online experiment where participants completed a writing task, received critical feedback on their work, and then edited and resubmitted their articles. We varied the use of positive affective language in the critical feedback and the identity of the feedback source.

Participants

We collected data from 205 participants (86 females) from MTurk, aged 18 to 65. A majority of participants were college educated at the bachelor degree level (54%). All participants were located in the US. All participants speak fluent English.

Writing task

Each participant wrote a newspaper article to promote a new product called Tree Change Dolls by Tasmanian artist Sonia Singh. Sonia collects discarded dolls from landfills, repairs them, and gives them a new, more down-to-earth look. We introduced the product and the artist to the participants, and provided a link to the product website (<http://treechangedolls.tumblr.com>). We asked participants to write a promotional article about Sonia's product for the front page of a local newspaper. We limited the newspaper article length to 150 words. Participants submitted their articles in a Google Form.

Feedback for the writing task

A blind-to-condition experimenter provided all participants with critical feedback on their newspaper articles. The critical feedback was selected from a pool of criticisms, similar to prior research [7]. Our feedback pool was generated through a pretest task where MTurk workers gave feedback on several short newspaper articles about Tree Change Dolls. We filtered the pool of critical feedback statements for statements that fit into one of the three categories drawn from literature about effective writing feedback [38]: comments about style, mechanics, and writing content. The experimenter selected three relevant feedback statements from the pool, one of each type, and connected them to make a feedback paragraph for each participant. Table 1 shows examples of the three types of feedback statements.

Experimental conditions

We manipulated the positive affective language in the

Type of feedback	Examples from the pool
About style	This piece is unprofessional and colloquial.
About mechanics	Wrong grammar in some sentences.
About content	Missed the recycled aspect of the dolls.

Table 1. The pre-authored feedback pool included criticisms on style, mechanics, and content.

	Feedback source's identity			
	Anonymous	Requester	Worker	Total
With-positive affective language	32	31	31	94
Without-positive affective language	37	38	33	108
Total	69	69	64	205

Table 2. Breakdown of participants across the six experimental conditions.

feedback provided to participants, and the identity of the feedback source. The experiment was a 2 (affective language conditions) x 3 (source identity conditions) factorial between-subjects design. Each participant was randomly assigned to one condition (see Table 2 for breakdowns).

Affective language manipulation

There were two affective language conditions: 1) feedback *with positive* language before it, and 2) feedback *without positive* language before it. All participants, regardless of condition, received feedback from the same pool of critical feedback. The only difference across conditions was the presence or absence of positive affective language. To create feedback prepended with positive affective language (with-positive-language condition), we created a pool of positive “wrapper” statements (see Figure 1b) that could be added to the beginning of the feedback. An example of writing feedback in the with-positive-language condition is: *“Good work so far. But the piece is boring and lacks enthusiasm. Wrong grammar in some sentences. Missed the recycling aspect of the dolls.”*

For the without-positive-language condition, we also prepended a “wrapper” statement before the feedback to ensure the overall length matched the with-positive-language condition. The without-positive-language “wrappers” are neutral in meaning (e.g. *“This needs revision”*, see Figure 1b).

Feedback source identity manipulation

Participants were randomly assigned to one of three source identity conditions. In the *Requester* condition, we told participants that the feedback came from the requester of

the HIT on Mechanical Turk. In the *Peer* condition, we told participants that the feedback came from another MTurk worker who also worked on the same HIT. In the *Anonymous* condition, the feedback was not attributed to a particular source (see Figure 1c).

Revision and resubmission of articles

After submitting the newspaper article for the writing task, participants received a link to a Google Form. This form showed them their original article submission, followed by the feedback for that article (see Figure 1a). We gave participants a place to edit their article, but we disabled their ability to copy and paste their original submission.

Post-experimental survey

After resubmitting their articles, participants answered an online post-experimental questionnaire. This survey consisted of questions about their emotions (see Table 4), questions about their attitude towards the feedback and the feedback source (see Table 3), and demographic questions.

Procedure

Our study consisted of two steps. In step 1, participants on MTurk completed the writing task HIT and were paid \$1.20. The experimenter then generated the feedback from the feedback pool, and used a Python script to randomly assign each participant to one of the six experimental conditions above. The feedback for each participant was formatted according to the experimental condition. We then sent participants a link to the feedback/editing task.

