“What’s going on inside their head?”: Students think aloud while processing feedback and an exemplar

Authors Blinded for Review

Abstract

Feedback is an essential part of the learning process, yet few studies have explored the cognitive processes used by students when reviewing authentic classroom feedback. Thus, we employed a think-aloud protocol where graduate students (*N*=35) reviewed assignment feedback with or without an exemplar. Participants who received both feedback and the exemplar not only engaged in high-level cognitive processing of their feedback and were more likely to make local inferences about the individual texts, they were also more likely to read all of the available texts, integrate information across them, and assess their performance. These findings suggest that inclusion of an exemplar may motivate students to more actively reevaluate their work, which may inform future feedback literacy research and interventions.

Introduction

Feedback is an essential part of the learning process, as it is believed to have one of the "highest influences on achievement" in academic settings (Hattie & Timperley, 2007, p. 83). It allows learners to assess their current understanding and make progress towards a larger end goal in a given domain. While its importance is clear, researchers are still trying to understand why students benefit differentially from the feedback they receive (Kluger & DeNisi, 1996; Wisniewski et al., 2020).

Previous research points to various factors that can contribute to or disrupt the benefits of feedback. Characteristics of the feedback itself (Hattie & Timperley, 2007), prior knowledge of the learner (Fyfe & Rittle-Johnson, 2016; Shute, 2008), affective responses (Mahfoodh, 2017; Pekrun et al., 2014), and overall engagement with the feedback (Winstone et al., 2017) are just some of the reasons that have been explored. While numerous frameworks or models of academic feedback exist (see Lipnevich & Panadero, 2021), two main views are prominent: those that center feedback as information an instructor provides learners to support their learning (i.e., the transmission view), and more contemporary frameworks where feedback is seen as a dialogue or a process where the student is an active agent and how they understand and engage with the information are the most important contributors to learning (i.e., the feedback literacy perspective). Feedback literacy includes the "understandings, capacities, and dispositions that learners need to maximize benefits from feedback processes" (Nieminen & Carless, 2023, p. 1382). For the educator, feedback literacy includes the ability to support students through this process, by designing and facilitating the proper uptake and use of feedback by their learners (Boud & Dawson, 2021; Carless & Winstone, 2020). This is an important shift because it

suggests that learners and educators alike can increase their skills as agents in the feedback process.

Current Study

Although work on feedback literacy is growing, there is still limited research in this domain. Further, to really understand what actions students take when interacting with feedback, it is important to observe that process, something that has been scarcely done (Byun et al., 2023; Mahfoodh, 2017; Máñez et al., 2019). In one of the few studies to explore the cognitive processes that are employed by students as they are processing feedback, Byun and colleagues (2023) observed that undergraduate students tended to rely on general text-processing strategies (i.e., reading/re-reading the feedback), rarely generating inferences or monitoring their own understanding. Further, when these higher-level processes occurred, students tended not to connect the feedback to their actual work/the assignment. These findings suggest that students use passive strategies when interacting with feedback, even when that feedback is important for their academic success.

The current study builds on Byun and colleagues' (2023) work. We utilized a think-aloud protocol during student processing of feedback on an authentic classroom assignment. This allowed us to compare students' cognitive processes when they receive their assignment feedback with an exemplar versus receiving feedback only. The inclusion of an exemplar may spur deeper processing of the text because it can resolve any confusion students may have about what was absent in their work and can provide a clear example of how their work could be improved (Hawe et al., 2017). Further, we hypothesized that when students read the exemplar, they would be more likely to integrate the material across the various texts, connecting the exemplar to their feedback and their assignment. This method can deepen our understanding of

the strategies that students employ when processing various forms of feedback, which can inform future feedback literacy interventions.

Methods

Participants

Participants were 45 graduate students recruited from a 5000-level educational psychology course at a large Midwestern university. Out of these students, one never booked a think-aloud session, one no-showed, four did not consent to participate, two were removed from analyses because they did not think aloud as prompted, and two were not coded because of researcher error, leaving 35 total participants in the final analysis. IRB approval was obtained.

Procedure

After completing an article summary assignment for the course, students signed up for a think-aloud session with one of six researchers. Students were randomly assigned to either a Feedback Only (FO; $n = 18$) or a Feedback & Exemplar (F&E; $n = 17$) condition. In the FO condition, students met with a researcher and received their feedback on the assignment. In the F&E condition, students met with a researcher and received their feedback on the assignment *and* an exemplar submission that modeled a successful assignment. The feedback was provided using in-text comments from the course instructor and teaching assistant. The researchers did not provide any additional feedback and were only there to administer the think-aloud. All think-aloud sessions took place over Zoom and were recorded. Researchers followed a script to ensure consistency across participants.