In step 2, participants followed the links to view their original articles and the feedback (see Figure 1). Then they edited their articles and resubmitted them. Finally, they filled out the post-experimental survey and were paid \$2.00. In total, each participant received \$3.20 for completing both step 1 and step 2.

Measures

We collected five dependent measures: participants’ emotions, perceptions of the feedback, perceptions of the feedback source, edit distance between the original articles and the revisions, and quality of the original and revised arti-

(a) Your original submission:

What better way to leave an impact on the world? A positive one at that! Sonia Singh creates an impression upon the youth (and older) citizens of the world by creating dolls from once thought used goods.

[SOURCE IDENTITY CONDITION]

[AFFECTIVE LANGUAGE CONDITION] This piece has grammatical mistakes and some expressions are too long.

Use the feedback and create a revision.

Write a revision of your work here.

(b) Affective language conditions

With-positive condition:

- "Good work so far, but..."
- "Good first revision, but..."
- "Good effort so far, but..."

Without-positive condition:

- "This needs revision; ..."
- "This needs more work; ..."
- "You need to put in more effort."

(c) Source identity conditions

Anonymous condition:

"Here is the feedback:"

Requester condition:

"Here is the feedback from the requester of the task:"

Peer condition:

"Here is the feedback from a fellow MTurker who also did the task."

Figure 1. (a) Feedback page with original submission, feedback, and a textbox for revision, (b) affective language conditions, and (c) source identity conditions

Measures	Indicate on a scale of 1 (strongly agree) to 7 (strongly disagree) how much you agree with the following:
Usefulness of feedback	<ul style="list-style-type: none"> - The feedback I received influenced my revision of the work. - I applied the feedback I received when I edited my work.
Fairness of feedback	<ul style="list-style-type: none"> - 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.
Fairness of feedback source	<ul style="list-style-type: none"> - 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.
Expertise of feedback source	<ul style="list-style-type: none"> - The person who gave me feedback is qualified. - The person who gave me feedback is an expert in the task. - The person who gave me feedback is knowledgeable.

Table 3. Survey items for perceptions of feedback and feedback source.

cles. We measured participants' self-efficacy and self-identification as a writer as control variables. We measured participants' perceptions of the feedback and feedback source with multi-item scales.

Emotions

We measured emotional response to feedback using an adapted version of the PANAS scale [41]. We asked participants to report their level of annoyance, frustration, happiness, and enthusiasm on a 7-point Likert scale.

Perceptions of feedback content

Participants rated their perceptions of the positivity or negativity of the feedback content on a 7-point Likert scale. Participants also rated the usefulness of the feedback with two items (Cronbach's α .78; see Table 3 for item phrasing). These items were based on previous studies about feedback receptivity [21, 22]. We measured perceived feedback fairness using four items adapted from Leung et al [21] (Cronbach's α .93; see Table 3).

Perceptions of feedback source

Participants rated their perceptions of the feedback source on a 7-point Likert scale. We measured perceived source fairness using four items adapted from Leung et al. [21] (Cronbach's α .92; see Table 3). We measured perceived source expertise using three items adapted from Nguyen et al. [31] (Cronbach's α .95; see Table 3).

Edit distance between original and revised articles

We measured revision effort after receiving feedback using the Levenshtein Distance (LD) [18] between participants' original and revised articles. LD indicates the number of character changes required to make two pieces of text equivalent [18]. LD is commonly used in studies about writing feedback to measure changes between texts [e.g., 8, 36]. Given the prior work and the experiment's size (> 400 submissions and revisions), LD provided a scalable measure of the editing effort between revisions.

To control for the confounding effect of article length on LD, an LD ratio was also calculated by dividing the LD by the character count of the original articles. However, our statistical analyses showed the same results for LD and the LD ratio. Therefore, we only report the analysis on LD measures in the results.