When participants arrived on Zoom, the researcher gave a brief overview of the experiment, instructed participants on how to engage in a think-aloud, and practiced with two unrelated warm-up questions. The think-aloud directions were modified from Ericsson and

Simon (1993). Those in the FO condition were instructed to open their assignment feedback and think aloud. Those in the F&E condition were instructed to open both their assignment feedback *and* the exemplar. They were told they could move freely between the documents while thinking aloud. If participants were silent for more than 5 seconds during the practice or experiment, they were prompted with the following: "What are you thinking right now?"

Think-Aloud Coding

All think-alouds were transcribed and quality-checked. Researchers noted when utterances were in reference to the original assignment, the feedback, or the exemplar. If participants were using their computer mouse to indicate where they were in the document, these actions were also noted in the transcription.

Prior to assigning codes to the participants' utterances, all think-alouds were broken into feedback units. Feedback units constituted one or more clauses containing a meaningful utterance. Units varied in length depending on the nature of what was said. Where appropriate, feedback units were associated with the source of the utterance (i.e., assignment, feedback, or exemplar). Think-alouds were coded by all authors, and consensus was reached for each think-aloud unit before proceeding to the next.

A codebook was created to categorize participants' responses. The codebook was first utilized by Byun et al. (2023) and drew on prior research, specifically the work on inferencing while processing scientific texts (Cromley et al., 2010), and the mental processes involved in processing feedback (Máñez et al., 2019). There were four main categories of responses: Processing Text, Meaning Making, Monitoring, and Self-Regulation. Although the current study utilized the same four categories, it added much more detailed subcategories such as affective reactions, different types of inferences, and intentionality of text-processing. Table 1 outlines

each overarching category and their additional sub-codes. Utterances that did not fit into any category were coded as “Other.”

Results

We first reviewed the **average proportion of utterances** in each think-aloud category to see if there were differences across conditions (See Table 2). Significant differences were only found in the Processing Text main category, specifically for Reading Feedback and Navigating with Intent. The FO group had a significantly greater average proportion of utterances for reading their feedback aloud compared to the F&E group, $F(1, 33) = 17.067$, $p < .001$, $\eta^2 = .341$. The F&E group had a significantly greater average proportion of utterances for navigating with intent compared to the FO group, $F(1, 33) = 14.568$, $p < .001$, $\eta^2 = .306$. No significant differences were found when disaggregating categories by their plus and minus distinctions (e.g., Affective Reactions by positive/negative or Assessing Comprehension/Memory by expressing understanding/lack thereof). Both groups had the highest average proportion of utterances for Processing Text and Monitoring (>30% for each), followed by Meaning Making (>20%), and Self-Regulation (<2%). This focus on overall proportions was consistent with the analysis conducted by Byun et al. (2023) and in some ways allows us to account for the different amounts of text across the two conditions. However, simply relying on proportions does obscure the actual behaviors of our participants, particularly of the F&E group. Relying on proportions instead of counts to represent the strategies employed by the group naturally inflates the *monitoring* proportions for the F&E group. Therefore, we also analyze the count variations across the two groups to truly illuminate how the participants were engaging with the different materials.

We calculated the **average number of utterances (counts)** in each think-aloud category to see if there were differences across conditions (See Table 3). Significant differences were found in the Processing Text main category and the Reading Assignment and Navigating with Intent subcategories. It is not surprising that the F&E group had higher counts of text processing and navigation, as the addition of an exemplar increases the amount of text and documents for review. Of importance is the F&E group having higher counts for reading their assignment. This suggests that the inclusion of an exemplar motivated the students to review their assignment to better understand the feedback in light of the exemplar.

Significant differences were also found in the Meaning Making main category and the Local Inference and Global Inference subcategories. That is, students in the F&E group were making more inferences within and across texts when an exemplar was provided. In terms of Monitoring, the only significant difference was found in the Personal Assessment subcategory, with F&E students uttering more verbalizations about their performance. No significant differences were found when disaggregating categories by their plus and minus distinctions (e.g., Affective Reactions by positive/negative or Assessing Comprehension/Memory by expressing understanding/lack thereof). No significant differences were found in the Self-Regulation and Other categories.