Quality of original and revised articles

In order to measure quality of the original and revised articles, we ran a task on MTurk where independent workers rated the original and revised article quality. Since each participant composed two articles (original and revised), we collected in total 410 articles from 205 participants. We randomly grouped these 410 articles into 41 blocks of 10 articles each, making sure that no two articles from the same participant belonged to the same block. Hence, for each participant, the original and revised articles were evaluated separately. For each of the 41 blocks, we recruited three MTurk workers to evaluate the articles on a scale of 1 (very poor) to 7 (very good). In total, we recruited 123 unique MTurk workers and none evaluated more than one block of articles. For each article, as we employed 3 raters to rate on an ordinal scale from 1 to 7, we computed Krippendorff's α [15] to access interrater reliability. The average Krippendorff's α is .84, indicating good reliability.

We ran a repeated-measure ANOVA comparing the ratings of the three workers in each block, and ratings among the blocks. We found no significant difference in the ratings of the three workers in each block ($F[2, 730] = 0.21$, n.s.), and no significant differences among the blocks ($F[41, 365] = 1.28$, $p = 0.12$). Therefore, we averaged the ratings of the three workers to get a quality measure for each article. We also computed the difference between the average rating of the revised article and the original article to get a measure of the change in work quality for each participant.

Self-efficacy

We measured self-efficacy using a four-item scale adapted from Chen et al. [6] (Cronbach's α .92).

Self-identification as a writer

We also asked participants to report on a 1 (strongly disagree) to 7 (strongly agree) Likert scale how much they identified themselves as writers.

Perception of the feedback source's identity

Participants in the anonymous conditions were asked to report who they assumed provided the feedback: the requester, a fellow worker, or someone else they specified.

	Condition					
	<i>With positive language conditions</i>			<i>Without positive language conditions</i>		
	Anonymous	Requester	Worker	Anonymous	Requester	Worker
Number of characters in original articles	703.53 (181.41)	703.48 (133.47)	738.74 (191.51)	705.59 (146.07)	718.30 (176.25)	673.23 (148.24)
Quality of original articles (1-low to 7-high)	4.38 (0.73)	4.17 (1.21)	4.54 (1.09)	4.03 (1.05)	4.66 (1.23)	4.40 (1.14)
Number of characters in revised articles	735.84 (155.20)	700.83 (195.41)	790.96 (208.89)	665.05 (212.08)	656.02 (244.86)	665.41 (217.39)
Quality of revised articles (1-low to 7-high)	4.76 (1.15)	4.31 (1.75)	4.89 (1.25)	4 (1.19)	4.49 (1.23)	4.21 (1.39)
Number of character changes from original to revised articles	432.15 (175.37)	374.45 (138.64)	423.93 (191.04)	460.94 (124.91)	514.33 (180.08)	436.52 (185.93)

Table 4. The means (and standard deviations) of the length and quality of original vs. revised articles, and the number of edits in each condition.

Dependent variable	Anonymous M (SE)	Requester M (SE)	Peer M (SE)
Annoyance	2.78 (.29)	3.52 (.35)	3.48 (.37)
Frustration	2.53 (.28)	3.34 (.33)	3.36 (.35)
Usefulness of feedback	5.98 (.16)	5.49 (.20)	5.91 (.21)

Table 6. Means and standard errors of participants' annoyance, frustration, and perceptions of feedback and feedback provider in 3 feedback source conditions.

RESULTS

Table 4 shows the average length and quality of the original articles and the revised articles, as well as average number of edits, in each condition.

We compared participants' emotions, perceptions of the feedback and the feedback source, the amount of changes they made to their writing, and the improvement in the quality of their writing across the six experimental conditions. We controlled for the participants' own perception of the feedback source's identity, their self-identification as a writer, and their self-efficacy.

Positive language improved reception & performance

Our first research question asks how the use of affective language in feedback influences feedback reception, behaviors, and performance. Participants in the with-positive language condition rated the feedback significantly more positive ($M = 3.75$, $SE = 0.16$) than those in the without-positive language condition ($M = 2.34$, $SE = 0.15$, $F[1, 197]$

Dependent variable	With M (SE)	Without M (SE)	F test F [1, 195]
Annoyance	2.93 (.29)	3.59 (.29)	7.03**
Frustration	2.71 (.18)	3.44 (.27)	9.70**
Happiness	4.92 (.22)	4.19 (.22)	14.57***
Enthusiasm	4.70 (.26)	3.83 (.27)	14.65***
Fairness of feedback	5.24 (.23)	4.67 (.24)	7.98***
Source's fairness	5.11 (.22)	4.23 (.22)	22.75***
Source's expertise	4.40 (.21)	3.79 (.21)	11.20***
Levenshtein distance	451.89 (27.51)	510.75 (27.09)	6.32**

Table 5. Better reception to feedback, but shorter Levenshtein (edit) distance in the with-positive-language condition than without-positive-language condition. 2-tailed p values: ** $p < .01$, * $p < .001$.**

Figure 2. Quality increased for revised articles in the positive language condition, and decreased for those in the condition without positive language

$= 43.94$, $p < 0.001$), confirming our manipulation.

After seeing the feedback, participants felt significantly less annoyed, less frustrated, more enthusiastic and happier when they received critiques with positive language than without positive language (see Table 5).

Participants perceived feedback in the with-positive-language condition as significantly fairer than those in the without-positive-language condition. Likewise, participants in the with-positive condition also perceived the feedback *source* as significantly more fair and knowledgeable about article writing than those in the without-positive condition (see Table 5).

In terms of their work performance, participants in the with-positive language condition made fewer edits on their work than those in the without-positive language condition (see Table 4). However, the quality of their writing improved (Mean change in quality = 0.48, $SE_{\text{change}} = 0.31$, $t[93] = 1.99$, $p = 0.05$) while the quality of participants' articles in the without-positive language condition decreased (mean change in quality = -0.29, $SE_{\text{change}} = 0.31$, $t[107] = 1.98$, $p = 0.05$, see Fig 2). There is a significant difference between the with-positive and without-positive conditions in terms of the change in writing quality ($F[1, 195] = 5.04$, $p = 0.01$).

Anonymous feedback improved reception

The second research question asks how the identity of the feedback source influences feedback reception, behaviors, and performance. Participants were less annoyed ($F[2,195] = 3.29, p = 0.03$) and less frustrated ($F[2,195] = 4.63, p = 0.01$) if they received anonymous feedback than if they were told the feedback came from the requester or peers (see Figure 3 & Table 6). We did not tell participants in the anonymous condition the identity of the feedback source, so we asked them to specify who provided the feedback. Most participants (74.29%) in the anonymous feedback condition believed the feedback came from the requester. Only 17.14% of these participants thought the feedback source came from a peer.

Anonymous feedback was perceived as more useful than requester feedback ($F[2,195] = 4.07, p = 0.01$, see Figure 3 & Table 6). Requesters were perceived as more fair ($M = 4.97, SE = 0.14$) than peers ($M = 4.16, SE = 0.19, F[1,195] = 4.61, p = 0.01$). All other differences were not significant.

We did not find any effect of feedback source identity (both who we told the participants, and who the participants thought was the feedback source) on the amount of changes participants made to their articles ($F[2,195] = 0.39, p = 0.89$ for feedback source condition, $F[2,195] = 2.23, p = 0.10$ for perceived identity of feedback source) or the change in quality ($F[2,195] = 1.36, p = 0.26$ for feedback source condition, $F[2,195] = 0.96, p = 0.35$ for perceived identity of source).

Near significant interaction between positive language and feedback source's identity

We found a near significant interaction effect ($F[2,195] = 2.88, p = 0.06$) between the affective language conditions and the feedback source identity on the amount of changes made to articles. Feedback from an anonymous source or a peer did not trigger significant changes to the writings, regardless of the language used in the feedback. Meanwhile, feedback with positive language from the requester triggered significantly fewer changes to the writings than feedback without positive language from the requester.

Figure 3. Participants exhibited less annoyance and frustration for anonymous feedback than from a requester or peer

No relationship between revision effort and quality improvements in writing

We conducted a linear regression to test the effect of Levenshtein edit distance on the level of change in writing quality. There was no significant association between how many changes a participant made to their work and the improvement in work quality ($F[1,203] = 0.11, p = 0.99$).

We also reviewed the edits of a random sample of 20 participants. We found that some participants added more content to the original submission or moved their sentences around instead of changing the meaning or framing of their writing. An example is changing “*Sonia strives to recreate beauty from discarded objects*” to “*Sonia, a Tasmanian artist with a passion for children’s toys, strives to recreate beauty from discarded dolls.*” These edits may not improve writing quality as much as reframing or conceptually pivoting (e.g. changing “*Sonia strives to recreate beauty from discarded objects*” to “*Sonia is transforming the ways children and adults interact with toys through her message of saving both the planet and childhood innocence*”).