Discussion

Prior research highlighted students' low-level processing strategies and lack of integration of information when reviewing feedback on an authentic classroom assignment (Byun et al., 2023). The current research extended this work by experimentally manipulating the context in which the feedback was provided. When students received the addition of an exemplar, we observed a significantly improved depth of processing. It is important to note that students were

not instructed how to use the exemplar, or even to use it, they were simply given feedback and the exemplar in tandem. Even though students in the F&E condition were "given" the extra task of reading the exemplar, this should not be seen as a burden or additional cost, as they did not report any more emotional distress or levels of confusion with the extra material.

Participants in the F&E group not only engaged in high-level cognitive processing of their feedback and were more likely to make local inferences about the individual texts, they were also more likely to read all of the available texts, integrate information across them (i.e., making global inferences), and assess their own performance. These differences illustrate a level of feedback literacy that was not observed in the FO condition or by the participants in Byun et al. (2023). This suggests that the addition of an exemplar motivates the students to review and reevaluate their performance, something that does not seem to occur as often in the FO condition, even though feedback alone should theoretically offer the same opportunities. This finding is consistent with research focused on multi-text comprehension which suggests that when there is conflicting information in text, readers will employ higher-level strategies to resolve that conflict (Stadtler et al., 2013). But again, both groups were provided with conflicting text (i.e., feedback and their assignment), but perhaps the exemplar more explicitly highlighted the conflict, ultimately shaping their motivated processing (Lyu et al., 2023; Schoor et al., 2023; van den Broek et al., 2001).

The current study provides greater insight into the cognitive processes of students as they process authentic feedback and can be used to inform ongoing feedback literacy interventions and research. Future work should explore the quality of inferences and monitoring of students, as this is not something that proportions or counts can illustrate.

References

Boud, D., & Dawson, P. (2021). What feedback literate teachers do: An empirically-derived competency framework. *Assessment & Evaluation in Higher Education, 48*(2), 158–171. https://doi.org/10.1080/02602938.2021.1910928

Byun, C., Steadman, C., Hufnagle, A. S., Weisen, S., Christianti, D., Troy, A., Wahyuni, I., Stewart, J., & Van Boekel. M. (2023, April 13-16). *Examining students' strategies for processing and using feedback: A think-aloud study* [Paper presentation]. American Education Research Association (AERA) Conference, Chicago, IL, United States. https://doi.org/10.3102/2004038

Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education, 28*(1), 150–163. https://doi.org/10.1080/13562517.2020.1782372

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology, 35*(1), 59-74. https://doi.org/10.1016/j.cedpsych.2009.10.002

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* MIT Press.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*(1), 82–97. https://psycnet.apa.org/doi/10.1037/edu0000053

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://psycnet.apa.org/doi/10.3102/003465430298487

Hawe, E., Lightfoot, U., & Dixon, H. (2017). First-year students working with exemplars: promoting self-efficacy, self-monitoring and self-regulation. *Journal of Further and Higher Education, 43*(1), 30–44. https://doi.org/10.1080/0309877X.2017.1349894

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://psycnet.apa.org/doi/10.1037/0033-2909.119.2.254

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.720195

Lyu, B., McCrudden, M. T., & Bohn-Gettler, C. (2023). The effects of purpose instructions and strategy-focused instructions on reading processes and products. *Reading and Writing*, 1-24. https://doi.org/10.1007/s11145-023-10464-w

Mahfoodh, O. H. A. (2017). "I feel disappointed": EFL university students' emotional responses towards teacher written feedback. *Assessing Writing, 31*, 53-72. https://doi.org/10.1016/j.asw.2016.07.001

Máñez, I., Vidal-Abarca, E., Kendeou, P., & Martínez, T. (2019). How do students process complex formative feedback in question-answering tasks? A think-aloud study. *Metacognition and Learning, 14*, 65–87. https://psycnet.apa.org/doi/10.1007/s11409-019-09192-w

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education, 85*(6), 1381-1400. https://doi.org/10.1007/s10734-022-00895-9

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction, 29*, 115-124. https://doi.org/10.1016/j.learninstruc.2013.09.002

Schoor, C., Rouet, J. F., & Britt, M. A. (2023). Effects of context and discrepancy when reading multiple documents. *Reading and Writing, 36*(5), 1111-1143. https://doi.org/10.1007/s11145-022-10321-2

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction, 31*(2), 130-150. https://doi.org/10.1080/07370008.2013.769996

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*(8), 1081-1087. https://doi.org/10.3758/BF03206376

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17-37. https://doi.org/10.1080/00461520.2016.1207538

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 487662. https://doi.org/10.3389/fpsyg.2019.03087