Perceptions of feedback provider mediates the effect of positive language on improvement in writing quality

For RQ1, we hypothesize that perceived source’s expertise mediates the effect of affective language on writing quality improvement. We conducted a linear regression on writing quality improvement, with both affective language and perceived source’s expertise as predictors. We found that the effect of affective language became insignificant ($F[1, 202] = 3.34, p = 0.07$), while the effect of perceived source’s expertise remained significant ($\beta = 0.20, SE = 0.10, F[1,202] = 4.10, p = 0.04$). Controlling for affective language, the more expertise participants attributed to the feedback source, the more improvement they achieved in their writing quality after feedback. We conducted a mediation test proposed by Preacher & Hayes [34] and found that the indirect effect of positive affective language through perceived source’s expertise was significant ($\beta = 0.11, 95\% CI = [0, .26], p = 0.05$). Therefore, consistent with our hypothesis, the with-positive language manipulation indirectly increased quality improvement through its positive effect on perceived feedback source’s expertise (Figure 4).

Figure 4. Effect of positive language on quality improvement mediated by perceived sources’ expertise.

DISCUSSION

Online feedback exchange is easier and more popular with many platforms that can shape the styles, language uses, and source transparency of feedback. This study investigated the effects of two factors—*affective language* and *feedback source identity*—on feedback receptivity, effort, and performance after feedback. We explored how feedback systems can be designed to improve the effectiveness and experience of feedback exchange and reduce negative reactions. Consistent with previous work about *affective language* [29, 30], we found that adding positive language improved participants' emotional responses and reception to feedback. Moreover, we found that feedback without positive language increased editing (compared to the *with-positive language condition*), but that these additional edits did not lead to better quality work. Indeed, participants who received feedback without positive language ended up with lower quality work.

More edits, but poorer performance

There are several possible explanations for why feedback without positive language increased editing but reduced writing quality compared to feedback prepended with positive language. Without the positive language, participants may have been more aroused and stressed than those who received feedback with positive language [9, 38]. Such activation might have amplified their reactions to the feedback, resulting in more edits. However, stress can degrade people's attention to detail [26]; greater negative affect requires cognitive resources for emotional regulation and task-irrelevant ruminations, thus reducing the focus on performance improvement [33]. Poorer revision quality may also reflect antisocial motivations or intentional counterwork behavior induced by greater annoyance and frustration in response to negatively framed feedback and poorer perceptions of feedback source [2, 33].

These findings highlight an important distinction in addressing the goal of constructive feedback: that quantity of work iteration is not necessarily indicative of quality. In fact, while some previous studies [e.g., 13] found that the amounts of edits were associated with increase in writing quality after feedback; others [e.g., 37] did not find any relationship between edits and writing quality. In line with Stevenson et al., we did not find any correlation between edits made and increase in writing quality. As the qualitative review of a sample of 20 pairs of original and edited articles in our study shows, some edits may be too superficial to effectively change the impact of the articles. Effective feedback should not be designed to induce greater behavioral responses per se, but rather oriented towards improving the reflective processes.

Positive language improved the impacts of feedback

Consistent with Nelson & Schunn [29], we found that participants who received feedback with positive *affective language* achieved better improvement in work quality than those who received feedback without positive language,

despite making fewer edits. The mediation analysis provides one explanation for why positive language led to greater improvement in work quality: positive statements increased participants' perception of the sources' expertise. Increased perception of the feedback sources' expertise in turn might have encouraged participants to improve their work further [21]. Our results confirm speculation in previous studies [10, 11], and bridge a gap in the literature about the relationship between participants' perceptions of a feedback source and their subsequent performance.

Anonymous feedback perceived as fairer & more useful than identified feedback

Our study compares the effects of anonymous sources and those of identified sources with different power statuses (requester or peer) on feedback reception. Consistent with previous work [10, 21], we found that when the identity of the feedback source was known, a feedback source of higher status (the requester) was perceived as fairer than one of lower status (another participant). However, unlike previous studies, we did not find any significant difference in the perceived usefulness or fairness of the feedback from the requester and from another participant of similar status. This may be because in our study, participants did not think the requester's expertise in the writing task was different from that of a peer.

We found that, regardless of who participants assumed as the feedback source, anonymous feedback sources reduced participants' annoyance and frustration after getting feedback, compared to knowing the feedback source's identity. Moreover, even feedback known to be from the requester was perceived as less useful than feedback from an anonymous source. Participants might have perceived anonymous feedback sources as more honest in giving the feedback than identified sources, as suggested by Lu & Bol [22]. Our results highlight the benefits of source anonymity when delivering feedback.

Design implications

Our study explores two feedback factors directly connected to the design features of feedback systems: *affective language* (framing feedback with or without a positive "wrapper" statement) and the source's identity (saying the feedback came from a requester, a peer, or an anonymous provider). First, we found that positive language improved participants' emotional response and reception towards the feedback. Moreover, the *with-positive language condition* led participants to quality improvements on their work. Future feedback systems can consider building a pool of positive "wrapper" statements such as the ones that we used and attach these statements (if necessary) to feedback before delivering critiques to participants to enhance the effectiveness of feedback.

Second, we found that if the feedback source was anonymous, participants were less annoyed and frustrated when receiving feedback, and regarded the feedback as more useful than when they knew the feedback source. This

result suggests that when delivering feedback, the power status of the sources should perhaps be hidden for better feedback reception. We also found that the effectiveness of feedback in helping participants improve their performance depended partly on the impressions participants had about the feedback source. Zhu et al. [44] recommended that in Wikipedia, senior members should deliver feedback to newcomers. We would like to add that feedback systems should carefully select features that display the feedback source's identity during feedback exchanges. This is because certain features about the feedback sources (such as member ranking, moderator status, or previous task-related experiences) can provide cues for feedback receiver to form impressions about the source's power status or expertise level [25]. Our results show that such impressions about the source's social status or expertise can influence feedback reception. More extensive research about this is needed to identify which features should or should not be presented and when.

LIMITATIONS & FUTURE WORK

Several aspects of the feedback interaction in our results suggest fruitful areas for future research.

Diverse settings

Our study was conducted in Mechanical Turk. The relationship between requesters and Turkers is short-term, which may mean that Turkers have personal investment in their work and less motivation to respond to feedback. Managers and employees in a traditional workplace have a longer term, sustained relationship. Different relationships (e.g. teacher/student, manager/employee, mentor/mentee) and associated settings may change responses to feedback. Future work should consider the influence of affective language and source identity on receptivity for other types of feedback relationships in diverse settings.

We recruited Mturk workers located in the US. According to several studies [16, 27, 32], Mturk workers in the US are more similar to the general US adult population than college students in US universities. However, these studies also found that Mturk workers are slightly younger and more technologically advanced than the general US population. Therefore, more studies on different populations and with different feedback platforms are needed before we can generalize the findings beyond Internet savvy US adults.

Other tasks

Our study's task—creative writing in response to a specific prompt—was very structured and represents only one of many possible types of tasks where people can receive feedback on their work. Future work should examine whether our results are generalizable to other task domains, such as visual design or project-based learning.

Learning versus work improvement

Our study measured the number of edits made in response to feedback and the quality of the work after revisions. However, we did not directly measure how the exact feedback statements affected participants' article edits. Future

studies should consider evaluating the influence of feedback on participants' post-feedback behaviors to better describe the mechanisms underlying feedback reception.

In addition, our current measures do not address the learning gains in response to feedback. Future work should examine the types of changes people make in response to feedback and how they are associated with skill development over time.

Linguistic structure for feedback

Compared to the positive “wrapper” statements, the without-positive “wrapper” statements were slightly more directive, which may explain the difference in edit actions by participants in the without-positive language condition. Future studies should consider less directive wrappers (such as negative statements like “this article is subpar” or neutral statements like “here is my feedback”) to better isolate the effect of positive language on feedback reception.

Social transparency

In the anonymous condition, where we did not tell participants the feedback's origin, most participants assumed the feedback came from the requester, perhaps because of the experimental setting of MTurk. We controlled for the perceived identity of the feedback source in our analysis. However, such control might have reduced the power of some of our analyses. Future studies can use a more realistic setting where most participants believe that anonymous feedback comes from unknown sources, such as a member of a large online community.