Table 1

Think-Aloud Coding Scheme Across Categories

Processing Text	
Reading and Re-Reading	Can be assignment, feedback, or exemplar verbatim
Skimming	Can be assignment, feedback, or exemplar but omitting words, using synonyms, etc.
Paraphrasing	Can be assignment, feedback, or exemplar but participant is using their own words to repeat the text without adding any additional meaning
Navigating the Text with Intent	Articulating a desire/intent to engage in movement between the text or texts with explicit statement of utility behind the movement
Meaning Making	
Local Inference	Inference about what a text (assignment, feedback, or exemplar) means; can be correct or incorrect
Global Inference	Integrating across multiple pieces of texts (assignment, feedback, exemplar, or multiple instances in the same text); can be correct or incorrect
Ambiguous Inference	Participant is clearly generating an inference but the source and/or its connections are not made transparent
Knowledge Elaboration	Activation of prior knowledge/experience that is not explicitly stated in the text (assignment, feedback, or exemplar)
Monitoring	
Performance Comparison	Verbalizations about the relationship between the participant's accuracy and a standard (e.g., rubric criteria, grammar, exemplar, feedback); can suggest utility/usage but needs to be explicitly stated
Personal Assessment	Verbalizations about the participant's own performance
Affective Reaction	Emotion-based verbalizations about the feedback or feelings because of receiving feedback; need to state the emotion/a feeling; can be positive or negative
Attribution	Verbalizations about possible explanations regarding the cause of the participant's performance
Assessing Comprehension/Memory	Active instances of participants attempting to comprehend or understand the feedback given and/or verbalizations of perceived comprehension; can be expressing understanding or lack thereof
Self-Regulation	
	Verbalizations about the participant's plans for how they will use their feedback moving forward; Includes setting goals and identifying strategies
Other	
	Participant's utterance did not fit into categories reported above, was unclear, or irrelevant to the experimental task

Table 2

By Proportions, Means and Standard Deviations of Think-Aloud Codes Across Conditions

	Feedback Only (n = 18)	Feedback & Exemplar (n = 17)
Processing Text	.374 (*.105*)	.392 (*.125*)
Reading Feedback	**.279 (*.079*)**	**.174 (*.071*)**
Navigating with Intent	**.002 (*.007*)**	**.032 (*.032*)**
Meaning Making	.212 (*.108*)	.252 (*.098*)
Local Inference	.131 (*.117*)	.150 (*.084*)
Global Inference	.048 (*.040*)	.073 (*.049*)
Ambiguous Inference	.001 (*.005*)	.001 (*.003*)
Knowledge Elaboration	.033 (*.035*)	.028 (*.037*)
Monitoring	.378 (*.145*)	.311 (*.110*)
Performance Comparison	.055 (*.042*)	.043 (*.048*)
Personal Assessment	.038 (*.057*)	.058 (*.053*)
Affective Reaction	.058 (*.064*)	.034 (*.041*)
Attribution	.107 (*.058*)	.079 (*.055*)
Assessing Comprehension/Memory	.118 (*.075*)	.097 (*.079*)
Self-Regulation	.019 (*.021*)	.013 (*.021*)
Other	.017 (*.026*)	.024 (*.028*)

Note. Bolded items show statistically significant differences across conditions (p <.001).

Table 3

By Counts, Means and Standard Deviations of Think-Aloud Codes Across Conditions for Significant Results

	Feedback Only (n = 18)	Feedback & Exemplar (n = 17)	ANOVA Results with Effect Size
Processing Text	14.72 (*5.17*)	23.24 (*10.62*)	$F(1, 33) = 9.257$, $p = .005$, $\eta^2 = .219$
Reading Assignment	1.50 (*1.50*)	4.00 (*4.72*)	$F(1, 33) = 4.571$, $p = .040$, $\eta^2 = .122$
Navigating with Intent	0.11 (*0.32*)	2.00 (*2.40*)	$F(1, 33) = 10.977$, $p = .002$, $\eta^2 = .250$
Meaning Making	8.28 (*4.14*)	14.82 (*4.86*)	$F(1, 33) = 18.448$, $p < .001$, $\eta^2 = .359$
Local Inference	4.56 (*2.62*)	9.24 (*4.41*)	$F(1, 33) = 14.780$, $p < .001$, $\eta^2 = .309$
Global Inference	2.11 (*1.84*)	4.00 (*2.81*)	$F(1, 33) = 5.601$, $p = .024$, $\eta^2 = .145$
Personal Assessment	1.44 (*1.76*)	3.59 (*3.69*)	$F(1, 33) = 4.901$, $p = .034$, $\eta^2 = .129$