We only experimented with one aspect of the identity of feedback sources to disclose to participants: power status. In future studies, we intend to further explore how knowledge about other aspects of feedback sources will influence receivers' attitudes and behaviors towards feedback and their performance post feedback. These other aspects of feedback sources can include: the sources' expertise in the task domain, time spent reviewing the work, resources used to inform critiques on the work, and factors regarding the feedback receivers, such as time and number of iterations the receivers went through on the task.

CONCLUSION

Feedback can sometimes be painful, but can reveal important weaknesses and directions for improvement. This study examined and explained the effects of affective language and source identity on people's receptivity to feedback, their behaviors in response to feedback, and their overall performance. We found that using positive language such as “good work so far” reduced participants' annoyance and frustration, and improved their attitudes towards both the feedback and the source, which in turn, helped improve participants' post-feedback work quality. We also found that anonymous feedback was perceived more positively than feedback from a peer or from a work requester. Our findings contribute to theoretical work about feedback

reception and task performance, and provide implications for the design of feedback systems.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their feedback on earlier versions of the paper. This work was supported by NSF grants SOCS-1208382, CYBR-1122320, CAREER-1122206, CNS-1221006, VOSS-1322278, CNS-1345305, CNS-1040757, CNS-1040800, and CNS-1040801.

REFERENCES

1. Chris Argyris. 1991. Teaching Smart People How to Learn. *Harvard Business Review* 69, 99-109.
2. Frank D. Belschak and Deanne N. Den Hartog. 2009. Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology* 58, 2: 274-303.
3. Crystal Project Inc. 2014. Crystal. Email Application. Retrieved Sep 4th 2015 from <https://www.crystalknows.com/>
4. Susan J. Ashford, and Anne S. Tsui. 1991. Self-Regulation for Managerial Effectiveness: The Role of Active Feedback Seeking. *Academy of Management Journal*, 34, 2: 251–80.
5. Jennifer N. Belding, Karen Z. Naufel, and Kentaro Fujita. 2015. Using High-Level Construal and Perceptions of Changeability to Promote Self-Change Over Self-Protection Motives in Response to Negative Feedback. *Personality & Social Psychology Bulletin* 41, 822-838.
6. Gilad Chen, Stanley M. Gully, and Dov Eden. 2001. Validation of a New General Self-Efficacy Scale. *Organizational Research Methods* 4, 1: 62-83.
7. Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction* 17, 4, Article 18, 24 pages.
8. Steven P. Dow, Anand Kulkarni, Scott Klemmer, & Bjorn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of CSCW '12*, 1013-1022.
9. Carol Sorich Dweck & Ellen L. Leggett. 1998. A Social Cognitive Approach to Motivation and Personality. *Psychological Review* 95, 256–73.
10. Donald B Fedor, Walter D Davis, John M Maslyn, and Kieran Mathieson. 2001. Performance improvement efforts in response to negative feedback: the roles of source power and recipient self-esteem. *Journal of Management* 27, 1: 79-97.
11. Donald B. Fedor, Robert W. Eder, Ronald M. Buckley. 1989. The contributory effects of supervisor intentions on subordinate feedback responses. *Organizational Behavior and Human Decision Processes* 44, 3: 396-414.
12. Stacey R. Finkelstein and Ayelet Fishbach. 2012. Tell Me What I Did Wrong: Experts Seek and Respond to Negative Feedback. *Journal of Consumer Research* 39, 1: 22-38.
13. Steve Graham, Charles MacArthur, and Shirley Schwartz. 1995. Effects of goal setting and procedural facilitation on the revising behavior and writing performance of students with writing and learning problems. *Journal of Educational Psychology* 87, 2: 230-240.
14. John Hattie and Helen Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 1: 81–112.
15. Andrew F. Hayes and Klaus Krippendorff (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 1, 1: 77–89.
16. Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 453-456.
17. Avraham Kluger and Angelo DeNisi. 1998. Feedback interventions: toward the understanding of a double-edged sword. *Current Directions in Psychological Science* 7, 3: 67–72.
18. Stavros Konstantinidis. 2005. Computing the Levenshtein distance of a regular language. In *Proceedings of IEEE ISOC Information Theory Workshop 2005 on Coding and Complexity*, 113-116.
19. Minjung Koo and Ayelet Fishbach. 2008. Dynamics of Self-Regulation: How (Un)Accomplished Goal Actions Affect Motivation. *Journal of Personality and Social Psychology* 94, 2: 83–95.
20. Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2015. PeerStudio: Rapid Peer Feedback Emphasizes Revision and Improves Performance. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S '15)*, 75-84.
21. Kwok Leung, Steven Su, and Michael W. Morris. 2001. When is Criticism Not Constructive? The Roles of Fairness Perceptions and Dispositional Attributions in Employee Acceptance of Critical Supervisory Feedback. *Human Relations* 54, 1155-1187.
22. Ruiling Lu and Linda Bol. 2007. A Comparison of Anonymous Versus Identifiable e-Peer Review on College Student Writing Performance and the Extent of Critical Feedback. *Journal of Interactive Online Learning* 6, 2: 100-115.

23. Kurt Luther, Amy Pavel, Wei Wu, Jari-lee Tolentino, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2014. CrowdCrit: crowdsourcing and aggregating visual design critique. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW Companion '14), 21-24.
24. Laura MacLeod. 1999. Computer-aided peer review of writing. *Business Communication Quarterly* 62, 3: 87-95.
25. Jennifer Marlow and Laura Dabbish. 2014. From rookie to all-star: professional development in a graphic design social networking site. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW '14), 922-933.
26. Jennifer Marlow and Laura Dabbish. 2013. Activity traces and signals in software developer recruitment and hiring. In *Proceedings of the 16th ACM conference on Computer supported cooperative work* (CSCW '13), 145-156.
27. Winter Mason and Siddharth Suri. Conducting behavioral research on Amazon's Mechanical Turk. 2012. *Behavior research methods* 44, 1: 1-23.
28. Bruce S. McEwen and Robert M. Sapolsky. 1995. Stress and cognitive function. *Current Opinion in Neurobiology* 5, 2: 205-216.
29. Melissa M. Nelson and Christian D. Schunn. 2008. The Nature of Feedback: How Different Types of Peer Feedback Affect Writing Performance. *Instructional Science* 37, 4: 375 – 401.
30. Christine M. Neuwirth, Ravinder Chandhok, David Charney, Patricia Wojahn, and Loel Kim. 1994. Distributed collaborative writing: a comparison of spoken and written modalities for reviewing and revising documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '94), 51-57.
31. Duyen T. Nguyen, Laura A. Dabbish, and Sara Kiesler. 2015. The Perverse Effects of Social Transparency on Online Advice Taking. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 207-217.
32. Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on amazon mechanical turk. 2010. *Judgment and Decision making* 5, 5: 411-419.
33. Reinhard Pekrun. 1992. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology* 41, 4: 359-376.
34. Kristopher J. Preacher and Andrew F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods* 40, 3: 79-891.
35. Sheng-Chau Tseng and Chin-Chung Tsai. 2007. Online peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education* 49, 4: 1161-1174.
36. R. William Soukoreff and I. Scott MacKenzie. 2001. Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic. In *CHI '01 Extended Abstracts*, 319-320.
37. Marie Stevenson, Rob Schoonen, & Kees de Glopper. 2006. Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing* 15, 3: 201-233.
38. Jakko van der Pol, Ineke van den Berg, Wilfried F. Admiraal, Robert Jan Simons. 2008. The nature, reception, and use of online peer feedback in higher education. *Computers & Education* 51, 4: 1804-1817.
39. Anna C. K. van Duijvenvoorde, Kiki Zanolie, Serge R. A. B. Rombouts, Maartje E. J. Raijmakers, and Eveline A. Crone. 2008. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *The Journal of Neuroscience* 28, 38: 9495-9503.
40. C. W. Von Bergen, Martin S. Bressler, and Kitty Campbell. 2014. The sandwich feedback method: not very tasty. *Journal of Behavioral Studies in Business* 7, 1-13.
41. David Watson, Lee A. Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6: 1063 - 1070.
42. Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. 2015. A Classroom Study of Using Crowd Feedback in the Iterative Design Process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 1637-1648.
43. Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW '14), 1433-1444.
44. Haiyi Zhu, Amy Zhang, Jiping He, Robert E. Kraut, and Aniket Kittur. 2013. Effects of peer feedback on contribution: a field experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 2253-2262.